Appunti di Statistica Matematica
Inferenza Statistica Multivariata
Anno Accademico 2019/20*

1 Campioni e modelli statistici

Siano $(\Omega, \mathcal{A}, \mathrm{P})$ uno spazio di probabilità e $\underline{X} = (X_1, \ldots, X_n)$ un vettore aleatorio su Ω a valori in un insieme $\mathcal{X}$. L'insieme $\mathcal{X}$ è detto *spazio campionario*, il vettore $\underline{X} = (X_1, \ldots, X_n)$ *campione* e n *taglia campionaria*. Spesso per ogni $i = 1, \ldots, n$ si ha $X_i \in \mathbb{R}$, ma non necessariamente. Una realizzazione di $\underline{X}$, indicata con $\underline{x} = (x_1, \ldots, x_n)$, è detta *valore campionario/campionato* e rappresenta un dato osservato. Gli elementi di Ω sono le unità (campionarie) su cui viene osservato il valore $\underline{x}$.

Definizione 1.1 *Un campione è detto* casuale *se le sue componenti sono mutualmente indipendenti*[1] *e seguono la stessa legge.*

L'espressione campione *indipendente ed identicamente distribuito*, in breve i.i.d., è sinonimo di campione casuale. Un'espressione alternativa è copie indipendenti della stessa variabile aleatoria. La legge di probabilità di $\underline{X}$ determina il *modello probabilistico* e talora può essere interpretata come il processo generatore dei dati. I dati osservati sono interpretati come una realizzazione del campione. Scopo dell'inferenza statistica è utilizzare i dati per ottenere (informazioni su) il processo che li genera.
[Ricordare i campioni scambiabili.]

Esempio 1.2 Sia X una variabile aleatoria di legge Poisson(λ) con $\lambda > 0$ e siano $X_1, \ldots, X_n$ copie indipendenti di X. Allora $\underline{X} = (X_1, \ldots, X_n)$ è un campione casuale. In questo esempio se è noto λ, allora è noto anche il modello probabilistico. Si supponga che per $i = 1, \ldots, n$, la variabile aleatoria (v.a.) X_i rappresenti il numero di prodotti difettosi in una linea di produzione il giorno i-esimo di un certo periodo. La legge di Poisson è una delle distribuzioni utilizzate per modelizzare "eventi rari". La "stima" del valore di λ fornisce contemporaneamente informazioni su quanti difettosi ci sono in media al giorno e dà una misura dello scostamento dal valor medio: $\mathrm{E}(X_1) = \mathrm{Var}(X_1) = \lambda$. L'ipotesi di campione casuale è un'ipotesi sulla raccolta di dati, per esempio può indicare che la tecnica di raccolta dati è la stessa ogni giorno, e anche sulla linea di produzione, per esempio può indicare che le condizioni di lavorazione sono identiche ogni giorno.

Esercizio 1.3 Per il campionamento da un'urna con reimmissione/senza reimmissione, individuare il campione. E' casuale? Mancano informazioni per individuare $(\Omega, \mathcal{A}, \mathrm{P})$? In particolare si verifichi che un campionamento senza reimmissione è identicamente distribuito, ovviamente non è indipendente.

Definizione 1.4 1. *Un* modello statistico *per $\underline{X}$ è una famiglia di distribuzioni di probabilità su/per $\underline{X}$.*

2. *Un modello statistico $\mathcal{F}$ è detto* parametrico *se esiste $\Theta \subseteq \mathbb{R}^p$ ed una mappa (biettiva) $\Theta \longrightarrow \mathcal{F}$.*

Compito dell'inferenza statistica è selezionare un modello probabilistico all'interno del modello statistico utilizzando i dati (cfr. ridurre l'incertezza, raffinare la conoscenza...).

*Per correzioni e suggerimenti riccomagno@dima.unige.it. Una versione aggiornata sarà disponibile alla fine delle lezioni.

[1] Gli eventi $A_i \in \mathcal{A}$, $i \in \mathbb{N}$ sono mutualmente o congiuntamente indipendenti rispetto a P se $\mathrm{P}(\cap_{k\in J} A_k) = \prod_{k\in J} \mathrm{P}(A_k)$ per ogni $J \subset \mathbb{N}$ finito. Si veda anche Appendice 6.4.1.

Un modello può essere espresso in vari modi, e.g. tramite funzioni di ripartizione congiunte per $\underline{X}$

$$\mathcal{F} = \left\{F_{\underline{X}}(x_1, \ldots, x_n) : (x_1, \ldots, x_n) \in \mathcal{X} \text{ e } F_{\underline{X}} \text{ funzione di ripartizione su } \underline{X}\right\}$$

o di densità congiunte

$$\mathcal{F} = \left\{f_{\underline{X}}(x_1, \ldots, x_n) : (x_1, \ldots, x_n) \in \mathcal{X} \text{ e } f_{\underline{X}} \text{ funzione di densità su } \underline{X}\right\}$$

o descrivendo qualche caratteristica della legge congiunta di $\underline{X}$. Spesso un modello statistico è dato in termini di densità rispetto ad una misura dominante (che nel seguito assumiamo essere la misura di Lebesgue o la misura che conta): $f_{\underline{X}}(\cdot;\theta)$ oppure $\mathrm{P}(\underline{X} = \cdot;\theta)$.
Anticipiamo che spesso 1. lo *spazio dei parametri* di un modello statistico parametrico è (in relazione biunivoca con) un (sotto insieme di uno) spazio vettoriale di dimensione finita e 2. il modello statistico $\mathcal{F}$ è *identificabile* poiché ad ogni $\theta \in \Theta$ è associato un, ed un solo, modello probabilistico in $\mathcal{F}$. Nelle applicazioni sono spesso utili i modelli semi-parametrici ovvero che sono solo parzialmente specificati da parametri.

Esempio 1.5

1. Per $(X_1, \ldots, X_n) \sim \mathcal{N}_n((\underbrace{\mu, \ldots, \mu}_{n}), \sigma^2 I_n)$, se σ^2 è noto, lo spazio dei parametri è $\mathbb{R}$; se μ è noto lo spazio dei parametri è $\mathbb{R}_{>0}$; se entrambi sono non noti allora $\Theta = (\mathbb{R}, \mathbb{R}_{>0})$ e $\theta = (\mu, \sigma^2)$.
2. Per $X \sim \text{Poisson}(\lambda)$, $\lambda \in \Theta = \mathbb{R}_{>0}$.
3. Siano $\underline{X} = (X_1, \ldots, X_n) \in \mathbb{R}^n$ e $||a||$ la norma euclidea del vettore $a \in \mathbb{R}^n$.
   (a) $\mathcal{F} = \{$tutte le leggi di probabilità su $(X_1, \ldots, X_n)\}$ è modello statistico non parametrico per $\underline{X}$.
   (b) $\mathcal{F} = \{$tutte le leggi di probabilità su $(X_1, \ldots, X_n)$ con marginali univariate indipendenti$\}$ è modello statistico non parametrico.
   (c) $\mathcal{F} = \{F_{\underline{X}}(\underline{x};\mu) = F_0(||\underline{x} - \mu||) : F_0$ funzione di ripartizione univariata e $\mu \in \mathbb{R}^n\}$ è modello statistico *semi-parametrico*.
4. Le famiglie di posizione e scala sono modelli semi-parametrici: si ipotizza che $\underline{X}$ è i.i.d e che $X_1 \sim \dfrac{1}{\sigma} f\left(\dfrac{x-\mu}{\sigma}\right)$ e $(\mu, \sigma) \in \mathbb{R} \times \mathbb{R}_{>0}$ con f densità, μ parametro di posizione e σ di scala (n.b. anche $x \in \mathbb{R}$). Un tipico modello di posizione e scala è dato da $\mathcal{N}(\mu, \sigma^2)$, $\mu \in \mathbb{R}$ e $\sigma^2 \in \mathbb{R}^2$ dove $f(x) = \exp(-x^2/2)/\sqrt{2\pi}$.
5. L'insieme delle densità simmetriche e continue su $\mathbb{R}$ è un modello semi-parametrico dove il parametro reale identifica l'asse di simmetria.

Esempio 1.6 $X \sim \mathcal{N}_1(\mu_1 + \mu_2, \sigma^2)$ con $\mu_1, \mu_2 \in \mathbb{R}$ e σ noto non è identificabile.

Esercizio 1.7 (standardizzazione) Verificare che per f densità univariata, $\mu \in \mathbb{R}$ e $\sigma \in \mathbb{R}_{>0}$ allora $g(x) = \frac{1}{\sigma} f(\frac{x-\mu}{\sigma})$ è una densità.

Esercizio 1.8 Siano $X_1, \ldots, X_n$ i.i.d. a valori reali e $g_1, \ldots, g_n$ funzioni da $\mathbb{R}$ a $\mathbb{R}$.

1. $g_1(X_1), \ldots, g_n(X_n)$ sono indipendenti? Sono i.i.d?
2. Sotto quali ipotesi valgono le seguenti uguaglianze? $\mathrm{E}\left(\sum_{i=1}^n (g_i(X_i))\right) = \sum_{i=1}^n (\mathrm{E}(g_i(X_i))) = n\,\mathrm{E}(g_1(X_1))$.
3. Sotto quali ipotesi valgono le seguenti uguaglianze? $\mathrm{Var}\left(\sum_{i=1}^n (g_i(X_i))\right) = \sum_{i=1}^n \mathrm{Var}(g_i(X_i)) = n\,\mathrm{Var}(g_1(X_1))$.

Predica su: dominio e codominio, quantità aleatorie e deterministiche, parametro e costante (parametro di disturbo)

1.1 Identificabilità e stimabilità

Anche se non vi faremo riferimento in seguito è opportuno approfondire il punto 2. della Definizione 1.4 e confrontarla con quella di funzione stimabile del parametro.

Definizione 1.9 *Siano $\underline{X}$ un campione e $\mathcal{F} = \left\{F_{\underline{X}}(\cdot;\theta) : \theta \in \Theta \subseteq \mathbb{R}^p\right\}$ un modello statistico parametrico.*

1. *La funzione $h(\theta)$ del parametro è identificabile se $h(\theta_1) \neq h(\theta_2)$ implica che $F_{\underline{X}}(\cdot;\theta_1) \neq F_{\underline{X}}(\cdot;\theta_2)$ per ogni $\theta_1, \theta_2 \in \Theta$.*

2. *La funzione $h(\theta)$ del parametro è stimabile se esiste una funzione g del solo campione $\underline{X}$ tale che $\mathrm{E}_\theta\left(g(\underline{X})\right) = h(\theta)$ per ogni $\theta \in \Theta$.*

In altre parole, h è identificabile se, al variare dei valori che assume, anche la distribuzione del vettore aleatorio osservabile $\underline{X}$ cambia e, anticipando della nomenclatura, è stimabile se esiste un suo stimatore corretto.

Proposizione 1.10 *Se h è stimabile allora è identificabile.*

Proof. Se h è stimabile allora esiste g tale che $\mathrm{E}_\theta\left(g(\underline{X})\right) = h(\theta)$ per ogni $\theta \in \Theta$. Se $h(\theta_1) \neq h(\theta_2)$ allora $\mathrm{E}_{\theta_1}\left(g(\underline{X})\right) \neq \mathrm{E}_{\theta_2}\left(g(\underline{X})\right)$. Si deduce che le due leggi probabilistiche rispetto le quali si sono calcolati i due valori attesi sono differenti. ∎

Esempio 1.11 Sia $p \in]0,1[$ e $X \sim \text{Bernoulli}(p)$ con $X \in \{0,1\}$ e $\mathrm{P}(X=1) = p$. La funzione $h(p) = \sqrt{p}$ è identificabile, ma non stimabile: infatti dovrebbe esistere una funzione g tale che $\mathrm{E}_\theta\left(g(\underline{X})\right) = (1-p)g(0) + pg(1) = \sqrt{p}$, che è impossibile.

Esercizio 1.12 Siano $X \sim \text{Bernoulli}(p)$ con $X \in \{0,1\}$ e $Y \sim \text{Bernoulli}(p)$ con $Y \in \{-1,1\}$ e $\mathrm{P}(X=1) = p = \mathrm{P}(Y=1)$. Calcolare $\mathrm{E}(X)$ e $\mathrm{Var}(X)$, $\mathrm{E}(Y)$ e $\mathrm{Var}(Y)$.

1.2 Modelli regolari

Definizione 1.13 *Sia $\underline{X}$ un vettore aleatorio con densità congiunta $f_{\underline{X}}(\cdot)$. Il* supporto *di $f_{\underline{X}}$, $\mathrm{supp}(f_{\underline{X}})$, è (la chiusura nella topologia euclidea di) $\{\underline{x} : f_{\underline{X}}(\underline{x}) > 0\}$. Un modello statistico parametrico è regolare se il supporto di ogni suo modello probabilistico non dipende dal parametro.*

Esempio 1.14 Per $n = 1$ e $X \sim \mathcal{N}(\mu, \sigma^2)$, il supporto è $\mathbb{R} \times \mathbb{R}_{>0}$ per ogni (μ, σ^2) ed il modello è regolare. Per $\theta \in \mathbb{R}_{>0}$ e $f_{\underline{X}} = \text{Uniforme}(]0,\theta[)$, il supporto è $[0,\theta]$. Quindi Uniforme$(]0,\theta[)$, $\theta \in \mathbb{R}_{>0}$, è un modello statistico non regolare per X.

Esercizio 1.15 Verificare che Poisson(λ), $\lambda > 0$, è un modello regolare.

Esempio 1.16 Un modello statistico $\{f_{\underline{X}} :$ densità di probabilità$\}$ per un campione $\underline{X} \in \mathcal{X}$ è detto *di classe esponenziale* se

$$f_{\underline{X}}(\underline{x}) = h(\underline{x}) \exp\left(\langle \theta, T(\underline{x})\rangle - \psi(\theta)\right) = h(\underline{x}) \exp\left(\sum_{i=1}^{p} \theta_i T_i(\underline{x}) - \psi(\theta)\right)$$

con $h : \mathcal{X} \to \mathbb{R}_{>0}$, $\theta \in \Theta \subset \mathbb{R}^p$, $T : \mathcal{X} \to \Theta$ e $\langle \cdot, \cdot \rangle$ prodotto scalare su Θ. La funzione ψ è detta *funzione dei cumulanti* e θ *parametro naturale*. I modelli di classe esponenziale sono regolari. Quasi sempre, ma non necessariamente, avremo $\Theta = \mathbb{R}^p$. Per una generalizzazione della definizione si veda per esempio Appendix 8 in Lauritzen, 1996.
Altre importanti (classi di) modelli statistici sono i modelli di regressione e i modelli grafici di indipendenza.

Esercizio 1.17 Siano $X_1, \ldots, X_n$ i.i.d. di legge $h \exp\left(\langle T, \theta\rangle - \psi(\theta)\right)$.

1. Verificare che $\int_{\mathcal{X}} h(\underline{x}) \exp\left(\langle T(\underline{x}), \theta\rangle\right) d\underline{x} = \exp\left(\psi(\theta)\right)$.

2. Calcolare la legge congiunta del campione. Qual è lo spazio campionario dell'n-campione?

3. La legge congiunta del campione è un modello di classe esponenziale?

4. La famiglia di leggi probabilità $\mathcal{N}(\mu, \sigma^2)$ al variare di μ e di σ^2 è un modello di classe esponenziale per una variabile aleatoria a valori reali?

Esercizio 1.18 (Mistura di Gaussiane) Sia $\pi \in]0,1[$ e per $i = 1,2$ sia $f_i(\cdot, \mu_i, \sigma_i^2)$ una densità normale univariata di media μ_i e varianza σ_i^2. Si consideri per una variabile aleatoria $X \in \mathbb{R}$ il modello statistico a cinque parametri dato da

$$f_X(x) = \pi f_1(\cdot, \mu_1, \sigma_1^2) + (1-\pi) f_2(\cdot, \mu_2, \sigma_2^2) \qquad x \in \mathbb{R}$$

Verificare che il modello è ben definito specificando gli spazi campionario e parametrico. Si verifichi che è regolare e non è di classe esponenziale. Generalizzare il modello a un campione causale di numerosità n. Nel codice R in Appendice 7.1 sono implementati diversi metodi per campionare da un modello di mistura.

Studieremo tre principi che sono alla base dell'inferenza statistica parametrica: 1) sufficienza: riassumere i dati senza perdere informazioni sul parametro, 2) verosimiglianza: identificare tutte le informazioni sul parametro contenute nei dati, 3) invarianza: trasformare lo spazio dei parametri e lo spazio campionario in sincronia/contemporaneamente.

2 Statistiche

Definizione 2.1 *Si dice* statistica *una funzione del (solo) campione* $\underline{X}$ *e* distribuzione campionaria *la sua legge di probabilità.*

Sia il vettore aleatorio $T(\underline{X})$ sia la sua realizzazione sono chiamati statistica. Una statistica è una funzione dei dati che fornisce una sintesi di interesse per lo studio di un particolare aspetto dei dati e/o del sistema che li genera. Una statistica non dipende dal modello statistico (e quindi dal parametro nel caso di modello statistico parametrico), ma solo dal campione. In particolare se il modello statistico è parametrico, una statistica non può dipendere dal parametro. E' un vettore aleatorio, spesso unidimensionale, la cui legge di probabilità può essere talora determinata dalla legge del campione. La distribuzione campionaria di una statistica (spesso) dipende dal modello statistico. Gli stimatori puntuali, che vedremo più avanti, sono statistiche utilizzate per stimare il parametro θ, o una sua funzione, di un modello statistico parametrico.
Una statistica T induce una partizione dello spazio campionario Ω: per esempio, se $\mathcal{X} = \mathbb{R}^n$ e $T \in \mathbb{R}$, la mappa $\Omega \overbrace{\longrightarrow}^{\underline{X}} \mathbb{R}^n \overbrace{\longrightarrow}^{T} \mathbb{R}$ definisce la partizione i cui elementi sono $\{\omega \in \Omega : T(\underline{X}) = t\}$ al variare di $t \in \mathbb{R}$. Più in generale se $\underline{X} \in \mathcal{X}$ e $T \in \mathcal{T}$, si ha $\Omega \overbrace{\longrightarrow}^{\underline{X}} \mathcal{X} \overbrace{\longrightarrow}^{T} \mathcal{T}$.

Esempio 2.2 Siano $X_1, \ldots, X_n$ variabili aleatorie a valori in $\mathbb{R}$. I seguenti sono esempi di statistiche.

Totale campionario: $T = \sum_{i=1}^n X_i$

Media campionaria: $\bar{X} = T/n$

Varianza campionaria: $S^2 = \sum_{i=1}^n (X_i - \bar{X})^2/(n-1)$ e $S^2(n-1)/n$

Massimo campionario: $X_{(n)} = \max\{X_1, \ldots, X_n\}$

Minimo campionario: $X_{(1)} = \min\{X_1, \ldots, X_n\}$

Moda campionaria: valore più frequente

Statistica d'ordine del campione (se si può ordinare): $\text{sort}\{X_1, \ldots, X_n\} = (X_{(1)}, X_{(2)}, \ldots, X_{(n)})$. La mediana è una statistica d'ordine[2]

[2]Per un campione i.i.d. di taglia n con densità univariata f_X e funzione di ripartizione F_X e per $j = 1, \ldots, n$ vale

$$F_{X_{(j)}}(x) = \sum_{k=j}^{n} F_X(x)^k \,(1 - F_X(x))^{n-k} \qquad \text{and} \qquad f_{X_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!} f_X(x) F_X(x)^{j-1}\,(1-F_X(x))^{n-j}$$

Il campione $\underline{X} = (X_1, \dots, X_n)$ è una statistica

In un modello di classe esponenziale le T_i, $i = 1, \dots, p$, sono statistiche

Esempio 2.3 1. Sia $X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ i.i.d. con $\mu \in \mathbb{R}$ e $\sigma \in \mathbb{R}_{>0}$ allora $R = \bar{X} - \mu$ non è una statistica. Se μ è noto allora R è una statistica.

2. Sia $X_1, \dots, X_n \sim \text{Unif}(]0, \theta[)$ con $\theta > 0$ allora $R = X_{(n)}/\theta$ non è una statistica. Ma è molto utile. Cosa rappresenta?

Esercizio 2.4 Verificare che per le seguenti statistiche e modelli statistici le distribuzioni campionarie sono quelle indicate.

1. Se $X_1, \dots, X_n$ campione Bernoulliano di parametro p i.i.d, allora $T = n\bar{X} \sim \text{Bin}(n, p)$
2. Se $X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ i.i.d. allora $\bar{X} \sim \mathcal{N}(\mu, \sigma^2/n)$
3. Sia $X_1, \dots, X_n$ un campione a valori reali, i.i.d e con funzione di ripartizione F. La distribuzione del massimo campionario è F^n. Determinare la legge del minimo campionario. Calcolare anche le funzioni di densità del massimo e del minimo campionario.

Esempio 2.5 1. Siano $(\Omega, \mathcal{A}, \mathrm{P})$, $A \in \mathcal{A}$ e

$$1_A(\omega) = (A)(\omega) = \begin{cases} 1 & \text{se } \omega \in A \\ 0 & \text{se } \omega \notin A \end{cases}$$

Allora $(A) \sim \text{Bernoulli}(\mathrm{P}(A))$. Si noti che $\mathrm{E}((A)) = \mathrm{P}(A)$.

2. Sia $\underline{X}$ un vettore aleatorio a valori in $\mathbb{R}^n$. Sia $A \subset \mathbb{R}^n$ un insieme Borel misurabile. Allora

$$(\underline{X} \in A)(\omega) = 1_{(\underline{X} \in A)}(\omega) = \begin{cases} 1 & \text{se } \underline{X}(\omega) \in A \\ 0 & \text{se } \underline{X}(\omega) \notin A \end{cases}$$

e $(\underline{X} \in A) \sim \text{Bernoulli}(\mathrm{P}(\underline{X} \in A))$.

Esempio 2.6 Siano $X_1, \dots, X_n \in \mathbb{R}$ i.i.d. e con funzione di ripartizione F_X. Sia $x \in \mathbb{R}$. La statistica $C(x) = \sum_{i=1}^n (X_i \leq x)$ conta quanti dati sono minori od uguali a x. Assume valori in $\{0, \dots, n\}$ e per $j \in \{0, \dots, n\}$ si ha

$$\mathrm{P}(C(x) \leq j) = \mathrm{P}\left(\sum_{i=1}^n (X_i \leq x) \leq j\right) = \text{ la probabilità che ci siano al più } j \text{ dati non maggiori di } x$$

La statistica $C(x)$ è la somma di n Bernoulli di parametro $F_X(x)$ indipendenti. Quindi la distribuzione campionaria di $C(x)$ è Binomiale$(n, F_X(x))$.

Esercizio 2.7 Siano $X_1, \dots, X_n \sim \text{Cauchy}(0, \sigma)$ per $\sigma > 0$ cioè $f_{X_1}(x) = \frac{1}{\pi\sigma} \frac{1}{1+(x/\sigma)^2}$, $x \in \mathbb{R}$.

1. Verificare che $f_{X_1}(x)$ è densità di probabilità e che $\mathrm{E}(X_1)$ non esiste.
2. E' un modello di posizione e/o scala?
3. Si verifichi che non è un modello di classe esponenziale.
4. Verificare che la funzione di ripartizione è $F_{X_1}(x) = \dfrac{1}{\pi} \arctan\left(\dfrac{x}{\sigma}\right) + \dfrac{1}{2}$ e la funzione caratteristica è $\Phi_{X_1}(t) = \exp(-\sigma|t|)$.
5. Nell'ipotesi che il campione sia i.i.d., verificare che $\frac{\sum_{i=1}^n X_i}{n} \sim \text{Cauchy}(0, \sigma)$. Infatti la funzione caratteristica di X_1 è $\Phi_{X_1}(t) = \exp(-\sigma|t|))$ e per a_i numeri reali e X_i variabili aleatorie indipendenti $(i = 1, \dots, n)$ vale $\Phi_{\sum_{i=1}^n a_i X_i}(t) = \prod_{i=1}^n \Phi_{X_i}(a_i t)$.

Esercizio 2.8 Siano $X_1, \dots, X_N$ i.i.d di legge $\dfrac{1}{\sigma} f\left(\dfrac{x-\mu}{\sigma}\right)$ con $\mu \in \mathbb{R}$ e $\sigma > 0$.

1. Individuare Z_i per cui $X_i = \sigma Z_i + \mu$ e determinarne la densità.
2. Sono gli $Z_1, \dots, Z_n$ i.i.d.?
3. Verificare che $\bar{X} = \sigma \bar{Z} + \mu$ e $\operatorname{Var}(\bar{X}) = \sigma^2 \operatorname{Var}(\bar{Z})$.

Esercizio 2.9 (Il modello di Cauchy di posizione non è di classe esponenziale) Siano $X_1, \dots, X_n \sim$ Cauchy$(\mu, 1)$ per $\mu \in \mathbb{R}$ e con leggi marginali univariate $f_X(x) = \frac{1}{\pi} \frac{1}{1+(x-\mu)^2}$, $x \in \mathbb{R}$.
E' un modello di posizione infatti $X_1 = Z + \mu$ con $Z \sim$ Cauchy$(0,1)$. Per assurdo supponiamo che sia un modello di classe esponenziale ovvero che esistono T statistica e θ e ψ funzioni di μ tali che $\log f_X(x) + \log \pi = -\log(1 + (x-\mu)^2) = T(x)\theta - \psi$ (*). Per $\mu = 0$ si trova $T(x) = \dfrac{\psi(\theta(0)) - \log(1+x^2)}{\theta(0)}$. Sostituendo in (*) si trova $\theta(\mu) = \theta(0) \dfrac{\psi(\mu) - \log(1+(x-\mu)^2)}{\psi(0) - \log(1+x^2)}$. Questa è una contraddizione perché θ è costante in x. Dimostrare che Cauchy$(0, \sigma)$ non è di classe esponenziale.

2.1 Statistiche sufficienti

Definizione 2.10 (Teorema della fattorizzazione) *Sia $\underline{X}$ un campione statistico su $(\Omega, \mathcal{A}, \mathrm{P})$ a valori in $\mathcal{X}$ e sia $\mathcal{F} = \left\{ f_{\underline{X}}(\cdot; \theta) : \theta \in \Theta \text{ con } f \text{ densità di probabilità} \right\}$ un modello statistico parametrico per $\underline{X}$. Una statistica $T = T(\underline{X})$ è detta* sufficiente *per θ se esistono due funzioni h e g tali che*

$$f_{\underline{X}}(\underline{x}; \theta) = h(\underline{x}) g(T(\underline{x}), \theta)$$

per ogni $\theta \in \Theta$ e per (quasi) ogni $\underline{x} \in \mathcal{X}$.

Chiaramente se $\mathcal{X}$ è un insieme numerabile allora $f_{\underline{X}}(\underline{x}; \theta) = \mathrm{P}(\underline{X} = \underline{x}; \theta)$. Scriveremo anche $\mathrm{P}_\theta(\underline{X} = \underline{x})$. La definizione suggerisce che una statistica è sufficiente se cattura tutte le informazioni sul parametro contenute nel campione. Più precisamente, se due insiemi di dati producono lo stesso valore della statistica sufficiente, allora con ragionamenti basati sulla legge congiunta del campione (detta anche funzione di verosimiglianza quando letta come funzione del parametro) dai due campioni si otterranno identiche deduzioni inferenziali sul parametro. Questa interpretazione è ribadita dal seguente teorema.

Theorem 2.11 (di Neyman-Fisher) *Una statistica T è sufficiente per θ se e solo se la distribuzione condizionata di $\underline{X}$ a $T(\underline{X}) = t$ non dipende da θ per ogni valore della statistica T.*

Proof. La dimostrazione nel caso continuo è analoga a quella del caso discreto. Nell'ipotesi che T sia sufficiente per θ, si noti che per ogni t valore di T vale

$$\begin{aligned} f_T(t) = \mathrm{P}_\theta(T = t) &= \sum_{\underline{x}: T(\underline{x}) = t} \mathrm{P}_\theta(\underline{X} = \underline{x}, T(\underline{X}) = t) = \sum_{\underline{x}: T(\underline{x}) = t} \mathrm{P}_\theta(\underline{X} = \underline{x}) \\ &= \sum_{\underline{x}: T(\underline{x}) = t} h(\underline{x}) g(T(\underline{x}), \theta) && \text{perché } T \text{ è sufficiente per } \theta \\ &= g(t, \theta) \sum_{\underline{x}: T(\underline{x}) = t} h(\underline{x}) = g(t, \theta) h^*(\underline{x}) && \text{con } h^*(\underline{x}) = \sum_{\underline{x}: T(\underline{x}) = t} h(\underline{x}) \end{aligned}$$

In particolare la legge di T dipende dal parametro solo tramite la funzione g. Ora si osservi che

$$f_{\underline{X}|T(\underline{X})=t}(\underline{x}, \theta) = \begin{cases} \dfrac{\mathrm{P}_\theta(\underline{X} = \underline{x}, T = t)}{\mathrm{P}_\theta(T = t)} = \dfrac{\mathrm{P}_\theta(\underline{X} = \underline{x})}{\mathrm{P}_\theta(T = t)} = \dfrac{g(t, \theta) h(\underline{x})}{g(t, \theta) h^*(\underline{x})} = \dfrac{h(\underline{x})}{h^*(\underline{x})} & \text{se } T(\underline{x}) = t \\ \text{non definita} & \text{altrimenti} \end{cases}$$

Viceversa la tesi segue dall'osservare che $f_{\underline{X}} = f_{(\underline{X}, T)} = f_{\underline{X}|T} f_T$. ■
Si veda l'Esempio 4.10.

Esempio 2.12 Siano $X_1, \ldots, X_n$ i.i.d. Bernoulli(p) con $p \in \Theta =]0,1[$. Verificare che $T = \sum_{i=1}^n X_i$ è statistica sufficiente per p. Utilizzando la definizione calcoliamo la legge congiunta del campione e cerchiamo di determinarne una fattorizzazione in due fattori di cui uno dipende dal parametro e l'altro dipende dal campione solo tramite T

$$\begin{aligned}
\mathrm{P}_p(\underline{X} = \underline{x}) &= \mathrm{P}_p(X_1 = x_1, \ldots, X_n = x_n) = \prod_{i=1}^n \mathrm{P}_p(X_1 = x_i) \\
&= \prod_{i=1}^n p^{x_i}(1-p)^{1-x_i} = p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i} = p^t(1-p)^{n-t}
\end{aligned}$$

da cui $h(\underline{x}) = 1$ e $g(t,p) = p^t(1-p)^{n-t}$. Facciamo la verifica utilizzando il teorema. Occorre considerare la legge condizionata del campione al valore assunto dalla statistica sufficiente quindi

$$\mathrm{P}_p(\underline{X} = \underline{x} | T = t) = \begin{cases} 0 & \text{se } \sum_{i=1}^n x_i \neq t \\ \dfrac{\mathrm{P}_p(\underline{X} = \underline{x}, T = t)}{\mathrm{P}_p(T = t)} = \dfrac{p^t(1-p)^{n-t}}{\binom{n}{t}p^t(1-p)^{n-t}} = \dfrac{1}{\binom{n}{t}} & \text{se } \sum_{i=1}^n x_i = t \end{cases}$$

Esempio 2.13

1. La statistica T nella definizione di modello statistico di classe esponenziale è sufficiente per il parametro naturale.
2. Il campione è statistica sufficiente per il parametro di ogni modello statistico parametrico sul campione stesso. (Il campione è statistica sufficiente).

Esempio 2.14 Sia $X_1, \ldots, X_n$ i.i.d. $\sim$ Unif$(\{1, 2, \ldots, \theta\})$ con $\theta \in \mathbb{Z}_{\geq 1}$. La legge congiunta del campione è

$$\begin{aligned}
f_{\underline{X}}(\underline{x}; \theta) &= \prod_{i=1}^n \frac{1}{\theta}\,(x_i \in \{1, \ldots, \theta\}) \\
&= \frac{1}{\theta^n} \prod_{i=1}^n (x_i \in \{1, \ldots, \theta\}) \\
&= \frac{1}{\theta^n}(x_{(1)} \geq 1)(x_{(n)} \leq \theta) \prod_{i=1}^n (x_i \in \mathbb{Z}_{\geq 1})
\end{aligned}$$

Allora $T = X_{(n)}$ è sufficiente per θ infatti $g(t,\theta) = \frac{(x_{(n)} \leq \theta)}{\theta^n}$ e $h(\underline{x}) = (x_{(1)} \geq 1) \prod_{i=1}^n (x_i \in \mathbb{Z}_{\geq 1})$.

Esempio 2.15 Sia $X_1, \ldots, X_n$ i.i.d. $\sim \mathcal{N}_1(\mu, \sigma^2)$ con $(\mu, \sigma^2) \in (\mathbb{R}, \mathbb{R}_{>0})$. Determiniamo statistiche sufficienti per (μ, σ^2), μ e σ^2.

1. La legge congiunta del campione è

$$f_{\underline{X}}(\underline{x}) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2}\right) = \frac{1}{\sqrt{2\pi\sigma^2}^n} \exp\left(-\sum_{i=1}^n \frac{(x_i - \mu)^2}{2\sigma^2}\right)$$

Si noti che

$$\sum_{i=1}^n (x_i - \mu)^2 = \sum_{i=1}^n (x_i - \bar{x} + \bar{x} - \mu)^2 = \sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2 + 0$$

e quindi

$$f_{\underline{X}}(\underline{x}) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{n-1}{2\sigma^2}s^2 - \frac{n}{2\sigma^2}(\bar{x} - \mu)^2\right)$$

Ponendo $h \equiv 1$ e $g = f_{\underline{X}}$ si evidenzia che $(\bar{X}, S^2)$ è una statistica sufficiente per $\theta = (\mu, \sigma^2)$. Per $g = \exp\left(-\frac{n}{2\sigma^2}(\bar{x} - \mu)^2\right)$ si ha che $\bar{X}$ è una statistica sufficiente per μ noto σ^2. Ponendo $h \equiv 1$ e $g = f_{\underline{X}}$ si deduce che $(\bar{X}, S^2)$ è sufficiente per σ^2.

2. Poiché $\bar{X} \sim \mathcal{N}_1(\mu, \sigma^2/n)$ si ha

$$
\begin{aligned}
f_{\underline{X}|\bar{X}}(\underline{x}) = \frac{f_{\underline{X}}(\underline{x})}{f_{\bar{X}}(\bar{x})} &= \frac{(2\pi\sigma^2)^{-n/2}\exp\left(-\dfrac{n-1}{2\sigma^2}s^2 - \dfrac{n}{2\sigma^2}(\bar{x}-\mu)^2\right)}{(2\pi\sigma^2/n)^{-1/2}\exp\left(-\dfrac{(\bar{x}-\mu)^2}{2\sigma^2/n}\right)} \\
&= \frac{1}{\sqrt{n(2\pi\sigma^2)^{n-1}}}\exp\left(-\sum_{i=1}^{n}(x_i-\bar{x})^2/(2\sigma^2)\right)
\end{aligned}
$$

Dal teorema di Neyman-Fisher si deduce che $\bar{X}$ è sufficiente per μ.

Esercizio 2.16 1. Dimostrare che una funzione invertibile di una statistica sufficiente per un parametro θ è ancora statistica sufficiente per θ. In particolare le statistiche sufficienti non sono uniche.

2. Sia $X_1, \ldots, X_n$ i.i.d. $\mathrm{Exp}(\lambda)$, $\lambda > 0$. Verificare che la media campionaria è una statistica sufficiente per λ.

3. Sia $X_1, \ldots, X_n$ i.i.d. $\mathrm{Uniform}(]0, \theta[)$, $\theta > 0$. Determinare una statistica sufficiente univariata per θ.

Esempio 2.17 Sia $\Theta = \{0, 1\}$, $\theta_0 = 0$ e $\theta_1 = 1$. Il rapporto di verosimiglianza è statistica sufficiente per θ, infatti

$$
f_{\underline{X}}(\underline{x}; \theta) = \left(\frac{f_{\underline{X}}(\underline{x}; \theta_1)}{f_{\underline{X}}(\underline{x}; \theta_0)}\right)^{\theta} f_{\underline{X}}(\underline{x}; \theta_0)
$$

2.2 Statistiche sufficienti minimali

Definizione 2.18 *Una statistica sufficiente per θ è* sufficiente minimale *per θ se è sufficiente ed è funzione di ogni altra statistica sufficiente per θ.*[3]

Esempio 2.19 Per $X_1, \ldots, X_n$ campione casuale $\mathcal{N}(\mu, \sigma^2)$, $\bar{X} = \sum_{i=1}^n X_i/n$ e $\underline{X} = (X_1, \ldots, X_n)$ sono sufficienti per μ. Poiché $\bar{X}$ è funzione di $\underline{X}$, $\underline{X}$ non è sufficiente minimale.

Per definizione, se T' è una statistica sufficiente e T è sufficiente minimale allora esiste una funzione h tale che $T = h(T')$ ed in particolare se $T'(\underline{x}) = T'(\underline{y})$ allora $T(\underline{x}) = T(\underline{y})$. Una statistica sufficiente minimale può essere calcolata (se nota h) a partire da ogni altra statistica sufficiente. Si intende che i dati non possono essere "compressi" ulteriormente senza perdere informazioni su θ. L'Esercizio 2.27 illustra un caso in cui due statistiche sufficienti e minimali hanno dimensione diversa. La nota successiva precisa quanto sopra.

Osservazioni 2.20 1. La partizione dello spazio campionario $\mathcal{X}$ indotta dalla statistica sufficiente T è tale per cui la legge congiunta del campione $f_{\underline{X}}$ può essere ricostruita (a meno di una costante moltiplicativa funzione del solo campione) conoscendo il valore di $f_{\underline{X}}$ in un punto di ogni elemento della partizione. La partizione (sia su Ω che su $\mathcal{X}$) indotta da una statistica sufficiente minimale è la meno fine tra quelle indotte da statistiche sufficienti.

2. Se ne deduce che diverse statistiche sufficienti minimali inducono la stessa partizione su $\mathcal{X}$ e che una funzione biunivoca di statistica sufficiente minimale è sufficiente minimale.

3. Una statistica sufficiente non minimale contiene informazioni "superflue" sul parametro per ricostruire la legge congiunta del campione. Fornisce la minima informazione necessaria per ricostruire il modello (θ) a partire dai dati.

4. Il seguente teorema dà una condizione sufficiente affinchè una statistica sia sufficiente minimale.

[3] $\Omega \xrightarrow{\underline{X}} \mathcal{X} \xrightarrow{T} \mathcal{T} \xrightarrow{T_M} \mathcal{T}_M$.

Theorem 2.21 *Se per ogni $\underline{x}, \underline{y} \in \mathcal{X}$ il rapporto $\dfrac{f_{\underline{X}}(\underline{x};\theta)}{f_{\underline{X}}(\underline{y};\theta)}$, detto rapporto di verosimiglianza,*[4] *è costante in θ se e solo se $T(\underline{x}) = T(\underline{y})$ allora T è sufficiente e minimale per θ.*

Proof. Per semplicità supponiamo che $f_{\underline{X}}(\underline{x};\theta) > 0$ per ogni $\underline{x}, \theta$.
Deduciamo prima che T è sufficiente per θ. Per ogni valore t assunto da T sia $\mathcal{A}_t = \{\underline{x} \in \mathcal{X} : T(\underline{x}) = t\}$. In particolare per ogni $\underline{x}, \underline{y} \in \mathcal{A}_t$ si ha $T(\underline{x}) = T(\underline{y})$ e per ipotesi $\dfrac{f_{\underline{X}}(\underline{x};\theta)}{f_{\underline{X}}(\underline{y};\theta)}$ non dipende da θ. Ora per $\underline{x} \in \mathcal{X}$ si consideri $\underline{y} \in \mathcal{A}_{T(\underline{x})}$ e la funzione $h(\underline{x}) = \dfrac{f_{\underline{X}}(\underline{x};\theta)}{f_{\underline{X}}(\underline{y}(\underline{x});\theta)}$ che non dipende da θ per costruzione. Si scriva

$$f_{\underline{X}}(\underline{x};\theta) = f_{\underline{X}}(\underline{x};\theta)\frac{f_{\underline{X}}(\underline{y}(\underline{x});\theta)}{f_{\underline{X}}(\underline{y}(\underline{x});\theta)} = \frac{f_{\underline{X}}(\underline{x};\theta)}{f_{\underline{X}}(\underline{y};\theta)} f_{\underline{X}}(y;\theta) = h(\underline{x})g(T(\underline{y});\theta)$$

dove l'ultima uguaglianza segue dall'arbitrarietà di $\underline{y}$ in $\mathcal{A}_t$.
Ora dimostriamo la minimalità di T. Sia T' un'altra statistica sufficiente per θ. In particolare esistono h' e g' tali che

$$f_{\underline{X}}(\underline{x};\theta) = h'(\underline{x})g'(T'(\underline{x});\theta)$$

Per $\underline{x}, \underline{y} \in \mathcal{X}$ e poiché T' è sufficiente si ha

$$\frac{f_{\underline{X}}(\underline{x};\theta)}{f_{\underline{X}}(\underline{y};\theta)} = \frac{h'(\underline{x})g'(T'(\underline{x});\theta)}{h'(\underline{y})g'(T'(\underline{y});\theta)} = \frac{h'(\underline{x})}{h'(\underline{y})}$$

per ogni $\underline{x}, \underline{y}$ tali che $T'(\underline{x}) = T'(\underline{y})$. Questo rapporto evidentemente non dipende da θ. Per ipotesi questo avviene se e solo se $T(\underline{x}) = T(\underline{y})$, da cui T è funzione di T'. ■

Esempio 2.22 Sia $X_1, \ldots, X_n$ un campione casuale $\mathcal{N}(\mu, \sigma^2)$ con $(\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_{>0}$. Verificare che $T = (\bar{X}, S^2)$ è sufficiente minimale per (μ, σ^2). Occorre caratterizzare il luogo dei punti $\underline{y}$ per cui

$$\frac{f_{\underline{X}}(\underline{x};\mu,\sigma^2)}{f_{\underline{X}}(\underline{y};\mu,\sigma^2)} = \frac{(2\pi\sigma^2)^{-n/2}}{(2\pi\sigma^2)^{-n/2}} \frac{\exp\left(-\left(n(\bar{x}-\mu)^2 + (n-1)s_{\underline{x}}^2/n\right)/2\sigma^2\right)}{\exp\left(-\left(n(\bar{y}-\mu)^2 + (n-1)s_{\underline{y}}^2/n\right)/2\sigma^2\right)}$$

dove $(n-1)s_{\underline{x}}^2 = \sum_{i=1}^{2}(x_i - \bar{x})^2$. Si ha

$$\begin{aligned}\frac{f_{\underline{X}}(\underline{x};\mu,\sigma^2)}{f_{\underline{X}}(\underline{y};\mu,\sigma^2)} &= \exp\left(\frac{-n}{2\sigma^2}\left(\bar{x}^2 - \bar{y}^2 - 2\mu(\bar{x}-\bar{y}) + (n-1)(s_{\underline{x}}^2 - s_{\underline{y}}^2)/n\right)\right) \\ &= \exp\left(\frac{-n}{2\sigma^2}\left((\bar{x}-\bar{y})(\bar{x}+\bar{y}-2\mu) + (n-1)(s_{\underline{x}}^2 - s_{\underline{y}}^2)/n\right)\right)\end{aligned}$$

da cui si deduce che il rapporto di verosimiglianza non dipende da σ^2 e da μ se e solo se l'argomento dell'esponenziale è zero cioè se e solo se

$$(\bar{x}-\bar{y})(\bar{x}+\bar{y}-2\mu) + (n-1)(s_{\underline{x}}^2 - s_{\underline{y}}^2) = 0$$

se e solo se $\bar{x} = \bar{y}$ e $s_{\underline{x}}^2 = s_{\underline{y}}^2$, equivalentemente se e solo se $T_{\underline{x}} = T_{\underline{y}}$.

Esercizio 2.23

1. Si osservi che dai calcoli nell'Esempio 2.22 segue che $T = (\bar{X}, S^2)$ è anche sufficiente minimale per $\sigma^2 > 0$ nell'ipotesi di modello casuale $\mathcal{N}(\sigma, \sigma^2)$.

2. Per l'n-campione casuale di legge di probabilità $\mathcal{N}(\sigma, \sigma^2)$ con $\sigma > 0$, una statistica sufficiente minimale per il parametro univariato θ è la statistica bidimensionale $(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2)$.

3. Si verifichi che il modello $\mathcal{N}(\sigma, \sigma^2)$, $\sigma^2 > 0$, non è di classe esponenziale.

[4] Confronta lo stesso modello probabilistico in diversi valori campionari/"dati". Confrontare con il rapporto di verosimiglianza nell'Esempio 2.17.

Esempio 2.24 Sia $X_1, \ldots, X_n$ un campione casuale Unif$(]\theta, \theta+1[)$ con $\theta \in \mathbb{R}$. Determinare una statistica sufficiente minimale per θ. La legge congiunta del campione è

$$f_{\underline{X}}(x_1, \ldots, x_n; \theta) = \prod_{i=1}^{n} (\theta < x_i < \theta + 1) = \left(\theta < \min_i x_i\right)\left(\max_i x_i < \theta + 1\right)$$
$$= \left(\theta < x_{(1)}\right)\left(x_{(n)} - 1 < \theta\right) = \left(x_{(n)} - 1 < \theta < x_{(1)}\right)$$

Si noti che la statistica Rango$(\underline{X}) = X_{(n)} - X_{(1)}$ appartiene a $]0,1[$ con probabilità uno. Ora siano $\underline{x}, \underline{y}$ possibili valori campionari

$$\frac{f_{\underline{X}}(\underline{x};\theta)}{f_{\underline{X}}(\underline{y};\theta)} = \frac{\left(x_{(n)} - 1 < \theta < x_{(1)}\right)}{\left(y_{(n)} - 1 < \theta < y_{(1)}\right)} = \begin{cases} 1 & \text{se } x_{(n)} = y_{(n)} \text{ e } x_{(1)} = y_{(1)} \\ \text{dipende da } \theta & \text{altrimenti} \end{cases}$$

Per accertarsi dell'ultima eguaglianza disegnare il grafico del rapporto di verosimiglianza. Per il Teorema 2.21 $T(\underline{X}) = (X_{(1)}, X_{(n)})$ è statistica suffiente minimale per θ.

Esempio 2.25 In un modello di classe esponenziale $f(\underline{x};\theta) = h(\underline{x}) \exp\left(\langle T(\underline{x}), \theta\rangle - \psi(\theta)\right)$ le statistiche T sono sufficienti minimali per il parametro naturale. Infatti

$$\frac{f_{\underline{X}}(\underline{x};\theta)}{f_{\underline{X}}(\underline{y};\theta)} = \frac{h(\underline{x})}{h(\underline{y})} \exp\left(\langle T(\underline{x}) - T(\underline{y}), \theta\rangle\right)$$

non dipende da θ se e solo se $T(\underline{x}) = T(\underline{y})$.

Esercizio 2.26 Determinare una statistica sufficiente minimale per θ in campione casuale uniforme su $\{1, \ldots, \theta\} \subset \mathbb{Z}$ con $\theta \in \mathbb{Z}_{\geq 1}$

Esercizio 2.27 Siano X_1, X_2 i.i.d. Bernoulli(p) con $p \in]0,1[$.

1. Descrivere lo spazio campionario $\mathcal{X}$. $\{(i,j) : i,j = 0,1\}$
2. Verificare che il totale campionario è statistica sufficiente minimale per p ed individuare la partizione di $\mathcal{X}$ corrispondente. $A_{t=0} = \{(0,0)\}$, $A_{t=1} = \{(0,1),(1,0)\}$, $A_{t=2} = \{(1,1)\}$.
3. Verificare che $T = \left(X_{(1)}, X_{(2)}\right)$ è sufficiente minimale per p. Stessa partizione del totale campionario.
4. Come risolviamo il fatto che T bidimensionale e totale campionario unidimensionale sono entrambi sufficienti minimali per p? Dedurre che statistiche sufficienti minimali non hanno necessariamente la stessa dimensione.
5. Verificare che la statistica $R = X_1 X_2$ non è sufficiente minimale per p.

2.3 Statistiche ancillari

Definizione 2.28 *Sia $\underline{X}$ un campione su $(\Omega, \mathcal{A}, \mathrm{P})$ e $\mathcal{F}$ un modello statistico su $\underline{X}$ indicizzato dal parametro $\theta \in \Theta$. Una statistica $T(\underline{X})$ è detta* ancillare *per θ se la distribuzione campionaria di T non dipende da θ.*

Esempio 2.29 Per un campione casuale normale $\mathcal{N}(\mu, \sigma^2)$, la statistica $S = X_1 - X_2$ è ancillare per μ, non è ancillare per σ^2.

Esempio 2.30 Sia F una funzione di ripartizione per una variabile aleatoria continua a valori reali. Sia $X_1, \ldots, X_n$ campione casuale di legge $F(\cdot - \theta)$ con $\theta \in \mathbb{R}$ e per $x \in \mathbb{R}$ valga $F(x - \theta) = F_{X_i}(x)$ per ogni $i = 1, \ldots, n$. Notare che se F non è nota, si tratta di un modello semiparametrico. Con F nota è un modello parametrico. Verificare che la statistica $R = X_{(n)} - X_{(1)} \geq 0$ è ancillare per θ. Si definiscano le variabili aleatorie indipendenti $Z_i = X_i - \theta$ per $i = 1, \ldots, n$. La funzione di ripartizione di Z_i è F per ogni i infatti per $z_i \in \mathbb{R}$ si ha

$$\mathrm{P}(Z_i \leq z_i) = \mathrm{P}(X_i \leq z_i + \theta) = F(z_i + \theta - \theta) = F(z_i)$$

In particolare la legge di Z_i non dipende da θ. Per $r \geq 0$ si ha

$$F_R(r) = \mathrm{P}(R \leq r) = \mathrm{P}(X_{(n)} - X_{(1)} \leq r) = \mathrm{P}(X_{(n)} - \theta + \theta - X_{(1)} \leq r) = \mathrm{P}(Z_{(n)} - Z_{(1)} \leq r)$$

Poiché $Z_{(n)} - Z_{(1)}$ è funzione degli Z_i la cui legge non dipende da θ, neppure la legge di R dipende da θ. Si conclude che R è ancillare per θ.

Esercizio 2.31 Sia $X_1, \dots, X_n$ un campione casuale di legge $F(\frac{\cdot}{\sigma})$ con $\sigma > 0$. Verificare che ogni statistica del tipo $S\left(\frac{X_1}{X_n}, \dots, \frac{X_{n-1}}{X_n}\right)$ è ancillare per σ. Utilizzare le variabili aleatorie ausiliarie $Z_i = \frac{X_i}{\sigma}$.

Osservazioni 2.32 (ancillarità per modello non parametrico) Siano $X_1, \dots, X_n$ i.i.d. continue univariate con mediana uguale a zero. Si osservi che $T = \sum_{i=1}^n (X_i > 0) \sim Bin(n, 1/2)$ e quindi la legge di probabilità di T è la stessa per ogni modello probabilistico nella famiglia statistica. Ovvero T è ancillare per il modello statistico. La statistica T può essere usata per verificare l'ipotesi di mediana nulla.

Esempio 2.33 (statistica ancillare può dare informazioni sul parametro) Sia X una variabile aleatoria uniforme a valori in $\theta, \theta+1, \theta+2$ con $\theta \in \mathbb{Z}$. Si ha

X	θ	$\theta+1$	$\theta+2$
$\mathrm{P}(X = x)$	$1/3$	$1/3$	$1/3$
$X - \theta$	0	1	2

Siano X_1 e X_2 copie indipendenti di X e si consideri la statistica $T = (R, M)$ dove $R = X_{(2)} - X_{(1)}$ e $M = \dfrac{X_{(1)} + X_{(2)}}{2} = \dfrac{X_1 + X_2}{2}$. Notare che

1. lo spazio campionario dipende dal parametro (modello non regolare);
2. $R \in \{0, 1, 2\}$ e $M \in \{\theta, \theta + 1/2, \theta + 1, \theta + 3/2, \theta + 2\}$;
3. la statistica R è ancillare per θ [Esercizio (2.30)];
4. T è sufficiente minimale per θ. Questo si deduce dal fatto che $(X_{(1)}, X_{(2)})$ è sufficiente minimale per θ.[5]
5. Sia (r, m) una realizzazione di T e si osserva che m è intero. Da ciò si deduce che $m \in \{\theta, \theta+1, \theta+2\}$ e quindi $\theta \in \{m, m-1, m-2\}$.
6. Si aggiunge l'informazione che il valore osservato della statistica ancillare R è 2. Da ciò segue che $x_{(2)} = \theta + 2$ e $x_{(1)} = \theta$. Riassumendo le informazione sulle statistiche sono

$$\begin{cases} \dfrac{x_{(1)} + x_{(2)}}{2} = m \\ x_{(2)} - x_{(1)} = 2 \end{cases}$$

   Sostituendo $x_{(2)} = \theta + 2$ e $x_{(1)} = \theta$ si ottiene $\theta = m - 1$.

2.4 Statistiche complete

Definizione 2.34 *La statistica T è completa per θ se $\mathrm{E}_\theta\left(g(T(\underline{X}))\right) = 0$ per ogni $\theta \in \Theta$ implica $g = 0$.*

Osservazioni 2.35

1. Una versione più precisa della Definizione 2.34 afferma che, se $\mathrm{E}_\theta\left(g(T(\underline{X}))\right) = 0$ per ogni $\theta \in \Theta$, allora $g = 0$ deve valere con probabilità uno per ogni modello probabilistico nel modello statistico: T è completa per θ se $\mathrm{E}_\theta\left(g(T(\underline{X}))\right) = 0$ implica $\mathrm{P}_\theta(g(T) = 0) = 1$ per ogni θ.
2. Una versione poco più debole della definizione richiede che l'implicazione valga per g limitata (completezza limitata).
3. Non è difficile adattare la definizione di statistica completa a modelli non-parametrici.

[5] La legge congiunta di (X_1, X_2) è $\mathrm{P}(X_1 = x_1, X_2 = x_2) = \frac{1}{9}(x_{(1)} \geq \theta)(x_{(n)} \leq \theta + 2)(x_1, x_2 \in \mathbb{Z})$.

4. In genere è difficile dimostrare che una statistica è completa. Nell'Esercizio 2.40 dimostreremo che la statistica canonica in un modello di classe esponenziale è completa per il parametro naturale, che abbiamo già dimostrato essere sufficiente minimale.

Esercizio 2.36 Dimostrare che una statistica che è funzione invertibile di statistica completa è completa. Cosa cambia se nella definizione si suppone $\mathrm{E}_\theta\left(g(T(\underline{X}))\right)$ costante?

Esempio 2.37 Sia $X_1, \ldots, X_n$ un campione i.i.d. Bernoulli(θ) con $\theta \in]0,1[$. Verifichiamo che $T = \sum_{i=1}^n X_i$ è completa per θ. Per g funzione di T supponiamo quindi

$$0 = \mathrm{E}_\theta\left(g(T(\underline{X}))\right) = \sum_{t=0}^{n} \binom{n}{t} g(t)\theta^t(1-\theta)^{n-t} = (1-\theta)^n \sum_{t=0}^{n} \binom{n}{t} g(t) r^t$$

dove $r = \frac{\theta}{1-\theta}$. Questo è un polinomio in r. Per il principio di identità dei polinomi è zero se e solo se tutti i suoi coefficienti sono zero, cioè $\binom{n}{t}g(t)$ e quindi $g = 0$ identicamente.

Esempio 2.38 Sia $X \sim$ Uniforme($]-\theta,\theta[$) con $\theta > 0$. Poiché $\mathrm{E}_\theta(X) = 0$ per ogni θ, la statistica $T(X) = X$ non è completa per θ.

Esempio 2.39 Sia $X_1, \ldots, X_n$ un campione i.i.d. Uniforme($]0,\theta[$) con $\theta > 0$. Verificare che $T = X_{(n)}$ è completa per θ. Occorre verificare che se $\mathrm{E}_\theta(g(T)) = 0$ allora $g \equiv 0$.

1. Valga
$$0 = \mathrm{E}_\theta(g(T)) = \int_0^\theta g(t)\frac{nt^{n-1}}{\theta^n}\,dt = \frac{1}{\theta^n}\underbrace{\int_0^\theta g(t)nt^{n-1}\,dt}_{0}$$

2. Si derivi in θ
$$\begin{aligned}0 =& \frac{d}{d\theta}\left(\frac{1}{\theta^n}\int_0^\theta g(t)nt^{n-1}\,dt\right) = \theta^{-n}\frac{d}{d\theta}\int_0^\theta g(t)nt^{n-1}\,dt + \left(\frac{d}{d\theta}\theta^{-n}\right)\int_0^\theta g(t)nt^{n-1}\,dt\\ =& \theta^{-n}ng(\theta)\theta^{n-1} + 0 \qquad \text{perché } \mathrm{E}_\theta(g(T)) = 0\\ =& ng(\theta)/\theta\end{aligned}$$
e quindi $\mathrm{E}_\theta(g(T)) = 0$ implica $g = 0$ per ogni g tale che i passaggi precedenti sono possibili.

3. Possiamo effettivamente concludere che T è completa?

Esercizio 2.40 Nel modello statistico di classe esponenziale $f(T(\underline{x})) = \bar{h}(T(\underline{x}))\exp\left(\langle\theta, T(\underline{x})\rangle - \phi(\theta)\right)$ la statistica T è completa per il parametro naturale $\theta \in \mathbb{R}^p$ oltre ad essere sufficiente minimale. Infatti sia g una funzione a valori reali tale che

$$0 = \mathrm{E}_\theta\left(g(T)\right) = \int_{\mathbb{R}^p} g(t)\bar{h}(t)\exp\left(\langle\theta,t\rangle - \phi(\theta)\right)\,dt$$

e si considerino la parte positiva e la parte negativa di g, $g(t) = g^+(t) - g^-(t)$. Allora

$$\int_{\mathbb{R}^p} g^+(t)\bar{h}(t)\exp\left(\langle\theta,t\rangle - \phi(\theta)\right)\,dt = \int_{\mathbb{R}^p} g^-(t)\bar{h}(t)\exp\left(\langle\theta,t\rangle - \phi(\theta)\right)\,dt \tag{1}$$

ed in particolare per $\theta = 0$ si ha

$$\int_{\mathbb{R}^p} g^+(t)\bar{h}(t)\exp\left(-\phi(0)\right)\,dt = \int_{\mathbb{R}^p} g^-(t)\bar{h}(t)\exp\left(-\phi(0)\right)\,dt$$

semplificando $\exp(-\phi(0))$ si ha $\int_{\mathbb{R}^p} g^+(t)\bar{h}(t)\,dt = \int_{\mathbb{R}^p} g^-(t)\bar{h}(t)\,dt$. Dividendo l'Equazione (1) per questa quantità si ha

$$\int_{\mathbb{R}^p} e^{\langle\theta,t\rangle}\frac{g^+(t)\bar{h}(t)}{\int g^+(t)\bar{h}(t)\,dt}e^{-\psi(\theta)}\,dt = \int_{\mathbb{R}^p} e^{\langle\theta,t\rangle}\frac{g^-(t)\bar{h}(t)}{\int g^-(t)\bar{h}(t)\,dt}e^{-\psi(\theta)}\,dt$$

Semplificando $e^{-\psi(\theta)}$ e riconoscendo in $\dfrac{g^-(t)\bar{h}(t)}{\int g^-(t)\bar{h}(t)\,dt}$ e $\dfrac{g^+(t)\bar{h}(t)}{\int g^+(t)\bar{h}(t)\,dt}$ densità di probabilità, si nota che la precedente è un'uguaglianza tra funzioni generatrici dei momenti.

Theorem 2.41 (Teorema di Basu) *Sia $\underline{X}$ un campione su uno spazio di probabilità $(\Omega, \mathcal{A}, \mathrm{P})$ e sia Θ lo spazio parametrico di un modello statistico per $\underline{X}$. Siano V una statistica ancillare per $\theta \in \Theta$ e T una statistica sufficiente e completa per θ. Allora V e T sono indipendenti per ogni $\theta \in \Theta$.*

Proof. Dimostriamo che la legge congiunta di V e T è il prodotto delle leggi marginali.

- Sia B tale che $\{\omega \in \Omega : V(\omega) \in B\} \in \mathcal{A}$ cosicchè
$$(V \in B)(\omega) = \begin{cases} 1 \text{ se } V(\omega) \in B \\ 0 \text{ altrimenti} \end{cases}$$
è variabile aleatoria.
- $\mathrm{P}_\theta\left(V \in B\right) = \mathrm{E}_\theta\left((V \in B)\right)$ non dipende da θ poiché V è ancillare per θ.
- Per la "proprietà della torre" del valore atteso condizionato vale
$$\mathrm{E}_\theta\left(\mathrm{E}_\theta\left((V \in B)\,|T\right)\right) = \mathrm{E}_\theta\left((V \in B)\right) = \mathrm{P}\left(V \in B\right)$$
ed inoltre $\mathrm{E}_\theta\left((V \in B)\,|T\right)$ non dipende da θ per il teorema di Neyman-Fisher poiché T è sufficiente.
- Da ciò segue che l'argomento del valore atteso
$$\mathrm{E}_\theta\left(\mathrm{E}\left((V \in B)\,|T\right) - \mathrm{P}\left(V \in B\right)\right) = 0$$
è una statistica poiché non dipende da θ ed è funzione di T. Poiché T è completa allora
$$\mathrm{E}\left((V \in B)\,|T\right) = \mathrm{P}\left(V \in B\right)$$

Ora siano A e B due qualunque insiemi tali che $(V \in B)$ e $(T \in A)$ sono variabili aleatorie. Allora

$$\begin{aligned}
\mathrm{P}_\theta\left(V \in B, T \in A\right) &= \mathrm{E}_\theta\left((V \in B)(T \in A)\right) = \mathrm{E}_\theta\left(\mathrm{E}_\theta\left((V \in B)(T \in A)|T\right)\right) \qquad \text{perché } (T \in A) \text{ è funzione di } T \\
&= \mathrm{E}_\theta\left((T \in A)\,\mathrm{E}_\theta\left((V \in B)|T\right)\right) = \mathrm{E}_\theta\left((T \in A)\,\mathrm{P}\left(V \in B\right)\right) \\
&= \mathrm{P}\left(V \in B\right)\mathrm{E}_\theta\left((T \in A)\right) = \mathrm{P}\left(V \in B\right)\mathrm{P}_\theta\left(T \in A\right)
\end{aligned}$$

■

Esempio 2.42 (Teorema di Cochran) Siano $X_1, \ldots, X_n$ i.i.d. $\mathcal{N}(\mu, \sigma^2)$. La media campionaria $\bar{X}_n \sim \mathcal{N}(\mu, \sigma^2/n)$ è sufficiente e completa per μ (statistica canonica in modello di classe esponenziale). La legge di $(n-1)S_n^2/\sigma^2 = \dfrac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2} \sim \chi^2_{n-1}$ non dipende da μ e quindi la legge di S_n^2 non dipende da μ. Allora per il teorema di Basu $\bar{X}_n$ e S_n^2 sono indipendenti.

Esempio 2.43 Sia $X_1, \ldots, X_n$ i.i.d. Esponenziale(λ), $\lambda > 0$
$$f(x; \lambda) = 1/\lambda \exp(-x/\lambda)\,(x > 0)$$
Calcolare $\mathrm{E}_\lambda\left(\frac{X_n}{X_1 + \ldots + X_n}\right)$. Si osservi che $f(\cdot; \lambda)$ è una famiglia di scala e quindi per l'Esercizio 2.31 $\frac{X_n}{X_1 + \ldots + X_n}$ è ancillare per λ e quindi il valore atteso da calcolare non dipende da λ. La statistica $T(\underline{X}) = \sum_{i=1}^n X_i$ è completa, sufficiente minimale per λ perché statistica canonica in un modello di classe esponenziale. Quindi per il teorema di Basu $T(\underline{X})$ è indipendente da $\frac{X_n}{X_1 + \ldots + X_n}$.
Ricordiamo che $\mathrm{E}_\lambda(X_1) = \lambda$, somma di n esponenziali i.i.d. è Gamma(n, λ) e per $X \sim$ Gamma(α, β) il valore atteso di X è $\alpha\beta$ e la sua varianza è $\alpha\beta^2$. Si deduce

$$\begin{aligned}
\lambda = \mathrm{E}_\lambda(X_n) &= \mathrm{E}_\lambda\left(\sum_i X_i \frac{X_n}{\sum_i X_i}\right) \\
&= \mathrm{E}_\lambda\left(\sum_i X_i\right)\mathrm{E}\left(\frac{X_n}{\sum_i X_i}\right) = n\lambda\,\mathrm{E}_\lambda\left(\frac{X_n}{\sum_i X_i}\right)
\end{aligned}$$

da cui $\mathrm{E}_\lambda\left(\frac{X_n}{\sum_i X_i}\right) = 1/n$. Abbiamo calcolato il valore atteso di $\frac{X_n}{\sum_i X_i}$ senza conoscerne la distribuzione campionaria.

Esercizio 2.44 (Necessità dell'ipotesi di completezza nel teorema di Basu) Nell'Esempio 2.33 la statistica R è ancillare per θ e (R, M) è sufficiente minimale, però non sono indipendenti (dimostrarlo e.g. per $R = 0$ e $M = \theta + 1/2$ oppure nell'Osservazione 2.33 si è utilizzata un'informazione su R per specificare quella data da M). In particolare ne segue che (R, M) non è completa. Seppure può sembrare che una statistica ancillare (la cui distribuzione non dipende dal parametro) ed una statistica sufficiente minimale (che 'riduce' i dati al massimo senza perdere informazioni sul parametro) debbano essere indipendenti/non 'relazionati', questo esempio indica che la nozione giusta per discutere l'indipendenza tra statistiche sufficienti e statistiche ancillari è la completezza e non la minimalità.

Esercizio 2.45 Dimostrare che una statistica T sufficiente minimale e completa per θ è indipendente da ogni statistica ancillare per θ.

3 Stimatori puntuali

Sia $\underline{X} = (X_1, \ldots, X_n) \in \mathcal{X} \subset \mathbb{R}^n$ un campione su $(\Omega, \mathcal{A}, \mathrm{P})$ e sia $\mathcal{F}$ un modello statistico parametrizzato da $\theta \in \mathbb{R}^p$. Uno *stimatore puntuale di* θ è una funzione del (solo) campione $\underline{X}$. Più precisamente uno stimatore puntuale di θ è una statistica utilizzata per individuare un modello probabilistico all'interno del modello statistico. Ragionevolmente avrà la stessa dimensione di θ. In particolare uno stimatore è una variabile o vettore aleatorio. Una *stima* è una realizzazione di uno stimatore. In Appendice 6.7 presentiamo un esempio di stimatore non puntuale. Studieremo tecniche per determinare stimatori e tecniche per valutare stimatori puntuali. Anticipiamo

Definizione 3.1 *Uno stimatore U di $\theta \in \Theta$ è detto* corretto *o* non distorto *(per θ) se $\mathrm{E}_\theta(U) = \theta$ per ogni $\theta \in \Theta$.*

3.1 Metodo dei momenti

Sia $\underline{X} = (X_1, \ldots, X_n)$ un campione casuale e su ciascuna componente di $\underline{X}$ si assuma un modello statistico parametrizzato da $\theta \in \Theta \subseteq \mathbb{R}^p$. Per $j \in \mathbb{Z}_{>0}$ il *momento teorico j-esimo* di un generico elemento del campione rispetto alla legge di probabilità individuata da θ è definito come $\mu_j(\theta) = \mathrm{E}_\theta\left(X^j\right)$. Mentre il *momento empirico j-esimo* del campione è definito come $\hat{\mu}_j = \dfrac{\sum_{i=1}^n X_i^j}{n}$. Si noti che i momenti teorici potrebbero non essere definiti.

Per il *principio di sostituzione* o *metodo plug-in* uno *stimatore dei momenti* di θ è un valore $\hat{\theta}$ soluzione del sistema $\mu_j(\hat{\theta}) = \hat{\mu}_j$ per $j \in J$ sottoinsieme finito di $\mathbb{Z}_{>0}$.

Esempio 3.2 Sia $X_1, \ldots, X_n$ i.i.d. Uniform($]\theta_1, \theta_2[$) con $\theta_1 < \theta_2$ numeri reali. Il vettore dei parametri è $\theta = (\theta_1, \theta_2)$. Vale $\mu_1(\theta) = (\theta_1 + \theta_2)/2$ e $\mu_2 = (\theta_1^2 + \theta_2^2 + \theta_1\theta_2)/3$ e ponendo a sistema $\hat{\mu}_1 = \bar{X} = \mu_1(\hat{\theta})$ con $\hat{\mu}_2 = \sum_{i=1}^n X_i^2/n = \mu_2(\hat{\theta})$ si ottiene $\hat{\theta_1} = \hat{\mu}_1 - \sqrt{3(\hat{\mu}_2 - \hat{\mu}_1^2)}$ e $\hat{\theta_2} = \hat{\mu}_1 + \sqrt{3(\hat{\mu}_2 - \hat{\mu}_1^2)}$ e in conclusione si ha $\hat{\theta} = (\hat{\theta_1}, \hat{\theta_2})$ stimatore puntuale di (θ_1, θ_2).

Si noti che $\hat{\mu}_2 - \hat{\mu}_1^2 \geq 0$. Infatti per una sequenza di valori in $\mathbb{R}^n$, diciamo $\underline{x}$, dimostriamo $\dfrac{\sum_i x_i^2}{n} \geq \left(\dfrac{\sum_i x_i}{n}\right)^2$ equivalentemente $n\sum_i x_i^2 \geq \left(\sum_i x_i\right)^2$. Ricordando il prodotto scalare standard in $\mathbb{R}^n$ (per $\underline{a}, \underline{b} \in \mathbb{R}^n$ si ha $\langle a, b\rangle = \sum_i a_i b_i$) la disuguaglianza da dimostrare diventa

$$||1||\,||\underline{x}|| = \langle 1, 1\rangle\langle \underline{x}, \underline{x}\rangle \geq \langle \underline{x}, 1\rangle^2,$$

ma questa è la disuguaglianza di Cauchy-Schwarz (si veda anche la dimostrazione del Teorema 4.32).

Esempio 3.3 Per un campione casuale $X_1, \ldots, X_n$ per cui $\mathrm{E}(X_1)$ e $\mathrm{Var}(X_1)$ esistono finiti, uno stimatore dei momenti di $\mathrm{Var}(X_1)$ è $\dfrac{\sum_{i=1}^n X_i^2}{n} - \bar{X}^2$. Giustificare il risultato. Confrontare con S^2 e con $\frac{n-1}{n}S^2$.

Esercizio 3.4 Determinare uno stimatore dei momenti per $X_1, \ldots, X_n$ i.i.d. e

1. Binomial(k, p) con $(k, p) \in \mathbb{Z}_{>0} \times]0, 1[$ [6]
2. Geometric(p) con $p \in]0, 1[$ ovvero $p(1-p)^x$ per $x = 0, 1, \ldots$
3. $f_X(x, \theta) = (\theta + 1)x^\theta$ $(0 < x < 1)$ per $\theta \in \mathbb{Z}_{>1}$. Commentare il risultato
4. Poisson(λ) con $\lambda > 0$. E' preferibile usare $\hat{\mu}_1$ o $\hat{\mu}_2$?
5. Verificare che non esiste stimatore dei momenti se $X_1 \sim$ Cauchy$(0, \sigma)$, $\sigma \in \mathbb{R}$. [wikipedia inglese per la Cauchy è fatto bene]

Notare l'importanza dell'ipotesi di campione casuale e la pratica di utilizzare momenti di ordine basso, motivata sia da ragioni pratiche e computazionali che teoriche. Come si potrebbe estendere il metodo dei momenti al caso di variabili campionarie multivariate?

---

[6] Si osservi che lo stimatore dei momenti di (k, p) ottenuto usando i momenti primo e secondo è dato da $\hat{k} = \dfrac{\bar{X}^2}{\bar{X} - \frac{1}{n}\sum(X_i - \bar{X})^2}$ e $\hat{p} = \dfrac{\bar{X}}{\hat{k}}$. Può capitare di ottenere delle stime di $\hat{k}$ negative (!), quando la media campionaria è più piccola della varianza campionaria, indice di forte variabilità nei dati.

3.2 Metodo di massima verosimiglianza

Definizione 3.5 *Sia $\underline{X}$ un campione statistico su $(\Omega, \mathcal{A}, \mathrm{P})$ e si consideri un modello statistico per $\underline{X}$ parametrizzato da $\theta \in \Theta \subset \mathbb{R}^p$. La legge congiunta del campione interpretata come funzione del parametro θ è chiamata* funzione di verosimiglianza*.*

***** ampliare l'interpretazione di verosimiglianza ***** Spesso si usa la lettera $\mathcal{L}$ per indicare la funzione di verosimiglianza. La distinzione tra variabili campionare e il parametro è indicata da ',' oppure ';'. Nel caso di campione discreto si ha per $\underline{x} \in \mathcal{X}$ e $\theta \in \Theta$

$$\mathcal{L}(\theta; \underline{x}) = f_{\underline{X}}(\underline{x}; \theta) = \mathrm{P}_\theta(\underline{X} = \underline{x}) \tag{2}$$

e per campione continuo con densità congiunta $f_{\underline{X}}$ si ha

$$\mathcal{L}(\theta; \underline{x}) = f_{\underline{X}}(\underline{x}; \theta)$$

che per campione casuale si semplifica a

$$\mathcal{L}(\theta; \underline{x}) = \prod_{i=1}^{n} f_{X_1}(x_i; \theta)$$

Se la funzione di verosimiglianza è strettamente positiva si definisce la *funzione di log-verosimiglianza* come

$$l(\theta) = \log \mathcal{L}(\theta, \underline{x})$$

Osservazioni 3.6
1. Dall'Equazione (2) si deduce che la funzione di verosimiglianza rappresenta la probabilità sotto il modello indicato da θ di osservare il valore $\underline{x}$ (i dati, le osservazione).
2. Se T è statistica sufficiente per θ, cioè $f_{\underline{X}}(x, \theta) = h(\underline{x}) g(T(\underline{x}), \theta)$ allora $\mathcal{L}(\theta; \underline{x}) \propto g(T(\underline{x}), \theta)$.

Esempio 3.7 Siano $X_1, \ldots, X_n$ indipendenti ed identicamente distribuite secondo una legge uniforme su $]0, \theta[$ per $\theta \in]0, 1[$. Allora

$$\mathcal{L}(\theta; \underline{x}) = f_{\underline{X}}(\underline{x}; \theta) = \prod_{i=1}^{n} \frac{1}{\theta} (0 < x_i < \theta) = \frac{(0 < \min_i x_i)\left(x_{(n)} < \theta\right)}{\theta^n}$$

Definizione 3.8 Stimatore di massima verosimiglianza di θ *è*

$$\hat{\theta}_{MV}(\underline{X}) = \arg\sup_{\theta \in \Theta} \mathcal{L}(\theta; \underline{X})$$

Osservazioni 3.9
1. Si dice anche stimatore *in* massima verosimiglianza di θ.
2. $\hat{\theta}_{MV}(\underline{X})$ è una variabile/vettore aleatorio, è una statistica, è uno stimatore.
3. $\hat{\theta}_{MV}$ potrebbe non essere unico, potrebbe non esistere (per esempio essere $+\infty$), potrebbe non appartenere a Θ (appartiene però alla sua chiusura euclidea).
4. Nel caso di campione discreto è evidente l'interpretazione dello stimatore di massima verosimiglianza come quel valore del parametro che seleziona il modello probabilistico, all'interno del modello statistico, per cui il valore osservato $\underline{x}$ è il più probabile.
5. Poichè la funzione logaritmo è strettamente crescente, vale $\hat{\theta}_{MV}(\underline{X}) = \arg\sup_{\theta \in \Theta} l(\theta; \underline{X})$.
6. Nel calcolo di $\hat{\theta}_{MV}$ si possono trascurare i fattori della verosimiglianza costanti in θ, per esempio se T è sufficiente per θ allora
$$\hat{\theta}_{MV} = \arg\sup_{\theta \in \Theta} g(T(\underline{X}); \theta)$$
e lo stimatore di massima verosimiglianza è funzione del (campione solo tramite) la statistica sufficiente.

Osservazioni 3.10 1. Il determinare uno stimatore di massima verosimiglianza è quindi un problema di massimizzazione di funzioni multivariate e soggetto alle difficoltà inerenti questi calcoli, comprese quelle di instabilità numerica.

2. Se $\mathcal{L}$ è differenziabile in θ, lo stimatore di massima verosimiglianza è soluzione delle *equazioni di verosimiglianza* definite come $\frac{\partial}{\partial\theta_i}\mathcal{L}(\theta) = 0$ per $i = 1, \ldots, p$ o, equivalentemente se l è definita, $\frac{\partial}{\partial\theta_i}l(\theta) = 0$.

3. Occorre verificare che i punti critici delle equazioni di verosimiglianza siano effettivamente punti di massimo. In due dimensioni, se esiste la matrice Hessiana della funzione di verosimiglianza (o log-verosimiglianza) e nel punto critico il suo determinante è positivo ed almeno un suo elemento diagonale è negativo, allora il punto critico è un massimo locale ed un candidato ad essere massimo globale e quindi stimatore di massima verosimiglianza. Nel caso di tre parametri si può studiare il segno degli autovalori della matrice Hessiana. E nel caso di spazio parametrico di dimensione maggiore di tre?

Esercizio 3.11 1. Per il modello dell'Esempio 3.7 e per $n = 1$ e $n = 2$ fare il grafico della funzione di densità e della funzione di verosimiglianza.

2. Le due densità di probabilità in Figura 1 sono gaussiane troncate in $[-4, 4]$ e formano il modello statistico, per un campione di taglia uno, $\mathcal{F} = \{f1, f2\}$ con spazio dei parametri $\Theta = \{1, 2\}$. Con una sola osservazione $x^{oss} \neq 0$ la stima di massima verosimiglianza è

$$\hat{\theta}_{MV} = \begin{cases} 1 & \text{se } x^{oss} < 0 \\ 2 & \text{se } x^{oss} > 0 \end{cases}$$

per $x^{oss} = 0$ sia 1 che 2 sono stime di massima verosimiglianza.

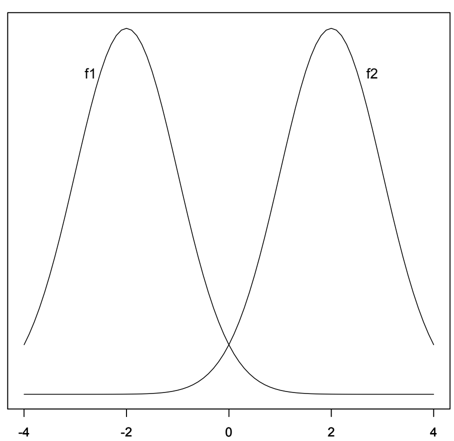

Figure 1: Esempio 3.11

Esempio 3.12 Si considerino le seguenti due densità di probabilità per $X \in \{0, 1, 2\}$ indicizzate da $\theta \in \{\theta_0, \theta_1\}$

$X = x$	0	1	2
θ_0	0.8	0.1	0.1
θ_1	0.2	0.3	0.5

Se si osserva $x = 0$ allora lo stimatore di massima verosimiglianza di θ è θ_0, altrimenti è θ_1.

Esercizio 3.13 1. Sia $X_1, \ldots, X_n$ i.i.d. Bernoulli(p) con $p \in]0,1[$. Utilizzando la funzione di log-verosimiglianza verificare che $\hat{p} = \bar{X}$ è stimatore di massima verosimiglianza di p. Notare che è anche statistica sufficiente. Verificare che il valore atteso di $\hat{p}$ è p. Sono queste proprietà generali degli stimatori di massima verosimiglianza?

2. Siano $X_1, \ldots, X_n$ copie indipendenti di $X \sim \text{Exp}(\lambda)$ con $\lambda > 0$. Si osservi che per $f_{X_1}(x_1) = \lambda \exp(-\lambda x_1)(x_1 > 0)$ si ha $\frac{\partial l}{\partial \lambda} = \frac{n}{\lambda} - \sum x_i = 0$ per $\hat{\lambda} = n / \sum x_i$ e che $\frac{\partial^2 l}{\partial \lambda^2} = -\frac{n}{\lambda^2} < 0$. Si ha quindi $\hat{\lambda}_{MV} = n / \sum X_i = 1/\bar{X}$. Si noti che $\hat{\lambda}_{MV}$ è stimatore distorto di λ infatti per la disuguaglianza di Jensen vale la maggiorazione stretta
$$\mathrm{E}(1/\bar{X}) > 1/\mathrm{E}(\bar{X}) = \frac{1}{\frac{1}{n}\frac{n}{\lambda}} = \lambda$$

In particolare gli stimatori di massima verosimiglianza possono non essere corretti. Indichiamo che sono asintoticamente corretti senza dimostrarlo (v. Esempio 3.18). L'esercizio successivo indica un metodo tramite il quale talora è possibile correggere uno stimatore.

Esercizio 3.14 (Stimatori di massima verosimiglianza possono essere distorti) Sia $X_1, \ldots, X_n$ i.i.d. Uniform$(]0, \theta[)$ con $\theta > 0$.

1. Dimostrare che lo stimatore di massima verosimiglianza di $\theta \in \mathbb{R}_{>0}$ è $\hat{\theta} = X_{(n)}$.
2. Verificare che la densità di probabilità di $X_{(n)}$ è $f(y) = \frac{n y^{n-1}}{\theta^n}$ con $y \in]0, \theta[$.
3. Verificare che $\mathrm{E}_\theta(\hat{\theta}) = \frac{n}{n+1}\theta$ e che $\mathrm{Var}_\theta(\hat{\theta}) = \frac{n\theta^2}{(n+2)(n+1)^2}$. Si noti che $\hat{\theta}$ è distorto per θ. Lo si può correggere ottenendo un altro stimatore $\hat{\theta}' = \frac{n+1}{n}\hat{\theta} = \frac{n+1}{n} X_{(n)}$.
4. Verificare che $\hat{\theta}'$ è statistica sufficiente per θ.
5. Confrontare $\hat{\theta}'$ con $\hat{\theta}_{MOM} = 2\bar{X}$.
   (a) Riconoscere in $\hat{\theta}_{MOM}$ uno stimatore dei momenti di θ, verificare che è corretto per θ. Notare che $\bar{X}$ non è statistica sufficiente per θ.
   (b) Calcolare la varianza e gli errori quadratici medi per entrambi gli stimatori.

Esempio 3.15 Siano $X_1, \ldots, X_n$ i.i.d. Uniform$(]\theta, \theta + 1[)$ con $\theta \in \mathbb{R}$. La funzione di verosimiglianza è

$$\mathcal{L}(\theta; \underline{x}) = (x_{(1)} > \theta)(x_{(n)} - 1 < \theta) = (x_{(n)} - 1 < \theta < x_{(1)})$$

Ogni variabile aleatoria nell'intervallo stocastico $[X_{(n)} - 1, X_{(1)}]$ è stimatore di massima verosimiglianza. Si osservi che $\mathrm{Prob}_\theta(X_{(n)} - 1 \geq X_{(1)}) = 0$ per ogni θ e quindi con probabilità uno esistono infiniti stimatori di massima verosimiglianza.

Esercizio 3.16 (IMPORTANTE) Calcolare gli stimatori di massima verosimiglianza nel caso di campioni casuali di taglia n

1. $\mathcal{N}(\mu, \sigma^2)$ per $(\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}^2$ supponendo σ^2 noto, μ noto, entrambi non noti;
2. Poisson(λ) con $\lambda > 0$.

Theorem 3.17 (di invarianza per stimatori di massima verosimiglianza) *Sia $\hat{\theta}$ di massima verosimiglianza per θ allora $g(\hat{\theta})$ è di massima verosimiglianza per $g(\theta)$ per ogni funzione g.*

Proof. Se g è invertibile, la dimostrazione è immediata. Definiamo $A_\eta = \{\theta \in \Theta : g(\theta) = \eta\}$ per ogni η valore assunto da g e la *funzione di verosimiglianza indotta o profilo*

$$\mathcal{L}^*(\eta; \underline{x}) = \sup_{\{\theta : g(\theta) = \eta\}} \mathcal{L}(\theta; \underline{x}) \in [0, +\infty]$$

Ora

$$
\begin{aligned}
\sup_\eta \mathcal{L}^*(\eta;\underline{x}) &= \sup_\eta \sup_{\{\theta: g(\theta)=\eta\}} \mathcal{L}(\theta;\underline{x}) && \text{per definizione di } \mathcal{L}^* \\
&= \sup_\theta \mathcal{L}(\theta;\underline{x}) && \text{per proprietà di sup} \\
&= \mathcal{L}(\hat\theta;\underline{x}) && \text{per definizione di MLE di } \theta / \hat\theta \text{ è arg sup} \\
&= \sup_{\{\theta: g(\hat\theta)=g(\theta)\}} \mathcal{L}(\theta;\underline{x}) = \mathcal{L}^*(g(\hat\theta);\underline{x})
\end{aligned}
$$

■

Esempio 3.18 1. Lo stimatore di massima verosimiglianza di p^2 in un campione casuale Bernoulliano di parametro $p \in]0,1[$ è $\hat p^2 = (\bar X)^2 = (\sum X_i)^2/n^2$.

2. Il valore atteso di $\hat p^2$ rispetto alle leggi di probabilità nel modello è

$$
\mathrm{E}(\hat p^2) = \mathrm{E}_p\left(\frac{(\sum X_i)^2}{n^2}\right) = \frac{1}{n^2}\left(\mathrm{Var}_p\left(\sum X_i\right) + \left(\mathrm{E}_p\left(\sum X_i\right)\right)^2\right) = \frac{np(1-p)}{n^2} + \frac{n^2p^2}{n^2} = p^2 + \frac{p(1-p)}{n}
$$

In particolare lo stimatore di massima verosimiglianza di p^2 esiste, unico, ma non è corretto. Ma il secondo termine nell'uguaglianza precedente (il fattore di distorsione) converge a zero se n tende a più infinito. Si dice che lo stimatore $\hat p^2$ è asintoticamente corretto per p^2. Questa è una proprietà condivisa dagli stimatori di massima verosimiglianza. Più in generale vale che gli stimatori di massima verosimiglianza sono asintoticamente consistenti (v. § 4.1).

3. Determinare lo stimatore di massima verosimiglianza di $\sqrt{p(1-p)}$.

Theorem 3.19 (Stima in massima verosimiglianza e modelli di classe esponenziale) *Sia*

$$
X \sim h(\underline{x}) \exp\left(\langle\theta, T(\underline{x})\rangle - \psi(\theta)\right)
$$

con $\theta \in \mathbb{R}^p$ *e* $\underline{x} \in \mathcal{X} \subseteq \mathbb{R}^v$ *e* $T = (T_1, \dots, T_p) \in \mathbb{R}^p$. *Siano* $X_1, \dots, X_n$ *copie indipendenti di* X.

1. *Il campione* $\underline{X} = (X_1, \dots, X_n)$ *è di classe esponenziale*

2. *e la funzione di verosimiglianza associata è*

$$
\mathcal{L}(\theta; \underline{x}_1, \dots, \underline{x}_n) = \prod_{i=1}^n h(\underline{x}_i) \exp\left(\langle\theta, \sum_{i=1}^n T(\underline{x}_i)\rangle - n\psi(\theta)\right)
$$

3. *Vale* $\mathrm{E}_\theta(T_i) = \dfrac{\partial\psi(\theta)}{\partial\theta_i}$ *per* $i = 1, \dots, p$

4. $\mathrm{Cov}_\theta(T_i, T_j) = \dfrac{\partial^2\psi(\theta)}{\partial\theta_i\theta_j}$ *per* $i, j = 1, \dots, p$

5. *Inoltre* $T(X)$ *è stimatore di massima verosimiglianza di* $\mathrm{Gradient}_\theta\, \psi = \left(\dfrac{\partial\psi(\theta)}{\partial\theta_1}, \dots, \dfrac{\partial\psi(\theta)}{\partial\theta_p}\right)$

6. *ed è stimatore corretto.*

Proof. Trascuriamo lo studio dell'esistenza dei valori attesi e covarianze. Si osservi che corrisponde a questioni di esistenza di derivate prime e seconde di una funzione del parametro. Si lascia al lettore la dimostrazione dei primi due punti. Per il resto della dimostrazione useremo

$$
1 = \int_{\mathcal{X}} h(\underline{x}) \exp\left(\langle\theta, T(\underline{x})\rangle - \psi(\theta)\right)\, d\underline{x} = \left(\int_{\mathcal{X}} h(\underline{x}) \exp\left(\langle\theta, T(\underline{x})\rangle\right)\, d\underline{x}\right) \exp(-\psi(\theta))
$$

da cui $\exp(\psi(\theta)) = \int_\chi h(\underline{x}) \exp\left(\langle\theta, T(\underline{x})\rangle\right) \, d\underline{x}$. Nel seguito supponiamo di poter commutare gli operatori di integrazione e derivazione coinvolti.[7] Per la dimostrazione del punto 3, da cui segue il punto 6., si consideri che

$$\begin{aligned}
\frac{\partial\psi}{\partial\theta_i} &= \frac{\frac{\partial}{\partial\theta_i}\int_\chi h(\underline{x})\exp\left(\langle\theta, T(\underline{x})\rangle\right)\, d\underline{x}}{\int_\chi h(\underline{x})\exp\left(\langle\theta, T(\underline{x})\rangle\right)\, d\underline{x}} = \frac{\int_\chi h(\underline{x})\frac{\partial}{\partial\theta_i}\left(\exp\left(\langle\theta, T(\underline{x})\rangle\right)\right)\, d\underline{x}}{\int_\chi h(\underline{x})\exp\left(\langle\theta, T(\underline{x})\rangle\right)\, d\underline{x}} \\
&= \frac{\int_\chi hT_i\exp(\langle\theta, T\rangle)\, d\underline{x}}{\exp(\psi(\theta))} = \int_\chi T_i h\exp(\langle\theta, T\rangle - \psi(\theta)\, d\underline{x}
\end{aligned}$$

La dimostrazione del punto 4. è simile: occorre derivare rispetto θ_j oltre che rispetto a θ_i. Per il quinto punto si ponga uguale a zero il gradiente della funzione di log-verosimiglianza (equazioni di log-verosimiglianza)

$$\frac{\partial l}{\partial\theta_i} = T_i - \frac{\partial\psi}{\partial\theta_i} = 0$$

e si verifichi che la matrice Hessiana è definita-negativa[8] nei punti critici di l sfruttando la relazione

$$\frac{\partial^2 l}{\partial\theta_i\theta_j} = -\frac{\partial^2\psi(\theta)}{\partial\theta_i\theta_j} = -\operatorname{VarCov}_\theta(T_i, T_j)$$

■

Esercizio 3.20 Alla luce di questo teorema, rifare l'esercizio (3.16).

Esercizio 3.21 Si ricordi che $\mathcal{N}(\mu, \sigma^2)$, con $(\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_{>0}$, è un modello di classe esponenziale. In particolare da

$$\exp\left(-\frac{x^2}{2\sigma^2} + \frac{\mu}{\sigma^2}x - \frac{\mu^2}{2\sigma^2} - \frac{1}{2}\log\sigma^2\right)$$

si deduce che $\theta = \left(\frac{\mu}{\sigma^2}, -\frac{1}{2\sigma^2}\right) \in \mathbb{R} \times \mathbb{R}_{<0}$ è parametro naturale e vale $(\mu, \sigma^2) = \left(-\frac{\theta_1}{2\theta_2}, -\frac{1}{2\theta_2}\right)$. Per un n-campione casuale $T = \left(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2\right)$ è statistica canonica e la funzione dei cumulanti è

$$\psi(\theta) = n\left(\frac{\mu^2}{2\sigma^2} + \frac{1}{2}\log\sigma^2\right) = n\left(-\frac{1}{2}\log(-2\theta_2) - \frac{\theta_1^2}{4\theta_2}\right)$$

Per il Teorema 3.19 T è stimatore di massima verosimiglianza di $\nabla_\theta\psi$. Calcoliamo

$$\begin{cases} \dfrac{\partial\psi}{\partial\theta_1} = -n\dfrac{\theta_1}{2\theta_2} = n\mu \\ \dfrac{\partial\psi}{\partial\theta_2} = n\left(-\dfrac{1}{2\theta_2} + \dfrac{\theta_1^2}{4\theta_2^2}\right) = n\left(\sigma^2 + \mu^2\right) \end{cases}$$

Per il teorema di invarianza degli stimatori di massima verosimiglianza si ha

$$\begin{aligned}
T_1 &= \sum X_i = n\hat{\mu} = n\hat{\mu} \\
T_2 &= \sum X_i^2 = n(\sigma^2 \hat{+} \mu^2) = n\left(\hat{\sigma^2} + \hat{\mu}^2\right)
\end{aligned}$$

[7] Spesso per gli esempi che facciamo si può applicare la regola di Leibnitz: $\frac{d}{d\theta}\int_{a(\theta)}^{b(\theta)} f(x,\theta)\, dx = f(b(\theta),\theta)\frac{d}{d\theta}b(\theta) - f(a(\theta),\theta)\frac{d}{d\theta}a(\theta) + \int_{a(\theta)}^{b(\theta)}\frac{d}{d\theta}f(x,\theta)\, dx$ con f, a, b differenziabili in θ. Per a e b costanti (v. modelli regolari) si semplifica a $\frac{d}{d\theta}\int_a^b f(x,\theta)\, dx = \int_a^b \frac{d}{d\theta}f(x,\theta)\, dx$.

[8] Matrici di varianza-covarianza sono semidefinite positive, infatti $\operatorname{VarCov}(\underline{X}) = \operatorname{E}\left((\underline{X} - \operatorname{E}(\underline{X}))(\underline{X} - \operatorname{E}(\underline{X}))^t\right) = \operatorname{E}\left(\underline{Y}\underline{Y}^t\right) = \operatorname{VarCov}(\underline{Y})$ where $\underline{Y} = \underline{X} - \operatorname{E}(\underline{X})$. Ora per ogni $\underline{x} \in \mathbb{R}^n$ vale $x^t\operatorname{VarCov}(\underline{Y})x = \operatorname{E}(\underline{x}^t\underline{Y} \cdot \underline{Y}^t\underline{x}) = \operatorname{E}((Y^t\underline{x})^t Y^t\underline{x}) = \operatorname{E}(Z^2)$ dove $Z = Y^t\underline{x}$.

e quindi gli stimatori di massima verosimiglianza dei parametri μ e σ^2 sono

$$\hat{\mu} = \sum X_i/n$$
$$\hat{\sigma^2} = \frac{\sum X_i^2}{n} - \hat{\mu}^2 = \frac{\sum X_i^2 - (\sum X_i)^2/n}{n} = \frac{\sum_{i=1}^n \left(X_i - \bar{X}\right)^2}{n}$$

Ricordando che $S^2 = \dfrac{\sum(X_i - \bar{X})^2}{n-1}$ è corretto per σ^2 si ha che lo stimatore di massima verosimiglianza di σ^2 non è corretto. Calcolare lo stimatore di massima verosimiglianza di θ.

Esempio 3.22 (stimatori di massima verosimiglianza possono essere distorti, un altro esempio) Sia $X_1, \ldots, X_n$ un campione casuale univariato con $\mathrm{E}(X_1)$ e $\mathrm{Var}(X_1)$ finiti (quadrato integrabile). Per la linearità dell'operatore 'valore atteso', la media campionaria è stimatore corretto di $\mathrm{E}(X_1)$. Mentre $\dfrac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}$ non è stimatore corretto di $\mathrm{Var}(X_1)$ (si ricordi che è stimatore di massima verosimiglianza) infatti:

$$\mathrm{E}_\theta\left(\sum_{i=1}^n (X_i - \bar{X})^2\right) = \mathrm{E}_\theta\left(\frac{n-1}{n}\sum_{i=1}^n X_i^2 - \frac{1}{n}\sum_{j\neq k} X_j X_k\right) = \ldots = (n-1)\sigma^2$$

Ne segue che $\dfrac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$ (lo stimatore dei minimi quadrati) è corretto per $\mathrm{Var}(X_1)$.

Esercizio 3.23 (Adimari e Pauli, Esercizio 12) Sia $X_1, \ldots, X_n$ un campione casuale di legge $\mathcal{N}(\theta, \sigma^2)$ con σ^2 noto e $\theta \in \Theta = \{-2, 0, 1\}$. Il modello è regolare, è sottomodello di un modello di classe esponenziale, ma non è di classe esponenziale perché lo spazio dei parametri non è uno spazio vettoriale. Si faccia un grafico rappresentante le tre funzione di densità nel modello e si stabilisca a cosa corrispondono i valori $x = -1$ e $x = 0.5$. Dedurre dal grafico che lo stimatore di verosimiglianza di θ è la funzione

$$\begin{cases} -2 & \text{se } \bar{x} < -1 \\ 0 & \text{se } -1 < \bar{x} < 0.5 \\ 1 & \text{se } 0.5 < \bar{x}. \end{cases}$$

Esercizio 3.24 Sia V stimatore di massima verosimiglianza di $\theta \in \mathbb{R}$ ed esistano $a, b \in \mathbb{R}$ con $a \neq 0$ tale che $\mathrm{E}_\theta(V) = a\theta + b$. Allora $(V - b)/a$ è stimatore corretto di θ.

3.3 Principio di invarianza/equivarianza

Si veda il libro di Casella Berger paragrafi 6.3 e 7.2.4. Finora abbiamo studiato la tecnica di riduzione dei dati per sufficienza e la sua relazione con la funzione di verosimiglianza: se T è statistica sufficiente per θ e $\underline{x}, \underline{y} \in \mathcal{X}$ sono tali che $T(\underline{x}) = T(\underline{y})$ allora si ha la stessa inferenza su θ per verosimiglianza utilizzando $\underline{x}$ o $\underline{y}$. Infatti per la sufficienza di T esistono due funzioni h, g tali che $\mathcal{L}(\theta; \underline{x}) = h(\underline{x})g(T(\underline{x}); \theta)$ e quindi $\mathcal{L}(\theta; \underline{x}) \propto g(T(\underline{x}); \theta) = g(T(\underline{y}); \theta) \propto \mathcal{L}(\theta; \underline{y})$.
Ora introduciamo un altro principio di inferenza basato sul concetto di invarianza.
Supponiamo un esperimento in cui è di interesse stimare il diametro medio degli alberi di una foresta. Il campione potrà essere misurato in centimetri o in piedi o in qualche altra unità di misura lineare. Se si misurano gli stessi alberi, si dovrà ottenere la stessa stima, cambierà l'unità di misura rispetto alla quale è espressa ma le deduzioni inferenziali sul valore medio del diametro dovranno essere le stesse. Questo è un esempio di *invarianza rispetto all'unità di misura (measurement invariance)* per cui l'inferenza non dipende dalla scala di misura.
Ora supponiamo che due problemi inferenziali abbiano la stessa struttura matematica. Allora dalla stessa procedura inferenziale si dovranno dedurre le stesse conclusioni per i due problemi. La struttura matematica di un problema inferenziale in questo contesto è il modello statistico. Si parla di *invarianza formale*. Per esempio per stimare l'altezza media degli alberi in una foresta, la lunghezza del collo delle giraffe, il costo medio delle uova è ragionevole considerare un campione identicamente distribuito, ragionevolmente non indipendente, e lo stesso modello statistico.

Secondo il principio di invarianza se $g(\underline{X})$ è un cambiamento di unità di misura tale che i modelli statistici per $\underline{X}$ e per $g(\underline{X})$ hanno la stessa struttura formale, allora una procedura di inferenza dovrebbe essere invariante sia rispetto al cambio di misura sia formalmente.

Definizione 3.25 *Un* gruppo di trasformazioni $\mathcal{G}$ *dello spazio campionario* $\mathcal{X}$ *è un insieme di funzioni da* $\mathcal{X}$ *a* $\mathcal{X}$ *tale che*

1. *per ogni* $g, h \in \mathcal{G}$ *anche* $g \circ h \in \mathcal{G}$ *(è chiuso rispetto alla composizione, posso trasformare il trasformato),*

2. *per ogni* $g, h, l \in \mathcal{G}$ *allora* $(h \circ g) \circ l = h \circ (g \circ l)$ *(proprietà associativa, è lo stesso ruotare ciò che è stato scalato e traslato o prima traslare e poi applicare un rotazione e traslazione)*

3. *esiste* $i_{\mathcal{G}} \in \mathcal{G}$ *per cui per ogni* $g \in \mathcal{G}$ *si ha* $i_{\mathcal{G}} \circ g = g \circ i_{\mathcal{G}} = g$ *(gruppo con identià)*

4. *per ogni* $g \in \mathcal{G}$ *esiste* $g^{-1} \in \mathcal{G}$ *tale che* $g \circ g^{-1} = g^{-1} \circ g = i_{\mathcal{G}}$ *(inverso, traslo di* θ *e di* $-\theta$ *e non ho cambiato lo spazio campionario).*

Definizione 3.26 *Il modello statistico* $\mathcal{F}$ *parametrizzato da* $\theta \in \Theta$ *è* invariante per (il gruppo di trasformazioni dello spazio campionario) $\mathcal{G}$ *se per* $\underline{X} \sim f(\cdot;\theta)$ *e per* $g \in \mathcal{G}$ *esiste unico* $\theta' \in \Theta$ *tale che* $\underline{Y} = g(\underline{X}) \sim f(\cdot;\theta')$.

Se $\mathcal{F}$ è invariante rispetto a $\mathcal{G}$ allora dato $g \in \mathcal{G}$ ad ogni θ è associato un unico $\theta' \in \Theta$. Sia $\bar{g}(\theta) = \theta'$ la funzione da Θ a Θ indotta da g.

Esempio 3.27 Siano $X \sim \text{Binomial}(n, \theta)$ con n noto e $\theta \in]0,1[= \Theta$ e $\mathcal{G} = \{g_1, g_2\}$ con $g_1(x) = n - x$ e $g_2(x) = x$ per $x \in \{0, 1, \dots, n\} = \mathcal{X}$. La variabile aleatoria $g_2(X) = X \sim \text{Binomial}(n, \theta)$ conta i successi mentre $g_1(X) \sim \text{Binomial}(n, 1 - \theta)$ conta i fallimenti. Il modello statistico determinato da $\text{Binomial}(n, \theta)$ con $\theta \in]0, 1[$ e n noto è invariante rispetto a $\mathcal{G}$ e $\bar{g}_2(\theta) = \theta$ e $\bar{g}_1(\theta) = 1 - \theta$ per $\theta \in]0, 1[$. Mentre se $\Theta =]1/2, 1[$ il modello statistico binomiale non è invariante rispetto a $\mathcal{G}$.

Esercizio 3.28 Siano $X_1, \dots, X_n$ i.i.d. $\mathcal{N}(\mu, \sigma^2)$ con $(\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_{>0}$ e $\mathcal{G} = \{g_a(\underline{x}) = (x_1 + a, \dots, x_n + a : a \in \mathbb{R}\}$. Verificare che il modello normale è invariante rispetto a $\mathcal{G}$. E' invariante anche rispetto al gruppo di trasformazioni $\mathcal{H} = \{g_a(\underline{x}) = (x_1 + a_1, \dots, x_n + a_n) : a = (a_1, \dots, a_n) \in \mathbb{R}^n\}$? Determinare un altro modello statistico invariante rispetto ad $\mathcal{H}$.

Esercizio 3.29 *Un modello statistico* $\mathcal{F}$ *per un campione casuale* $X_1, \dots, X_n$ *è invariate rispetto al gruppo di permutazione delle* X_i *ovvero degli indici* $1, \dots, n$*?*

Definizione 3.30 *Sia* $\mathcal{F}$ *invariante rispetto a* $\mathcal{G}$ *e sia* $W(\underline{X})$ *uno stimatore puntuale di* θ*. Lo stimatore* W è invariante rispetto a $\mathcal{G}$ *se per ogni* $\underline{X}$, $\theta \in \Theta$ *e* $g \in \mathcal{G}$ *vale* $W(g(\underline{x})) = \bar{g}(W(\underline{x}))$.

Questa definizione equaglia l'invarianza in misura, $\bar{g}(W(\underline{x}))$ stima di $\bar{g}(\theta)$, con l'invarianza formale, $W(g(\underline{x}))$ stima di $\bar{g}(\theta)$.

Esempio 3.31 (Esercizio 3.27) Sia $T(X)$ uno stimatore di θ e $T^*(x)$ uno stimatore di $1-\theta$. L'invarianza per misura impone $T(x) = 1-T^*(n - x)$ e quella formale impone la stessa procedura e quindi

$$T(x) = 1 - T^*(n - x) = 1 - T(n - x)$$

In particolare per uno stimatore invariante rispetto alla trasformazione successi/fallimenti, noto $T(x)$ è noto anche $T(n-x)$. Verificare che la stima della probabilità di successo data dalla proporzione di successi osservati $\frac{x}{n}$ è invariante per $\mathcal{G}$.

Esempio 3.32 Siano $X_1, \dots, X_n$ i.i.d. e si consideri un modello statistico di posizione con funzione di ripartizione $F(\dots - \theta)s$ e $\theta \in \mathbb{R}$ e sia $\mathcal{G} = \{g_a(\underline{x}) = (x_1 + a, \dots, x_n + a) : a \in \mathbb{R}, \underline{x} \in \mathbb{R}^n\}$.

1. Si consideri
$$\underline{Y} = g_a(\underline{X}) = \underline{X} + a = (X_1 + a, \ldots, X_n + a)$$
Il campione trasformato $\underline{Y}$ ha legge determinata da
$$\begin{aligned} \mathrm{P}(\underline{Y} \leq \underline{y}) &= \mathrm{P}(Y_1 \leq y_1, \ldots, Y_n \leq y_n) = \mathrm{P}(X_1 \leq y_1 - a, \ldots, X_n \leq y_n - a) \\ &= \mathrm{P}(X_1 \leq y_1 - a) \ldots \mathrm{P}(X_n \leq y_n - a) = \prod_{i=1}^{n} F(y_i - a - \theta) \end{aligned}$$
da cui la trasformazione indotta sullo spazio dei parametri da g_a è $\bar{g}(\theta) = \theta + a$.

2. Dunque uno stimatore $W(\underline{X})$ di θ è invariante rispetto a $\mathcal{G}$ se per ogni a, $\underline{x}$ e θ $W(g_a(\underline{x})) = g_a(W(\underline{x}))$ equivalentemente se $W(\underline{X} + a) = W(\underline{X}) + a$. Ne segue che se uno stimatore è invariante per $\mathcal{G}$ allora $\mathrm{E}_\theta(W(\underline{X} + a)) = \mathrm{E}_\theta(W(\underline{X})) + a$ e $\mathrm{Var}_\theta(W(\underline{X} + a)) = \mathrm{Var}_\theta(W(\underline{X}))$.

3. Oltre all'invarianza rispetto a $\mathcal{G}$ richiediamo che lo stimatore di θ sia corretto per θ e quindi imponiamo
$$\begin{aligned} \theta &= \mathrm{E}_\theta(W(\underline{X})) = \mathrm{E}_\theta(W(\underline{X} + a)) - a && \text{per invarianza} \\ &= \mathrm{E}_\theta(W(\underline{X} - \theta)) + \theta && \text{scegliendo } a = -\theta \end{aligned}$$
da cui semplificando si ha $\mathrm{E}_\theta(W(\underline{X} - \theta)) = 0$. Ricordando che la legge di $T = \underline{X} - \theta$ non dipende da θ (modello di posizione) si ha $\mathrm{E}(W(T)) = 0$ per ogni θ.

4. Per esempio sia $W(\underline{X}) = \bar{X}$ e si supponga che $\mathrm{E}_0(W(X_1)) = 0$. Allora W soddisfa entrambe le condizioni dei punti precedenti.

Esercizio 3.33 Siano $X_1, \ldots, X_N$ i.i.d. $\frac{1}{\sigma} f(\frac{x-\theta}{\sigma})$ con $\theta, x \in \mathbb{R}$ e $\sigma > 0$. Sia $\mathcal{G}$ come nell'esempio precedente. Stimare σ^2 con uno stimatore invariante rispetto a $\mathcal{G}$. Dedurre che $\mathcal{G}$ non ha effetto su σ^2.

Esempio 3.34 Sia $\underline{X} \sim \mathcal{N}_n(\underline{\mu}, \sigma^2 I_n)$ con $\mu = (\mu, \ldots, \mu)$ e $\mu \in \mathbb{R}$ e $\mathcal{G}$ come nell'esempio precedente. Si ha che $T_1(\underline{X}) = 0.9\bar{X}$ non è invariante per $\mathcal{G}$, mentre $T_2(\underline{X}) = \bar{X}$ e $T_3(\underline{X}) = \frac{X_{(1)} + X_{(n)}}{2}$ sono invarianti per $\mathcal{G}$.

Esempio 3.35 Segue dal teorema di invarianza per stimatori di massima verosimiglianza che questi sono invarianti rispetto a ogni $\mathcal{G}$.

Esempio 3.36 (Pace e Salvan Esempio 7.25) Siano $X_1, \ldots, X_n$ un campione i.i.d. $\mathrm{Unif}(]\theta, \theta + 1[)$ con $\theta \in]0, 1[$. Sia $T(X_1, \ldots, X_n) = (X_{(1)}, X_{(n)})$ la statistica minimale. Sia $\mathcal{G}$ il gruppo delle traslazioni come negli esempi precedenti.

1. Verificare che il modello statistico indicato è invariante per $\mathcal{G}$.

2. Per ogni $g \in \mathcal{G}$ individuare $\bar{g}$.

3. Dimostrare che T è invariante rispetto a $\mathcal{G}$.

4 Tecniche per valutare uno stimatore

4.1 Il linguaggio delle decisioni statistiche, stimatori corretti, errore quadratico medio e stimatori consistenti

La teoria delle decisioni fornisce un linguaggio appropriato per discutere sulla bontà degli stimatori statistici. Ne diamo un brevissimo accenno. Sia A un insieme di decisioni su una certa questione, e.g.

1. $A = \{a_1, \ldots, a_v\}$, per esempio prendo o non prendo l'ombrello
2. $A = [0, 1]$, se vogliamo stimare la probabilità che oggi piova oppure $\mathrm{P}(X_1 < 0)$
3. $A = \{\text{stimatori di } \theta \in \Theta\}$, se Θ è lo spazio dei parametri di un certo modello statistico.

Sotto 3. si definisce *costo* della decisione a, quando la decisione corretta è θ oppure sotto il modello probabilistico individuato da θ, un numero reale (usualmente positivo) indicato con $C(\theta, a)$. Per convenzione più $C > 0$ è alto, più la decisione a è costosa. Se la decisione riguarda la scelta di un modello probabilistico all'interno di un modello statistico non necessariamente parametrico $\mathcal{F}$, si scrive $C(a, F)$ con $F \in \mathcal{F}$.
Più specificamente, per un modello statistico parametrico indicizzato da $\theta \in \Theta$ e per un campione statistico definito su $(\Omega, \mathcal{A}, \mathrm{P})$, la *decisione* è una variabile aleatoria

$$\begin{aligned} U : \Omega &\longrightarrow A \\ \omega &\longmapsto a(\omega) \end{aligned}$$

ed il costo dipende da θ e da ω tramite U, $C(\theta, a) = C(\theta, U(\omega))$.
Il *rischio* associato a θ e U è definito come costo medio

$$R(\theta, U) = \mathrm{E}_\theta(C(\theta, U)) = \int_\Omega C(\theta, U(\omega))\, dP_\theta(\omega)$$

Chiaramente sia rischio sia costo si possono formulare in termini di una certa $g(\theta)$ funzione del parametro anzichè di θ.

Esempio 4.1 Per $\theta \in \mathbb{R}$, $U(\omega) \in \mathbb{R}$ per ogni $\omega \in \Omega$ si possono considerare

1. il costo L^1 definito come $C_1(\theta, U(\omega)) = |\theta - U(\omega)|$ ed il costo L^2: $C_2(\theta, U(\omega)) = (\theta - U(\omega))^2$.
2. La funzione rischio associata a C_2 è detta *errore quadratico medio (o mean square error)* ed è definita come $R_2(\theta, U(\omega)) = \mathrm{MSE}(\theta, U(\omega)) = \mathrm{E}_\theta(U - \theta)^2$. Mentre la funzione rischio associata al costo L^1 è maggiore od uguale a $\mathrm{Bias}(\theta, U) = |\,\mathrm{E}_\theta(U) - \theta|$.

Esercizio 4.2 Sia U uno stimatore univariato di $\theta \in \mathbb{R}$ quadrato integrabile

$$\begin{aligned} \underline{\mathrm{MSE}(\theta, U)} &= \mathrm{E}_\theta\left((\theta - \mathrm{E}_\theta(U))^2 + (\mathrm{E}_\theta(U) - U)^2 + 2(\theta - \mathrm{E}_\theta(U))(\mathrm{E}_\theta(U) - U)\right) \\ &= (\theta - \mathrm{E}_\theta(U))^2 + \mathrm{Var}_\theta(U) + 2(\theta - \mathrm{E}_\theta(U))^2 0 \\ &= \underline{\mathrm{Bias}(U, \theta)^2 + \mathrm{Var}_\theta(U)} \end{aligned}$$

In particolare se $\theta = \mathrm{E}_\theta(U)$ allora $\mathrm{MSE}(\theta, U) = \mathrm{Var}_\theta(U)$ cioè la varianza di stimatori corretti di θ è l'errore quadratico medio. Per esercizio si determini l'analoga relazione per $U, \theta \in \mathbb{R}^p$.

Esercizio 4.3 Generalizzare gli Esempi 4.1 e 4.2 al caso $\theta, U(\omega) \in \mathbb{R}^p$.

Sia $\mathcal{D}$ un insieme di decisioni aleatorie (di statistiche) e $U, V \in \mathcal{D}$. Sia C un costo e R il rischio associato. Si dice che

1. U è <u>preferibile</u> a V se $\mathrm{R}(\theta, U) \leq \mathrm{R}(\theta, V)$ per ogni θ
2. U è <u>migliore</u> a V se è preferibile e esiste $\bar{\theta}$ t.c. $\mathrm{R}(\bar{\theta}, U) < \mathrm{R}(\theta, V)$

3. U è <u>ammissibile</u> in $\mathcal{D}$ se non esiste in $\mathcal{D}$ una decisione migliore di U

4. U è <u>ottimale</u> in $\mathcal{D}$ se è preferibile a ogni altro $V \in \mathcal{D} \setminus \{U\}$.

Non sempre esistono stimatori ottimali. Scopo di questi paragrafi è dimostrare, sotto opportune ipotesi, che uno stimatore non distorto ottimale, rispetto al rischio quadratico medio, è quello di varianza minima.
Un'altra strategia, che non approfondiamo, per scegliere in $\mathcal{D}$ è la seguente: $U \in \mathcal{D}$ è <u>minmax</u> per θ se

$$\sup_{\theta} \mathrm{R}(\theta, U) = \inf_{V \in \mathcal{D}} \sup_{\theta \in \Theta} \mathrm{R}(\theta, V)$$

(la migliore decisione per controllare "il peggior caso".... il meglio del peggio: Marcello Marchesi).

Esercizio 4.4 Siano $X_1, \dots, X_n$ i.i.d. Bernoulli(p) con $p \in]0,1[$, $A = [0,1]$ e $C(p,a) = \dfrac{(p-a)^2}{p(1-p)}$. Si disegni il grafico di C in funzione di p e si noti che C è maggiore per $p \sim 0,1$, ovvero penalizza di più gli errori se $p \sim 0,1$ che per $p \sim a$.

Esempio 4.5 (cont. Esercizio 4.4) Si scelga la decisione statistica $\bar{X}$. Allora

$$\mathrm{R}(p, \bar{X}) = \mathrm{E}_p\left(\frac{(p-\bar{X})^2}{p(1-p)}\right) = \frac{1}{p(1-p)}\,\mathrm{E}_p\left((p-\bar{X})^2\right)$$

notare che $\mathrm{E}_p(\bar{X}) = p$

$$= \frac{1}{p(1-p)}\,\mathrm{E}_p\left((\mathrm{E}_p(\bar{X}) - \bar{X})^2\right) = \frac{1}{p(1-p)}\,\mathrm{Var}_p(\bar{X}) = \frac{1}{p(1-p)}\frac{p(1-p)}{n} = \frac{1}{n} \longrightarrow 0 \qquad \text{se } n \mapsto +\infty$$

Esercizio 4.6 Sia $\Theta = \mathbb{R}$ e si considerino i costi

1. $C(\theta, a) = |\theta - a|$

2. $C(\theta, a) = (\theta - a)^2$

3. $C(\theta, a) = \begin{cases} (\theta-a)^2 & \text{se } a \leq \theta \\ 10(a-\theta)^2 & \text{se } a > \theta \end{cases}$

4. $C(\theta, a) = \dfrac{(a-\theta)^2}{1+|\theta|}$.

Si noti che i costi 1. e 2. sono tanto più alti quanto più a è distante da θ e che il costo 3. penalizza maggiormente sovrastime di θ che sottostime. Calcolare i rischi associati a questi costi e per $\bar{X}$ per il modello nell'Esercizio 4.4. Indicare sotto quali ipotesi sono ben definiti.

4.2 Teorema di Rao-Blackwell

Esempio 4.7 (modello gerarchico a due livelli) Siano X, Y variabili aleatorie su $(\Omega, \mathcal{A}, \mathrm{P})$ a valori reali e quadrato integrabili, cioè con valore atteso e varianza finiti e a valori reali. Allora

$$\mathrm{Var}(X) = \mathrm{E}(\mathrm{Var}(X|Y)) + \mathrm{Var}(\mathrm{E}(X|Y))$$

infatti

$$\begin{aligned}\mathrm{Var}(X) &= \mathrm{E}((X-\mathrm{E}(X))^2) \\ &= \mathrm{E}\left((X-\mathrm{E}(X|Y))^2\right) + \mathrm{E}\left((\mathrm{E}(X|Y)-\mathrm{E}(X))^2\right) + 2\,\mathrm{E}\left((X-\mathrm{E}(X|Y))(\mathrm{E}(X|Y)-\mathrm{E}(X))\right)\end{aligned}$$

Ora $\mathrm{Var}(\mathrm{E}(X|Y)) = \mathrm{E}\left((\mathrm{E}(X|Y) - \mathrm{E}\left(\mathrm{E}(X|Y)\right))^2\right) = \mathrm{E}\left((\mathrm{E}(X|Y) - \mathrm{E}(X))^2\right)$ e

$$\mathrm{E}(\mathrm{Var}(X|Y)) = \mathrm{E}\left(\mathrm{E}\left((X-\mathrm{E}(X|Y))^2|Y\right)\right) = \mathrm{E}\left((X-\mathrm{E}(X|Y))^2\right)$$

mentre il doppio prodotto vale zero. Infatti

$$
\begin{aligned}
\mathrm{E}\left((X-\mathrm{E}(X|Y))(\mathrm{E}(X|Y)-\mathrm{E}(X))\right) &= \mathrm{E}\left(\mathrm{E}\left((\cdots)(\cdots)\right)|Y\right) \\
&= \mathrm{E}\left((\mathrm{E}(X|Y)-\mathrm{E}(X))\,\mathrm{E}\left(X-\mathrm{E}(X|Y)|Y\right)\right) \\
&= \mathrm{E}\left((\mathrm{E}(X|Y)-\mathrm{E}(X))\left(\mathrm{E}(X|Y)-\mathrm{E}(\mathrm{E}(X|Y)|Y)\right)\right) \\
&= \mathrm{E}\left((\cdots)(0)\right)) = 0
\end{aligned}
$$

Ne segue che $\mathrm{Var}(X) \geq \mathrm{Var}(\mathrm{E}(X|Y))$.

Il teorema di Rao-Blackwell generalizza a stimatori la disuguaglianza precedente e relaziona sufficienza e stimatori di rischio minore. Siano $\theta \in \Theta \subseteq \mathbb{R}^p$ e $g(\theta) \in \mathbb{R}$ una funzione del parametro. Si definiscano

$\mathcal{D}_2 = \{U \text{ stimatore di } g(\theta) \in \mathbb{R} \text{ t.c. } \mathrm{Var}_\theta(U) < +\infty \text{ per ogni } \theta \in \Theta\}$

$\mathcal{D}_{2c} = \{U \in \mathcal{D}_2 : \mathrm{Bias}(g(\theta), U) = 0 \text{ per ogni } \theta \in \Theta\}$.

Theorem 4.8 (di Rao-Blackwell) *Siano $U \in \mathcal{D}_2$ e T una statistica sufficiente per θ. Allora*

1. $\mathrm{E}_\theta(U|T)$ *è stimatore di* $g(\theta)$.
2. *Se U è corretto per $g(\theta)$ allora anche $\mathrm{E}(U|T)$ lo è.*
3. $\mathrm{MSE}(U, g(\theta)) \geq \mathrm{MSE}(\mathrm{E}(U|T), g(\theta))$.

Nel teorema si considerano la funzione costo $(a - g(\theta))^2$ e il rischio $\mathrm{E}_\theta\left((U-g(\theta))^2\right)$.
Proof. Si noti che T è sufficiente anche per $g(\theta)$.

1. Poiché T è sufficiente, per il Teorema 2.11 di Neyman-Fisher la legge condizionata a T di U non dipende da θ e quindi $\mathrm{E}(U|T)$ non dipende da θ ed è funzione dei dati solo tramite T.
2. E' proprietà della speranza condizionata che $\mathrm{E}_\theta\left(\mathrm{E}(U|T)\right) = \mathrm{E}_\theta(U) = g(\theta)$.
3. Sia $U' = \mathrm{E}(U|T)$ allora

$$
\begin{aligned}
\mathrm{MSE}_\theta(U', g(\theta)) &= \mathrm{E}_\theta((U'-g(\theta))^2) \\
&= \mathrm{E}_\theta((\mathrm{E}(U|T)-g(\theta))^2) && g(\theta) \text{ è costante per l'integrazione} \\
&= \mathrm{E}_\theta((\mathrm{E}(U-g(\theta)|T))^2) && \text{per la disuguaglianza di Jensen} \\
&\leq \mathrm{E}_\theta(\mathrm{E}\left(U-g(\theta)\right)^2|T) = \mathrm{E}_\theta((U-g(\theta))^2) && \text{per la proprietà della torre} \\
&= \mathrm{MSE}_\theta(U, g(\theta))
\end{aligned}
$$

■

Esercizio 4.9 Sia $U = f(T)$. Quanto vale $\mathrm{MSE}_\theta(\mathrm{E}_\theta(U|T), \theta)$?

Esempio 4.10 (necessità della sufficienza di T) Siano X_1 e X_2 copie indipendenti di legge $N(\theta, 1)$ con $\theta \in \mathbb{R}$. Sia $U = \bar{X} = (X_1+X_2)/2$ e $T = X_1$. Si ha che T non è sufficiente per θ ed inoltre $\mathrm{E}(U|X_1) = ... = \dfrac{X_1+\theta}{2}$ non può essere stimatore, non essendo statistica.

In pratica E_θ(U|T) è difficile da calcolare, trovare una statistica sufficiente è (relativamente) facile. Tanto vale partire da stimatori che sono funzioni di statistiche sufficienti. Alla luce di questo esempio si riveda il Teorema 2.11.

Esempio 4.11 Siano $X_1, \dots, X_n$ i.i.d. Binomial(k, θ) con k noto. Stimare $g(\theta) = \mathrm{P}_\theta(X_n = 1) = k\theta(1-\theta)^{k-1}$. Si noti che $T = \sum_{i=1}^n X_i \sim$ Binomial(kn, θ) è statistica sufficiente per θ. E' anche completa (segue dalla teoria dei modelli di classe esponenziale). Ma non è corretta per $k\theta(1-\theta)^{k-1}$ (verificarlo). Cerchiamo dunque uno stimatore corretto

$$U = h(X_1) = \begin{cases} 1 & \text{se } X_1 = 1 \\ 0 & \text{altrimenti} \end{cases}$$

Per il teorema di Rao-Blackwell $V = \mathrm{E}_\theta\left(h(X_1)|\sum_{i=1}^n X_i\right)$ ha rischio quadratico inferiore a quello di $h(X_1)$. Ovviamente $\mathrm{E}_\theta(V) = \mathrm{E}_\theta(h(X_1)) = k\theta(1-\theta)^{k-1}$. Eccezionalmente sappiamo calcolare V

$$\begin{aligned}
\mathrm{E}\left(h(X_1)|\sum_{i=1}^n X_i = t\right) = \mathrm{P}\left(X_1 = 1|\sum_{i=1}^n X_i = t\right) &= \frac{\mathrm{P}_\theta\left(X_1 = 1, \sum_{i=1}^n X_i = t\right)}{\mathrm{P}_\theta\left(\sum_{i=1}^n X_i = t\right)} \\
&= \frac{\mathrm{P}_\theta\left(X_1 = 1, \sum_{i=2}^n X_i = t-1\right)}{\mathrm{P}_\theta\left(\sum_{i=1}^n X_i = t\right)} \\
&= \frac{\mathrm{P}_\theta\left(X_1 = 1\right)\mathrm{P}_\theta\left(\sum_{i=2}^n X_i = t-1\right)}{\mathrm{P}_\theta\left(\sum_{i=1}^n X_i = t\right)} && \text{per l'indipendenza delle } X_i \\
&= \frac{\binom{k}{1}\binom{k(n-1)}{t-1}}{\binom{kn}{t}}
\end{aligned}$$

Osservazioni 4.12 In generale stimatori ottenuti con il metodo dei momenti non sono funzioni di statistiche sufficienti e quindi possono essere migliorati nello spirito del teorema di Rao-Blackwell condizionando su statistiche sufficienti, mentre gli stimatori di MV sono già funzione di statistiche sufficienti.

Theorem 4.13 (di Lehmann-Scheffé) *Alle ipotesi del teorema di Rao-Blackwell si aggiunga T completa e U corretta per $g(\theta)$. Allora*

$$\mathrm{Var}_\theta\left(\mathrm{E}(U|T)\right) \le \mathrm{Var}_\theta\left(V\right)$$

per ogni $V \in \mathcal{D}_{2,c}$.

Proof. Abbiamo già visto che $\mathrm{E}(U|T)$ è stimatore corretto per $g(\theta)$ ed è funzione di T. Sia V un altro stimatore corretto di $g(\theta)$. Per il teorema di Rao-Blackwell V può essere scelto funzione della statistica sufficiente T. Si ha

$$\mathrm{E}_\theta(V) = \mathrm{E}_\theta\left(\mathrm{E}(U|T)\right) = g(\theta)$$

da cui

$$\mathrm{E}_\theta\left(V - \mathrm{E}(U|T)\right) = 0$$

Dal fatto che T è completa e $V - \mathrm{E}(U|T)$ è funzione di T, segue che $V = \mathrm{E}(U|T)$ quasi certamente in $(\Omega, \mathcal{A}, \mathrm{P})$. ■

Osservazioni 4.14

1. Il teorema di Lehmann-Scheffé afferma che $\mathrm{E}_\theta(U|T)$ ha varianza minima in $\mathcal{D}_{2,c}$, cioè è ottimale nella classe di stimatori con varianza finita.
2. Collega sufficienza, correttezza e completezza.
3. $\mathrm{E}(U|T)$ è detto UMVUE: uniform, minimum variance, unbiased estimator di $g(\theta)$.
4. Il Teorema di Lehmann-Scheffé garantisce l'unicità degli stimatori UMVUE, cioè uno stimatore non distorto e basato su una statistica sufficiente e completa è unico a meno di insiemi di misura P zero.

Esercizio 4.15

1. Sia T una statistica completa e sufficiente per θ e si consideri $f(T)$. Si dimostri che $f(T)$ è l'unico stimatore UMVUE di $\mathrm{E}_\theta(f(T))$ (si ricordi che $\mathrm{E}(f(T)g|T) = f(T)\,\mathrm{E}(g|T)$.)
2. Sia $X_1, \ldots, X_n$ un campione casuale $\mathcal{N}(\mu, \sigma^2)$ con σ^2 noto e $\mu \in \mathbb{R}$. Poiché $\bar{X}$ è statistica sufficiente e completa per μ e $\bar{\mathrm{X}} = \mu$, $\bar{X}$ è UMVUE per μ.
3. Lo stimatore $\bar{X}^2$ non è UMVUE per μ^2 ed è funzione di statistica sufficiente e completa. Il suo valore atteso è $\sigma^2/n + \mu^2$. Si deduce che $\bar{X}^2 - \sigma^2/n$ è UMVUE per μ^2. Notare che $(\bar{X})^2$ è stimatore di massima verosimiglianza per μ^2.
4. (UMVUE e modelli di classe esponenziale). Nel modello di classe esponenziale

$$h(t)\exp\left(\langle\theta, T\rangle - \psi(\theta)\right)$$

la statistica $T = \mathrm{E}(T|T)$ è sufficiente e completa per il parametro naturale, è stimatore corretto di $\nabla_\theta\psi(\theta)$, quindi è UMVUE per $\nabla_\theta\psi(\theta)$ (è anche stimatore di massima verosimiglianza).

Esercizio 4.16 (Necessità dell'ipotesi di completezza nel teorema di Lehmann-Fisher) Si consideri una sola osservazione da $X \sim \text{Uniform}(]\theta, \theta+1[)$ $(n = 1)$. Si dimostri che

1. $X - \frac{1}{2}$ è stimatore corretto di θ e $\text{Var}_\theta(X - \frac{1}{2}) = \frac{1}{12}$.
2. $X - \frac{1}{2}$ è sufficiente per θ.
3. Sia $g(x) = \sin 2\pi x$. Verificare che $\int_\theta^{\theta+1} g(x)\, dx = 0$, ovvero che $\sin 2\pi X$ è stimatore corretto di zero, si parla anche di rumore aleatorio. In particolare ne segue che $X - \frac{1}{2}$ non è completa per θ.
4. $T = X - \frac{1}{2} + \frac{\sin 2\pi X}{2\pi}$ è stimatore corretto di θ.
5. Verificare, e.g. integrando per parti, che $\text{Cov}_\theta(X - \frac{1}{2}, \sin 2\pi X) = -\dfrac{\cos 2\pi\theta}{2\pi}$. Quindi $X - \frac{1}{2}$ è correlato con $\sin 2\pi X$.
6. Verificare che $\text{Var}_\theta(T) = 0.071 < \frac{1}{12} = \text{Var}_\theta(X - \frac{1}{2})$.
7. In generale si può dimostrare che W è stimatore UMVUE (del suo valore atteso) se e solo se non è correlato con stimatori corretti di zero (cf. Casella Berger Teorema 7.3.4, edizione del 1990).

Esercizio 4.17 Sia un campione di taglia uno $X \sim \text{Poisson}(\lambda)$ con $\lambda > 0$. La statistica X è statistica sufficiente e completa per λ. Dal fatto che

$$\text{E}_\lambda\left((-1)^X\right) = \sum_{k=0}^{+\infty} (-1)^k \frac{e^{-\lambda}\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{+\infty} \frac{(-\lambda)^k}{k!} = e^{-2\lambda}$$

si deduce che lo stimatore $T = (-1)^X$ è UMVUE per $e^{-2\lambda}$. Ma T può assumere valori negativi, mentre $e^{-2\lambda} > 0$. Lo stimatore di massima verosimiglianza di $e^{-2\lambda}$ sembra "preferibile".

Esercizio 4.18 Si riveda alla luce del teorema di Lehmann-Scheffé l'ultimo punto dell'Esercizio 3.14

4.3 Informazione secondo Fisher

In questo paragrafo si determina un limite inferiore alla varianza di statistiche. Siano $\underline{X}$ un campione statistico su uno spazio di probabilità $(\Omega, \mathcal{A}, \text{P})$ a valori in $\mathcal{X} \subseteq \mathbb{R}^n$ e un modello statistico parametrico $\mathcal{F} = \{f_{\underline{X}}(\cdot, \theta) : \theta \in \Theta \subseteq \mathbb{R}^p\}$ con $f_{\underline{X}}$ densità di probabilità. Siano quindi $\mathcal{L}(\theta, \underline{x}) = f_{\underline{X}}(\underline{x}, \theta)$ la funzione di verosimiglianza e $l(\theta, \underline{x}) = \log \mathcal{L}(\theta, \underline{x})$ la log-verosimiglianza, se esiste.
Supponiamo che il modello soddisfi le seguenti ipotesi, dette *condizioni di regolarità di Cramér-Rao*:

1. sia regolare
2. per ogni $i = 1, \ldots, p$, esista $\dfrac{\partial}{\partial \theta_i} l(\theta; \underline{x})$ per ogni $\theta \in \Theta$ e per quasi ogni $\underline{x} \in \mathcal{X}$
3. $\displaystyle\int_{\mathcal{X}} \frac{\partial}{\partial \theta_i} \mathcal{L}(\theta; \underline{x})\, d\underline{x} = \frac{\partial}{\partial \theta_i} \int_{\mathcal{X}} \mathcal{L}(\theta; \underline{x})\, d\underline{x}$.

Osservazioni 4.19

1. Notare che il punto 2. richiede che l sia funzione continua in θ_i per ogni i.
2. Dal punto 3. segue che il valor atteso di $\dfrac{\partial l(\theta; \underline{X})}{\partial \theta_i}$ è nullo, infatti

$$\text{E}_\theta\left(\frac{\partial \log \mathcal{L}(\theta; \underline{X})}{\partial \theta_i}\right) = \int_{\mathcal{X}} \frac{\partial}{\partial \theta_i} \mathcal{L}(\theta; \underline{x})\, d\underline{x}$$

3. Il gradiente di l in θ, $\left(\dfrac{\partial l(\theta; \underline{X})}{\partial \theta_i}\right)_{i=1,\ldots,p}$, è chiamato *score function*.

Osservazioni 4.20

1. Calcolare, se possibile, $\int_{\mathcal{X}} \frac{d}{d\theta} \mathcal{L}(\theta; \underline{x})\, d\underline{x}$ e $\frac{d}{d\theta} \int_{\mathcal{X}} \mathcal{L}(\theta; \underline{x})\, d\underline{x}$ per un campione casuale Unif$(]0, \theta[)$.
2. Scrivere le condizioni di regolarità di Cramér Rao per un campione discreto.

Definizione 4.21 *Sotto le condizioni di regolarità di Cramér Rao per $\theta \in \Theta$ e $i, j = 1, \ldots, p$ si definisce*

$$I(\theta)_{ij} = \mathrm{E}_\theta \left(\frac{\partial l(\theta; \underline{X})}{\partial \theta_i} \frac{\partial l(\theta; \underline{X})}{\partial \theta_j} \right)$$
$$I(\theta) = [I(\theta)_{ij}]_{i,j=1,\ldots,p}$$

quest'ultima detta matrice di informazione (attesa) di/secondo Fisher.

Osservazioni 4.22

1. $I(\theta)_{ij}$ non è una statistica.
2. $I(\theta)_{ij} = \mathrm{Cov}_\theta \left(\frac{\partial l(\theta; \underline{X})}{\partial \theta_i}, \frac{\partial l(\theta; \underline{X})}{\partial \theta_j} \right)$.
3. I è matrice simmetrica di dimensione p.
4. Se $\theta \in \mathbb{R}$ allora $I(\theta) = \mathrm{E}_\theta \left(\left(\frac{dl(\theta; \underline{X})}{d\theta} \right)^2 \right) = \mathrm{Var}_\theta \left(\frac{dl(\theta; \underline{X})}{d\theta} \right)$, ovvero l'informazione di Fisher è la varianza della funzione score.
5. L'informazione di Fisher osservata è utile in pratica: sia $\hat{\theta}$ stimatore di massima verosimiglianza di θ, l'informazione di Fisher osservata è definita come $I(\hat{\theta})_{ij} = \frac{\partial}{\partial \theta_i} l(\theta; \underline{x})_{|\theta=\hat{\theta}} \frac{\partial}{\partial \theta_j} l(\theta; \underline{x})_{|\theta=\hat{\theta}}$.

Osservazioni 4.23 (Informazione di Fisher in diverse parametrizzazioni) Siano $\theta \in \mathbb{R}$ e $g(\theta)$ una riparametrizzazione, regolare, del modello. Allora

$$\frac{dl}{d\theta} = \frac{dl}{dg(\theta)} \frac{dg(\theta)}{d\theta}$$

e quindi

$$I(\theta) = \mathrm{E}_\theta \left(\left(\frac{dl}{dg(\theta)} \right)^2 \left(\frac{dg(\theta)}{d\theta} \right)^2 \right) = \left(\frac{dg(\theta)}{d\theta} \right)^2 I(g(\theta))$$

Esercizio 4.24

1. Quali sono le ipotesi di regolarità su g necessarie per l'Osservazione 4.23?
2. Generalizzare l'Osservazione 4.23 per θ multidimensionale.

Esempio 4.25

1. Sia $X_1, \ldots, X_n$ un campione casuale Bernoulli(p) con $p \in]0, 1[$. Si ha $l(p; \underline{x}) = s_n \log p + (n - s_n) \log(1 - p)$ con $S_n \sim$ Binomial(n, p), da cui segue
$$\frac{dl}{dp} = \frac{S_n}{p} - \frac{n - S_n}{1 - p} = \frac{S_n - np}{p(1 - p)}$$
$$I(p) = \mathrm{E}_p \left(\left(\frac{dl}{dp} \right)^2 \right) = \frac{\mathrm{E}_p \left((S_n - \mathrm{E}_p(S_n))^2 \right)}{p^2(1 - p)^2} = \frac{np(1 - p)}{p^2(1 - p)^2} = \frac{n}{p(1 - p)}$$
2. Calcolare l'informazione di Fisher rispetto al parametro $\theta = p/(1 - p)$.
3. Calcolare l'informazione di Fisher rispetto al parametro $\theta = \log(p/(1 - p))$.

Esempio 4.26 (Informazione di Fisher per modelli di classe esponenziale) Se a meno di costante additiva $l(\theta; T(\underline{x})) = \langle \theta, T(\underline{x}) \rangle - \psi(\theta)$ con $\theta \in \mathbb{R}^p$, si ha

$$\begin{aligned}
\frac{\partial l}{\partial \theta_i} &= T_i - \frac{\partial \psi}{\partial \theta_i} = T_i - \mathrm{E}_\theta(T_i(\underline{X})) \\
I(\theta)_{ij} &= \mathrm{E}_\theta \left((T_i - \frac{\partial \psi}{\partial \theta_i})(T_j - \frac{\partial \psi}{\partial \theta_j}) \right) = \mathrm{Cov}_\theta(T_i, T_j) = \frac{\partial^2 \psi}{\partial \theta_i \partial \theta_j}
\end{aligned}$$

Nel modello di classe esponenziale la matrice di informazione attesa di Fisher è la matrice Hessiana della funzione dei cumulanti. Rifare il punto 3. dell'Esercizio 4.25.

Lemma 4.27 (Calcolo dell'informazione di Fisher) *Nell'ipotesi in cui sia possibile scambiare derivata seconda in θ con integrazione in $d\underline{x}$, vale*

$$I(\theta)_{ij} = -\,\mathrm{E}_\theta \left(\frac{\partial^2}{\partial \theta_i \partial \theta_j} \log \mathcal{L}(\theta; \underline{X}) \right)$$

Osservazioni 4.28 Il Lemma 4.27 esprime l'informazione di Fisher come la curvatura media locale della funzione log-verosimiglianza. Si noti che per modelli di classe esponenziale è descritta come la curvatura locale della funzione dei cumulanti.

Proof.

$$\begin{aligned}
\frac{\partial^2}{\partial \theta_i \partial \theta_j} \log \mathcal{L}(\theta; \underline{x}) &= \frac{\partial}{\partial \theta_i} \left(\frac{1}{\mathcal{L}} \frac{\partial}{\partial \theta_j} \mathcal{L}(\theta; \underline{X}) \right) \\
&= \frac{1}{\mathcal{L}} \frac{\partial^2}{\partial \theta_i \partial \theta_j} \mathcal{L}(\theta; \underline{x}) - \frac{1}{\mathcal{L}^2} \frac{\partial}{\partial \theta_i} \mathcal{L}(\theta; \underline{X}) \frac{\partial}{\partial \theta_j} \mathcal{L}(\theta; \underline{X}) \\
&= \frac{1}{\mathcal{L}} \frac{\partial^2}{\partial \theta_i \partial \theta_j} \mathcal{L}(\theta; \underline{x}) - \frac{\partial}{\partial \theta_i} l(\theta; \underline{X}) \frac{\partial}{\partial \theta_j} l(\theta; \underline{X})
\end{aligned}$$

Per concludere è sufficiente verificare che $\mathrm{E}_\theta \left(\frac{1}{\mathcal{L}} \frac{\partial^2}{\partial \theta_i \partial \theta_j} \mathcal{L}(\theta; \underline{x}) \right) = 0$. Si ha

$$\begin{aligned}
\mathrm{E}_\theta \left(\frac{1}{\mathcal{L}} \frac{\partial^2}{\partial \theta_i \partial \theta_j} \mathcal{L}(\theta; \underline{x}) \right) &= \int_{\mathcal{X}} \frac{1}{\mathcal{L}} \frac{\partial^2}{\partial \theta_i \partial \theta_j} \mathcal{L}(\theta; \underline{x}) f_{\underline{X}}(\underline{x}; \theta)\, d\underline{x} \\
&= \frac{\partial^2}{\partial \theta_i \partial \theta_j} \int_{\mathcal{X}} \mathcal{L}(\theta; \underline{x})\, d\underline{x} = 0 \qquad \text{qui si usa l'ipotesi aggiuntiva}
\end{aligned}$$

■

Esercizio 4.29 Verificare il lemma 4.27 per un campione casuale Bernoulliano.

Lemma 4.30 (Campione casuale) *Siano $\underline{X}_1, \ldots, \underline{X}_n$ i.i.d. e valgano le ipotesi del Lemma 4.27. Sia*

$$I^1(\theta)_{ij} = \mathrm{E}_\theta \left(\frac{\partial}{\partial \theta_i} l(\theta; \underline{X}_1) \frac{\partial}{\partial \theta_j} l(\theta; \underline{X}_1) \right)$$

Allora $I(\theta)_{ij} = nI^1(\theta)_{ij}$.

Proof. Poichè il campione è casuale si ha $\mathcal{L}(\theta; \underline{x}) = \prod_{i=1}^n f_{X_1}(x_i, \theta)$ e $l(\theta; \underline{x}) = \sum_{i=1}^n \log \mathcal{L}(\theta; x_i)$. Ora

$$\begin{aligned}
I(\theta)_{ij} &= -\,\mathrm{E}_\theta \left(\frac{\partial^2}{\partial \theta_i \partial \theta_j} \log \mathcal{L}(\theta; \underline{X}) \right) \\
&= -\,\mathrm{E}_\theta \left(\frac{\partial^2}{\partial \theta_i \partial \theta_j} \sum_{k=1}^n \log \mathcal{L}(\theta; \underline{X}_k) \right) \\
&= \sum_{k=1}^n (-1)\, \mathrm{E}_\theta \left(\frac{\partial^2}{\partial \theta_i \partial \theta_j} \log \mathcal{L}(\theta; \underline{X}_k) \right) = nI^1(\theta)_{ij}
\end{aligned}$$

■

Esercizio 4.31 Dimostrare il lemma precedente senza utilizzare l'ipotesi aggiuntiva del Lemma 4.27.

4.4 Limite inferiore di Cramér-Rao

Theorem 4.32 (Disuguaglianza di Cramér-Rao) *Sia $\underline{X}$ un campione a valori in $\mathcal{X}$. Sia $\mathcal{F}$ un modello statistico per $\underline{X}$ parametrizzato da $\theta \in \Theta$ con Θ intervallo di $\mathbb{R}$ che soddisfa le condizioni di regolarità di Cramér-Rao. Sia $T(\underline{X}) \in \mathbb{R}$ una statistica con varianza finita ed il cui valore atteso ammette derivata prima in θ. Allora*

$$\mathrm{Var}_\theta\left(T(\underline{X})\right) \geq \frac{\left(\dfrac{d}{d\theta}\mathrm{E}_\theta\left(T(\underline{X})\right)\right)^2}{I(\theta)} \geq 0$$

Proof. Facciamo la dimostrazione nel caso continuo. Per la terza condizione di regolarità di Cramér-Rao si ha

$$\begin{aligned}\frac{d}{d\theta}\mathrm{E}_\theta\left(T(\underline{X})\right) &= \int_{\mathcal{X}} T(\underline{x})\frac{d}{d\theta}f_{\underline{X}}(\underline{x};\theta)\,d\underline{x} = \int_{\mathcal{X}} T(\underline{x})\frac{d}{d\theta}\log\mathcal{L}(\underline{x};\theta)\,f_{\underline{X}}(\underline{x};\theta)\,d\underline{x} = \mathrm{E}_\theta\left(T(\underline{X})\frac{d}{d\theta}\log\mathcal{L}(\underline{x};\theta)\right)\\ &= \mathrm{E}_\theta\left(T(\underline{X})\,\frac{d}{d\theta}\log\mathcal{L}(\underline{x};\theta)\right) - \mathrm{E}_\theta\left(T(\underline{X})\right)\mathrm{E}_\theta\left(\frac{d}{d\theta}\log\mathcal{L}(\underline{x};\theta)\right) = \mathrm{Cov}_\theta\left(T(\underline{X}),\frac{d}{d\theta}\log\mathcal{L}(\underline{x};\theta)\right)\end{aligned}$$

L'ultima uguaglianza segue dal fatto che il valor medio della funzione score è nullo (cf. Osservazione 4.19). Ricordiamo che per X e Y variabili aleatorie sullo stesso spazio di probabilità con momenti secondi finiti vale

$$0 \leq \frac{\mathrm{Cov}(X,Y)^2}{\left(\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}\right)^2} \leq 1$$

e applichiamolo a $T(\underline{X})$ e $\dfrac{d}{d\theta}\log\mathcal{L}(\underline{X};\theta)$

$$\begin{aligned}0 &\leq \left(\frac{d}{d\theta}\mathrm{E}_\theta\left(T(\underline{X})\right)\right)^2 = \left(\mathrm{Cov}_\theta\left(T(\underline{X}),\frac{d}{d\theta}\log\mathcal{L}(\underline{x};\theta)\right)\right)^2 \leq \mathrm{Var}_\theta(T)\,\mathrm{Var}_\theta\left(\frac{dl}{d\theta}\right)\\ &= \mathrm{Var}_\theta(T)\,\mathrm{E}_\theta\left(\left(\frac{dl}{d\theta}\right)^2\right) = \mathrm{Var}_\theta(T)I(\theta)\end{aligned}$$

Notare l'utilizzo delle condizioni di regolarità di Cramér-Rao. ■

Osservazioni 4.33

1. Se T è corretto per θ allora il limite inferiore di Cramér-Rao è l'inverso dell'informazione di Fisher e dipende solo dal modello.

2. (a) Per campioni casuali di taglia n la disuguaglianza di Cramér-Rao diventa

   $$\mathrm{Var}_\theta\left(T(\underline{X})\right) \geq \frac{\left(\dfrac{d}{d\theta}\mathrm{E}_\theta\left(T(\underline{X})\right)\right)^2}{nI_1(\theta)}$$

   dove I_1 è l'informazione di Fisher contenuta in un (solo) elemento del campione.

   (b) Si semplifica ulteriormente a $\mathrm{Var}_\theta\left(T(\underline{X})\right) \geq \dfrac{1}{nI_1(\theta)}$ se T è corretto per θ.

   (c) Il limite inferiore di Cramér-Rao non dipende dalla parametrizzazione, infatti per $\theta \mapsto g(\theta)$ vale $\left(\dfrac{d}{d\theta}\mathrm{E}_\theta\left(T(\underline{X})\right)\right)^2 = \left(\dfrac{d}{dg(\theta)}\mathrm{E}_\theta\left(T(\underline{X})\right)\right)^2\left(\dfrac{dg(\theta)}{d\theta}\right)^2$ and $I(\theta) = I(g(\theta))\left(\dfrac{dg(\theta)}{d\theta}\right)^2$. Quindi il limite inferiore di Cramér-Rao è una caratteristica del modello statistico e non dipende dalla parametrizzazione del modello.

3. Il limite inferiore di Cramér-Rao è zero se e solo se $\mathrm{E}_\theta(T)$ è costante in θ (o $I(\theta) = +\infty$ ma questa è un'aspirazione: informazione infinita) ovvero lo stesso valore per ogni legge di probabilità nel modello statistico ovvero in media T non discrimina nel modello statistico ovvero il valor medio di T non aiuta a fare inferenza. Una statistica ancillare per θ è un esempio estremo, infatti neanche $\mathrm{Var}(T)$ dipende da θ.

4. Di una statistica T che raggiunge il limite inferiore di Cramér-Rao ed è usata come stimatore di $a(\theta)$, si dice che è stimatore *efficiente* per $a(\theta)$. Per stimatori efficienti vale quindi $\mathrm{Var}_\theta\left(T(\underline{X})\right) I(\theta) = \left(\frac{d}{d\theta}\mathrm{E}_\theta\left(T(\underline{X})\right)\right)^2$. Inoltre si noti che ogni stimatore è corretto per il suo valore atteso.
5. Uno stimatore efficiente e corretto è UMVUE, cioè ha varianza minima (possibile) ed è corretto.
6. Per uno stimatore efficiente (migliore in MSE) l'informazione di Fisher è l'inverso della varianza. Questo si riflette nell'espressione 'informazione di Fisher racchiusa nel campione'.

Esercizio 4.34 1. Scrivere e dimostrare la disuguaglianza di Cramér-Rao per campioni discreti. E' ragionevole/possibile considerarla se l'insieme dei parametri Θ è numerabile?

2. Esiste una relazione tra stimatori efficienti e statistiche sufficienti? E tra stimatori UMVUE e statistiche sufficienti?

Esercizio 4.35 (Disuguaglianza di Cramér-Rao multi-dimensionale) Per $\theta \in \mathbb{R}^p$ e $T \in \mathbb{R}^q$ con $q \leq p$, si ha $\mathrm{VarCov}(T) \in \mathbb{R}^{q\times q}$, $I(\theta) \in \mathbb{R}^{p\times p}$ e $\nabla_\theta \mathrm{E}_\theta \in \mathbb{R}^{q\times p}$. Analogamente al caso $p = q = 1$ si dimostra che

$$\mathrm{VarCov}_\theta(T) - \nabla_\theta \mathrm{E}_\theta(T)\, I(\theta)^{-1}\, \nabla_\theta \mathrm{E}_\theta(T)^t$$

è semi-definita positiva. Per $q = 1$ si semplifica a $\mathrm{Var}_\theta\left(T(\underline{X})\right) \geq \langle \nabla_\theta \mathrm{E}_\theta\left(T\right), I(\theta)^{-1}\nabla_\theta \mathrm{E}_\theta\left(T\right)\rangle$.

Corollario 4.36 Esiste al più un unico stimatore corretto ed efficiente di $g(\theta) \in \mathbb{R}$.

Proof. Stimatori efficienti e corretti sono UMVUE e questi ultimi sono unici. Presentiamo un'altra dimostrazione. Supponiamo per assurdo che T_1 e T_2 siano distinti stimatori corretti ed efficienti di $g(\theta)$ e consideriamo $T_3 = (T_1 + T_2)/2$. Vale che $\mathrm{E}(T_3) = g(\theta)$ e

$$\mathrm{Var}(T_3) = \frac{\mathrm{Var}(T_1) + \mathrm{Var}(T_2) + 2\,\mathrm{Cov}(T_1, T_2)}{4} \leq \frac{\mathrm{Var}(T_1) + \mathrm{Var}(T_2)}{4} + \frac{1}{2}\sqrt{\mathrm{Var}(T_1)}\sqrt{\mathrm{Var}(T_2)}$$

per la diseguaglianza di Cauchy-Schwartz. Dal fatto che T_1 e T_2 sono efficienti segue $\mathrm{Var}(T_1) = \mathrm{Var}(T_2)$ e questa varianza coincide con il limite inferiore di Cramér-Rao. In particolare la diseguaglianza precedente è un'uguaglianza e quindi $\mathrm{Cov}(T_1, T_2) = \mathrm{Var}(T_1)$. L'uguaglianza nella diseguaglianza di Cauchy-Schwartz si ha se, e solo se, esistono due funzioni a valori reali $a(\theta) \neq 0$ e $b(\theta)$ tali che $T_2 = aT_1 + b$. Da $\mathrm{Var}(T_1) = \mathrm{Cov}_\theta(T_1, aT_1 + b) = a\,\mathrm{Cov}(T_1, T_1) = a\,\mathrm{Var}(T_1)$ si deduce $a = 1$ e da $\mathrm{E}(T_1) = \mathrm{E}(T_2)$ si deduce $b = 0$. In conclusione si ha $T_1 = T_2$. ■

Esempio 4.37 (Necessità dell'ipotesi di modello regolare) 1. Sia $X \sim \mathrm{Unif}]0, \theta[$ con $\theta > 0$. Si ha

$$\mathcal{L}(\theta; x) = \frac{1}{\theta}(0 < x < \theta) \qquad \text{e} \qquad \mathcal{L}(\theta; x) = \frac{(0 < x < \theta)}{\theta}$$

Sebbene la log-verosimiglianza non sia definibile, per $x \in]0, \theta[$ si può scrivere $\frac{d}{d\theta} l(\theta; x) = -1/\theta$, la derivata non esiste per $x = 0, \theta$ ed è zero altrimenti. Quindi si potrebbe concludere $\mathrm{E}_\theta\left(\left(\frac{d}{d\theta} l(\theta; X)\right)^2\right) = 1/\theta^2$, l'informazione di Fisher per l'n-campione casuale dovrebbe essere $(\theta^2/n)^{-1}$ e il limite inferiore di Cramér-Rao per stimatori corretti di θ dovrebbe essere θ^2/n. Però esistono stimatori corretti di θ con varianza inferiore. Per esempio si consideri $V = \frac{n+1}{n} X_{(n)}$ la cui varianza è $\frac{\theta^2}{n(n+2)}$.

2. Individuare in quale punto della dimostrazione della disuguaglianza di Cramér-Rao si utilizza l'ipotesi di modello regolare.

Corollario 4.38 (determinare stimatori efficienti) *Nelle ipotesi del Teorema 4.32, sia T uno stimatore corretto di $a(\theta) \in \mathbb{R}$. Allora T raggiunge il limite inferiore di Cramér-Rao se e solo se esiste una funzione $h(\theta)$ tale che*

$$\frac{d \log(\mathcal{L}(\theta; \underline{x}))}{d\theta} = h(\theta)\left(T(\underline{x}) - a(\theta)\right)$$

Proof. La dimostrazione del teorema di Cramér-Rao si basa sul fatto che il coefficiente di correlazione tra due variabili aleatorie reali X e Y su $(\Omega, \mathcal{A}, \mathrm{P})$ è compreso tra -1 e 1, equivalentemente sulla disuguaglianza di Cauchy-Schwartz. Il coefficiente di correlazione è ± 1 se e solo se esistono $a, b \in \mathbb{R}$ con $a \neq 0$ tale che $Y = aX + b$ quasi certamente, cioè $\mathrm{P}(Y = aX + b) = 1$. Applichiamo questo fatto a $Y = \frac{d \log(\mathcal{L}(\theta;\underline{x}))}{d\theta}$ e $X = T$. ■

Osservazioni 4.39 (semplificare)

1. Perché h e a nel Corollario 4.38 dipendono da θ?
2. Si confrontino i Corollari 4.36 e 4.38. Dedurre dal Corollario 4.38 un'altra dimostrazione del fatto che uno stimatore efficiente, se esiste, è unico.
3. Calcolando il valore atteso di entrambi i membri dell'uguaglianza nel Corollario 4.38 si ha $a(\theta) = \mathrm{E}_\theta\left(T(\underline{X})\right)$ poiché il valor medio della funzione score è nullo.
4. Calcolando la varianza di entrambi i membri dell'uguaglianza nel Corollario 4.38 si ha

$$\mathrm{Var}_\theta\left(\frac{d \log(\mathcal{L}(\theta;\underline{X}))}{d\theta}\right) = h(\theta)^2 \,\mathrm{Var}_\theta\left((T(\underline{X}) - a(\theta))\right) = h(\theta)^2 \,\mathrm{Var}_\theta\left(T(\underline{X})\right)$$

   Poiché per definizione $I(\theta) = \mathrm{Var}_\theta\left(\frac{d \log(\mathcal{L}(\theta;\underline{x}))}{d\theta}\right)$, si deduce $I(\theta) = h^2(\theta)\,\mathrm{Var}_\theta(T)$.

5. Utilizzare la (dis)uguaglianza di Cramér-Rao per concludere che se T è corretto ed efficiente per $\theta \in \mathbb{R}$ allora $\mathrm{Var}(T) = 1/|h(\theta)| = 1/I(\theta)$.
6. In conclusione, per stimatori efficienti del loro valore atteso vale $\dfrac{dl}{d\theta} = I(\theta)(T - \mathrm{E}_\theta(T))$. Questa relazione fornisce un modo per calcolare $I(\theta)$ noto T stimatore efficiente di $\mathrm{E}_\theta(T)$.

Esempio 4.40

1. Sia $X_1, \ldots, X_n$ un campione i.i.d. Poisson(λ), con $\lambda > 0$. Si ha

$$\begin{aligned} l &= -n\lambda + \log\lambda \sum_{i=1}^n x_i + costante \\ \frac{dl}{d\lambda} &= -n + \frac{\sum_{i=1}^n x_i}{\lambda} = \frac{n}{\lambda}(\bar{x} - \lambda) \end{aligned}$$

   da cui $\bar{X}$ è efficiente per λ e $\mathrm{Var}_\lambda\left(\bar{X}\right) = \dfrac{\lambda}{n} = I(\lambda)^{-1}$. Da $\frac{dl}{d\lambda} = \dfrac{1}{\lambda}\left(\sum_{i=1}^n x_i - n\lambda\right)$ si ha $I(n\lambda) = 1/\lambda$.

2. Sia $X_1, \ldots, X_n$ un campione i.i.d. Bernoulli(p), con $p \in]0, 1[$. Si ha

$$\begin{aligned} l &= \sum_{i=1}^n x_i \log p + (n - \sum x_i)\log(1-p) + costante \\ \frac{dl}{dp} &= \frac{\sum x_i}{p} - \frac{n - \sum x_i}{1-p} = \frac{n}{p(1-p)}(\bar{x} - p) \end{aligned}$$

   da cui $\bar{X}$ è efficiente per p e $\mathrm{Var}_p\left(\bar{X}\right) = \dfrac{p(1-p)}{n} = I(p)^{-1}$.

3. Sia $X_1, \ldots, X_n$ un campione i.i.d. Cauchy(θ), con $\theta \in \mathbb{R}$. Si ha

$$\begin{aligned} \mathcal{L}(\theta, x_1) &= \frac{1}{\pi\left(1 + (x_1 - \theta)^2\right)} \\ \frac{dl}{d\theta} &= 2\sum_i \frac{x_i - \theta}{1 + (x_i - \theta)^2} \end{aligned}$$

   Non può scriversi come $h(\theta)(T - \theta)$ per nessun h, T e quindi non esiste stimatore efficiente di θ. Perchè questo esempio è comunque poco sensato?

Esempio 4.41 (limite inferiore di Cramér-Rao per modelli di classe esponenziale) Sia $\underline{X}$ un campione e sia $l(\theta;\underline{x}) = \theta T(\underline{x}) - \psi(\theta)$ con $\theta \in \mathbb{R}$ la log-verosimiglianza di un modello statistico per $\underline{X}$. Per l'Osservazione 4.39 vale $\psi'(\theta) = \mathrm{E}_\theta\left(T(\underline{X})\right)$ e $I(\theta) = \psi(\theta)''$, quindi il limite inferiore di Cramér-Rao è $\psi''^2/\psi'' = \psi''$. Estendere ed interpretare il risultato per $\theta \in \mathbb{R}^p$.

Esempio 4.42 1. Sia $X_1, \ldots, X_n$ un campione i.i.d. $\mathcal{N}(0, \theta^2)$, con $\theta \in \mathbb{R}$. Si ha

$$\frac{dl}{d\theta^2} = -\frac{n}{2\theta^2} + \frac{\sum x_i^2}{2\theta^4} = \frac{n}{2\theta^4}\left(\frac{\sum x_i^2}{n} - \theta^2\right)$$

da cui $\dfrac{\sum X_i^2}{n}$ è stimatore efficiente per θ^2 e ha varianza $\dfrac{2\theta^4}{n}$. Si osservi che la varianza campionaria $S^2 = \dfrac{\sum X_i^2}{n-1}$ non raggiunge il limite inferiore di Cramér-Rao ed è stimatore corretto per θ^2.

2. Sia $X_1, \ldots, X_n$ un campione i.i.d. $\mathcal{N}(\mu, \theta^2)$, con $(\mu, \theta) \in \mathbb{R} \times \mathbb{R}_{>0}$. Si ha

$$\frac{dl}{d\theta^2} = -\frac{n}{2\theta^2} + \frac{\sum(x_i - \mu)^2}{2\theta^4} = \frac{n}{2\theta^4}\left(\frac{\sum(x_i - \mu)^2}{n} - \theta^2\right)$$

da cui se μ è noto lo stimatore efficiente di σ^2 è $\dfrac{\sum(X_i - \mu)^2}{n}$ mentre non esiste se μ non è noto.

Esercizio 4.43 (stimatore UMVUE non efficiente) Sia $X_1, \ldots, X_n$ un campione casuale Exponential(λ), $\lambda > 0$. Lo stimatore $T = (n-1)/\sum X_i$ è corretto e di minima varianza in $\mathcal{D}_{2,c}$ per λ, ma non raggiunge il limite inferiore di Cramér-Rao. Infatti $\sum X_i \sim \Gamma(n, \lambda)$[9] da cui

$$\mathrm{E}_\lambda\left(\left(\frac{1}{\sum X_i}\right)^2\right) = \int_0^{+\infty} \frac{1}{t^2}\frac{\lambda^n e^{-\lambda t} t^{n-1}}{(n-1)!}\,dt = \frac{\lambda^2}{(n-1)(n-2)}\int_0^{+\infty}\frac{\lambda^{n-2}e^{-\lambda t}t^{n-3}}{(n-3)!}\,dt = \frac{\lambda^2}{(n-1)(n-2)},$$

$$\mathrm{E}_\lambda\left(\frac{1}{\sum X_i}\right) = \int_0^{+\infty}\frac{1}{t}\frac{\lambda^n e^{-\lambda t}t^{n-1}}{(n-1)!}\,dt = \ldots = \frac{\lambda}{n-1}$$

Quindi T è stimatore corretto per λ e ha varianza $\frac{\lambda^2}{n-2}$. La varianza è minima per il teorema di Lehmann-Scheffè perchè T è funzione della statistica sufficiente $\sum X_i$.
Da $l(\lambda) = n\log(\lambda) - \lambda\sum x_i$ si ha $\frac{dl}{d\lambda} = \frac{n}{\lambda} - \sum x_i$ e $\frac{d^2l}{d\lambda^2} = \frac{-n}{\lambda^2}$ e si ottiene che l'informazione di Fisher è

$$I(\lambda) = -\,\mathrm{E}\left(\frac{d^2 l}{d\lambda^2}\right) = n/\lambda^2$$

Quindi il limite inferiore di Cramér-Rao λ^2/n è strettamente minore della varianza di T.

Esercizio 4.44 (esercizio riassuntivo, cf. Gasparini in bibliografia) Sia $X_1, \ldots, X_n$ un campione casuale Poisson(λ), $\lambda > 0$.

- Stimare λ in massima verosimiglianza.

$$\begin{aligned}
\mathcal{L}(\lambda, \underline{x}) &\propto e^{-n\lambda}\lambda^{s_n} \qquad \text{con } S_n = \sum_{i=1}^n X_i \sim \text{Poisson}(n\lambda)\\
l(\lambda, \underline{x}) &= -n\lambda + s_n \log\lambda + \textit{costante in } \lambda\\
\frac{dl}{d\lambda} &= -n + s_n/\lambda = 0 \qquad \text{sse } \lambda = s_n/n\\
\frac{d^2 l}{d\lambda^2} &= -s_n/\lambda^2 \leq 0 \qquad \text{per ogni } \lambda, s_n
\end{aligned}$$

Quindi $\hat{\lambda}_{MV} = S_n/n$. Si osservi inoltre che Poisson(λ) è un modello di classe esponenziale con parametro canonico $\log\lambda$, statistica canonica S_n e funzione dei cumulanti $n\lambda$.

[9] Per $X \sim \Gamma(\alpha, \lambda)$, $\alpha, \lambda > 0$ e $x > 0$, $f_X(x) = \dfrac{\lambda^\alpha x^{\alpha-1}e^{-\lambda x}}{\Gamma(\alpha)}$, $\mathrm{E}(X) = \alpha/\lambda$ e $\mathrm{Var}(X) = \alpha/\lambda^2$.

1. Calcolare l'informazione di Fisher in λ.
$$I(\lambda) = \mathrm{E}_\lambda\left(\left(\frac{dl}{d\lambda}\right)^2\right) = \mathrm{E}_\lambda\left((-n+S_n/\lambda)^2\right) = \frac{1}{\lambda^2}\mathrm{E}_\lambda\left((-n\lambda+S_n)^2\right) = \frac{\mathrm{E}_\lambda(S_n)}{\lambda^2} = n/\lambda$$
o più velocemente
$$I(\lambda) = -\mathrm{E}_\lambda\left(\left(\frac{dl^2}{d^2\lambda}\right)^2\right) = -\mathrm{E}_\lambda\left((S_n/\lambda\right) = \frac{n\lambda}{\lambda^2} = n/\lambda$$

2. Calcolare il limite inferiore di Cramér-Rao per stimatori corretti di λ. E' $I(\lambda)^{-1} = \lambda/n$.

3. Calcolare il limite inferiore di Cramér-Rao per $\hat{\lambda}_{MV}$. E' l'inverso dell'informazione di Fisher perchè $\mathrm{E}_\lambda\left(\hat{\lambda}_{MV}\right) = n\lambda/n = \lambda$ ovvero $\hat{\lambda}_{MV}$ è corretto per λ.

4. Verificare che $\hat{\lambda}_{MV}$ è efficiente per λ.
$$\mathrm{Var}_\lambda(S_n/n) = \frac{1}{n^2}\mathrm{Var}_\lambda(S_n) = n\lambda/n^2 = \lambda/n = \text{limite inferiore di Cramér-Rao}$$

- Stimare $g(\lambda) = \mathrm{P}_\lambda(X_1 = 0) = \exp(-\lambda)$ con $T_1(\underline{X}) = \frac{1}{n}\sum_{i=1}^n(X_i = 0)$, la frequenza di zeri nel campione. E' uno stimatore plug-in o per analogia (cf. stimatori dei momenti). Si ha $nT_1 \sim \text{Binomial}(n, e^{-\lambda})$.

5. Si verifichi che T_1 è corretto e se ne calcoli la varianza.
$$\mathrm{E}(T_1) = \frac{ne^{-\lambda}}{n} = e^{-\lambda}\,(= g(\lambda))$$
$$\mathrm{Var}(T_1) = \frac{1}{n^2}(ne^{-\lambda}(1-e^{-\lambda})) = \frac{e^{-\lambda}(1-e^{-\lambda})}{n}$$

6. Calcolare il limite inferiore di Cramér-Rao di T_1.
$$\text{l.i. di C-R} = \frac{a'(\lambda)^2}{I(\lambda)}$$
$$I(\lambda) = -\mathrm{E}\left(\frac{d^2l}{d\lambda^2}\right) = \mathrm{E}\left(\frac{S_n}{\lambda^2}\right) = \frac{n\lambda}{\lambda^2} = \frac{n}{\lambda}$$
$$\frac{a'(\lambda)^2}{I(\lambda)} = \frac{(-e^{-\lambda})^2}{\frac{n}{\lambda}} = \frac{\lambda e^{-2\lambda}}{n} < \mathrm{Var}(T_1)$$
Quindi T_1 non è efficiente per $e^{-\lambda}$. Il limite inferiore di Cramér-Rao per T_1 si può anche ottenere come
$$\frac{\left(\frac{d}{de^{-\lambda}}\mathrm{E}_{e^{-\lambda}}(T_1)\right)^2}{I(\exp(-\lambda))} = \frac{\left(\frac{d}{d\lambda}\mathrm{E}_\lambda(T_1)\right)^2}{I(\lambda)} = \frac{e^{-2\lambda}}{n/\lambda} = \lambda e^{-2\lambda}/n$$

- Sia $T_2 = \left(\frac{n-1}{n}\right)^{\sum_{i=1}^n X_i}$. E' noto che $\sum_{i=1}^n X_i \sim \text{Poisson}(n\lambda)$ ed è statistica canonica in un modello di classe esponenziale con parametro canonico $\log\lambda$, in particolare è sufficiente e completa per λ nell'n-campione casuale Poisson. Quindi si può far rifermento al teorema di Lehmann-Scheffè per verificare se T_2 è UMVUE per $e^{-\lambda}$. Per calcolo diretto si può verificare che non raggiunge il limite inferiore di Cramér-Rao.

7. Calcolare $\mathrm{E}_\lambda(T_2)$.
$$\begin{aligned}\mathrm{E}_\lambda(T_2) &= \sum_{k=0}^{+\infty}\left(\frac{n-1}{n}\right)^k \mathrm{P}\left(\sum_{i=1}^n X_i = k\right) = \sum_{k=0}^{+\infty}\left(\frac{n-1}{n}\right)^k \frac{e^{-n\lambda}(n\lambda)^k}{k!} = \sum_{k=0}^{+\infty}\frac{(n-1)^k\lambda^k e^{-n\lambda}}{k!}\\ &= \sum_{k=0}^{+\infty}\frac{((n-1)\lambda)^k}{k!}e^{-(n-1)\lambda}e^{-\lambda} = e^{-\lambda}\end{aligned}$$
Quindi T_2 è corretto per $e^{-\lambda}$.

8. Verificare che $\mathrm{Var}_\lambda(T_2) = e^{-2\lambda}(e^{\lambda/n} - 1)$. [...omissis...] Si osservi che

$$\mathrm{Var}_\lambda(T_1) > \mathrm{Var}_\lambda(T_2) > \frac{a'(\lambda)^2}{I(\lambda)} = \frac{\lambda e^{-2\lambda}}{n}$$

quindi T_2 è preferibile a T_1 in $\mathcal{D}_{2,c}$. Inoltre

$$\frac{e^{-\lambda}(1-e^{-\lambda})}{n} > \frac{e^{\lambda/n}-1}{\lambda/n}\frac{\lambda e^{-2\lambda}}{n} > \frac{\lambda e^{-2\lambda}}{n}$$

convergenti a zero per n a più infinito.

- Si noti che $T_3(\underline{X}) = e^{-\bar{X}}$ è lo stimatore di massima verosimiglianza di $g(\lambda)$. Vale

$$\mathrm{E}_\lambda\left(e^{-\bar{X}}\right) > e^{-\mathrm{E}_\lambda(\bar{X})} = e^{-\lambda}$$

per la diseguaglianza di Jensen. Quindi T_3 è distorto per $e^{-\lambda}$.

9. Calcolare valore atteso e varianza di T_3 e confrontarli con quelli di T_1 e T_2.

Esercizio 4.45 (esercizio riassuntivo, normale) 1. Verificare che $\bar{X}$ in un n-campione casuale $\mathcal{N}(\mu, \sigma^2)$, con σ noto, è stimatore efficiente per μ e che il limite inferiore di Cramér-Rao è σ^2/n. Si può procedere in vari modi: direttamente, utilizzando la teoria dei modelli di classe esponenziale, e direttamente il Corollario 4.38.

METODO DIRETTO

$$l_1(\mu, \underline{x}) \propto -\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$
$$\frac{dl}{d\mu} = \frac{\sum_{i=1}^n (x_i-\mu)}{\sigma^2}$$

da cui

$$I(\mu) = \mathrm{E}_\mu\left(\left(\frac{dl}{d\mu}\right)^2\right) = \mathrm{E}_\mu\left(\left(\frac{\sum_i (X_i-\mu)^2}{\sigma^2}\right)\right) = \frac{1}{\sigma^4}\sum_i \mathrm{E}_\mu\left((X_i-\mu)^2\right) = \frac{1}{\sigma^4}n\sigma^2 = \frac{n}{\sigma^2}$$

inoltre

$$a'(\mu) = \frac{d}{d\mu}\mathrm{E}_\mu(\bar{X}) = \frac{d}{d\mu}\mu = 1 \qquad \bar{X} \text{ è corretto per } \mu$$

quindi il limite inferiore di Cramér-Rao è $\frac{a'(\mu)^2}{I_n(\mu)} = \frac{1}{\frac{n}{\sigma^2}} = \frac{\sigma^2}{n}$. Infine

$$\mathrm{Var}_\mu(\bar{X}) = \mathrm{E}_\mu\left((\bar{X}-\mu)^2\right) = \frac{\sigma^2}{n} = \text{limite inferiore di Cramér-Rao}$$

quindi $\bar{X}$ è efficiente per μ. Provare gli altri due metodi.

2. Calcolare il limite inferiore di Cramér-Rao per W stimatore corretto di σ^2, con μ non noto. Occorre calcolare l'informazione di Fisher:

$$\frac{\partial^2}{(\partial\sigma^2)^2}\log\left(\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}\right) = \frac{1}{2\sigma^4} - \frac{(x-\mu)^2}{\sigma^6}$$

e quindi

$$-\mathrm{E}\left(\frac{\partial^2}{(\partial\sigma^2)^2}l\right) = -\mathrm{E}\left(\frac{1}{2\sigma^4} - \frac{(X-\mu)^2}{\sigma^6}\right) = -\frac{1}{2\sigma^4} + \frac{1}{\sigma^6}\sigma^2 = \frac{1}{2\sigma^4} = \text{informazione di Fisher per 1-campioni}$$

da cui per W corretto per σ^2 il limite inferiore di Cramér-Rao è $\frac{2\sigma^4}{n}$ e vale $\mathrm{Var}_{\mu,\sigma^2}(W) \geq \frac{1}{2\sigma^4}$.

Si noti che la varianza campionaria $S^2 = \dfrac{\sum(X_i - \bar{X})^2}{n-1}$ non è stimatore efficiente di σ^2 infatti $\mathrm{Var}_{\mu,\sigma^2}(S^2) = \frac{2\sigma^4}{n-1}$. Dedurlo da $\mathrm{E}\,\chi^2(k) = k$ e $\mathrm{Var}\,\chi^2(k) = 2\,k$.

3. Esiste uno stimatore corretto ed efficiente di σ^2? Utilizziamo il Corollario 4.38.

$$\mathcal{L}(\mu, \sigma^2|\underline{x}) \propto \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left(-\frac{\sum_i (x_i - \mu)^2}{2\sigma^2}\right)$$
$$\frac{\partial}{\partial(\sigma^2)} \log \mathcal{L} = \frac{n}{2\sigma^4}\left(\frac{\sum_i (x_i - \mu)^2}{n} - \sigma^2\right) \qquad = b(\sigma^2)\left(T(\underline{x}) - \sigma^2\right)$$

Esistono b e T per cui valga l'ultima uguaglianza? Si ha $b(\sigma^2) = \frac{n}{2\sigma^4}$, ma l'unico candidato per T dipende da μ sconosciuto. Non esiste stimatore efficiente di σ^2 se μ è sconosciuto, se μ è noto $\frac{\sum_i (X_i - \mu)^2}{n}$ è efficiente per σ^2.

MODELLO NORMALE COME MODELLO DI CLASSE ESPONENZIALE

1**-CAMPIONE**

4. Per $X \sim \mathcal{N}(\mu, \sigma^2)$ (conti leggermente diversi da quanto sopra)

$$\mathcal{L}(\mu, \sigma^2|x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2} + \frac{x\mu}{\sigma^2} - \frac{\mu^2}{2\sigma^2}\right) = \exp\left((-\frac{1}{2}\log(2\sigma^2) - \frac{\mu^2}{2\sigma^2}) + (\frac{\mu}{\sigma^2}x - \frac{1}{2\sigma^2}x^2)\right)$$
$$l(\mu, \sigma^2|x) = -\frac{1}{2}\log(2\sigma^2) - \frac{\mu^2}{2\sigma^2} + (\frac{\mu}{\sigma^2}x - \frac{1}{2\sigma^2}x^2)$$

Scegliamo i parametri naturali $\theta_1 = \frac{\mu}{\sigma^2} \in \mathbb{R}$ e $\theta_2 = -\frac{1}{2\sigma^2} \in \mathbb{R}_{<0}$ con trasformazione inversa $\mu = -\frac{\theta_1}{2\theta_2}$ e $\sigma^2 = -\frac{1}{2\theta_2}$. La statistica canonica è $T = (X, X^2)$ e la funzione dei cumulanti è

$$\psi(\mu, \sigma^2) = \frac{1}{2}\log(2\sigma^2) + \frac{\mu^2}{2\sigma^2} = -\frac{1}{2}\log(-\theta_2) - \frac{\theta_1^2}{4\theta_2}$$

5. Calcolare il valore atteso di T: vale $\text{gradiente}_\theta\,\psi(\theta) = \mathrm{E}_\theta(T)$ e

$$\frac{\partial}{\partial\theta_1}\psi = -\frac{2\theta_1}{4\theta_2} = \mu \qquad \frac{\partial}{\partial\theta_2}\psi = -\frac{1}{2}\frac{1}{\theta_2} + \frac{\theta_1^2}{4}\frac{1}{\theta_2^2} = \sigma^2 + \mu^2$$

6. Calcolare la varianza di T: $\mathrm{Var}_{\theta_1,\theta_2}(T) = \mathrm{Hessian}_{\theta_1,\theta_2}(\psi) = I(\theta)$ e

$$\frac{\partial^2\psi}{\partial\theta_1^2} = -\frac{1}{2\theta_2}$$
$$\frac{\partial^2\psi}{\partial\theta_1\theta_2} = \frac{\theta_1}{2\theta_2^2}$$
$$\frac{\partial^2\psi}{\partial\theta_2^2} = \frac{1}{2\theta_2^2}\left(1 - \frac{\theta_1^2}{\theta_2}\right)$$

da cui

$$\mathrm{Hessian}_\theta\,\psi(\theta) = -\frac{1}{2\theta_2}\begin{bmatrix} 1 & -\frac{\theta_1}{\theta_2} \\ -\frac{\theta_1}{\theta_2} & \left(\frac{\theta_1}{\theta_2}\right)^2 - -\frac{1}{\theta_2} \end{bmatrix}$$

e quindi

$$\mathrm{Var}_\theta(T) = I(\theta)^{-1} = \begin{bmatrix} \mathrm{Var}(X) & \mathrm{VarCov}(X, X^2) \\ \mathrm{VarCov}(X, X^2) & \mathrm{Var}(X^2) \end{bmatrix} = I(\mu, \sigma^2) = \begin{bmatrix} \sigma^2 & 2\mu\sigma^2 \\ 2\mu\sigma^2 & 4\mu^2\sigma^2 + 2\sigma^4 \end{bmatrix}$$

n-CAMPIONE

7. Verificare che $\bar{X}, S^2$ è stimatore non distorto di (μ, σ^2). Si ha

$$l(\mu, \sigma^2; x_1, \ldots, x_n) = -n\psi(\theta) + \langle \theta, \sum_i T(x_i) \rangle$$

$$\mathrm{E}\left(\sum_i X_i\right) = n\mu \qquad \mathrm{E}\left(\sum_i X_i^2\right) = n\mu$$

$$\mathrm{Var}(T) = \begin{bmatrix} n\sigma^2 & 2n\mu\sigma^2 \\ 2n\mu\sigma^2 & 2n\sigma^2(\mu^2 + 2\sigma^2) \end{bmatrix}$$

Quindi $\bar{X}$ è stimatore non distorto di μ e per $S^2 = \frac{\sum X_i^2}{n-1} - \frac{1}{n(n-1)}(\sum_i X_i)^2$ si ha $\mathrm{E}\left(S^2\right) = \frac{\mathrm{E}(\sum X_i^2)}{n-1} - \frac{1}{n(n-1)}\left(\mathrm{E}\left(\sum X_i^2\right) + \mathrm{E}\left(\sum_{i \neq j} X_i X_j\right)\right) = ... = \sigma^2$

8. Il limite inferiore di Cramér-Rao di $(\bar{X}, S^2)$ può essere calcolato nella parametrizzazione θ o (μ, σ^2). ....

GAUSSIANA MULTIVARIATA

9. Per $\mu \in \mathbb{R}^n$ e $\Sigma \in \mathcal{S}_n^+$ (insieme delle matrici definite positive di taglia n) sia $\underline{X} \sim \mathcal{N}_n(\mu, \Sigma)$ ovvero per $x \in \mathbb{R}^n$ valga

$$\begin{aligned} f_{\underline{X}}(x) &= \frac{1}{\sqrt{(2\pi)\det(\Sigma)}} \exp\left(-\frac{1}{2}(x-\mu)^t \Sigma^{-1}(x-\mu)\right) \\ &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^t\Sigma^{-1}x}{2} + \mu^t\Sigma^{-1}x - \frac{\mu^t\Sigma^{-1}\mu}{2} - \frac{1}{2}\log(\det(\Sigma))\right) \\ &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\operatorname{trace}(xx^t\Sigma^{-1}) + \mu^t\Sigma^{-1}x - \psi\right) \end{aligned}$$

con $\psi(\mu, \Sigma) = \frac{\mu^t\Sigma^{-1}\mu}{2} + \frac{1}{2}\log\det(\Sigma)$. Verificare $x^t\Sigma^{-1}x = \operatorname{trace}(xx^t\Sigma^{-1})$.

10. Verificare che si tratta di un modello di classe esponenziale con parametri canonici $\xi = \Sigma^{-1}\mu$ e $K = \Sigma^{-1}$ e con statistiche sufficienti $\underline{X}$ e $S = \underline{XX^t}$ e che quindi si può scrivere

$$f_{\underline{X}}(x) = -\frac{1}{2}\operatorname{trace}(sK) + \xi^t x = -\frac{1}{2}\sum_{i,k=1}^{n} s_{ik}K_{ik} + \sum_{i=1}^{n} \xi_i x_i - \psi(\mu, \Sigma)$$

11. Determinare lo spazio dei parametri per i parametri canonici, esprimere $\psi(\mu, \Sigma)$ nei parametri canonici e verificare che $\psi(\xi, K) = -\frac{1}{2}\log\det(K) + \frac{1}{2}\xi^t K^{-1}\xi$.

12. Un v-campione casuale da $\mathcal{N}_n(\mu, \Sigma)$ è di classe esponenziale, con statistica canonica $\left(\bar{X}, S\right)$ dove $\bar{X} = \dfrac{\sum_{i=1}^{v} \underline{X}_i}{v}$ e $S = \dfrac{\sum_{i=1}^{v} \underline{X}_i \underline{X}_i^t}{v} = \dfrac{\sum_{i=1}^{v} S_i}{v}$, parametro canonico $\theta = (\mu^t\Sigma^{-1}, \Sigma^{-1})^t$, funzione dei cumulanti $\dfrac{\psi_n(\theta)}{n} = \dfrac{\mu^t\Sigma^{-1}\mu}{2} + \dfrac{n}{2}\log\det(\Sigma)$ e funzione di log-verosimiglianza

$$l = -\frac{1}{2}\operatorname{trace}(SK) + \xi^t\bar{X} - \frac{1}{2}\xi^t K^{-1}\xi + \frac{v}{2}\log\det(K)$$

5 Statistica Bayesiana parametrica (da sistemare)

Sia $\underline{X} = (X_1, \ldots, X_n)$ un campione e $\mathcal{F} = \left\{ f_{\underline{X}}(\cdot;\theta) : \theta \in \Theta \subset \mathbb{R}^k \right\}$ un modello statistico parametrico per $\underline{X}$ espresso in termini di densità di probabilità indicizzate dal parametro θ.
Si suppone che θ sia un vettore aleatorio con legge $\pi(\theta)$ (densità a priori o iniziale) che esprime l'incertezza sperimentale e/o informazioni a priori sul parametro θ.
Sia $f_{(\underline{X},\theta)}(\underline{x},\theta) = \pi(\theta) f_{\underline{X}}(\underline{x};\theta) = \pi(\theta)\mathcal{L}(\theta,\underline{x})$ la legge congiunta di $\underline{X}$ e θ. La funzione di verosimiglianza è interpretata come densità condizionata a $\underline{X} = \underline{x}$ di θ e la legge congiunta è fattorizzata applicando la regola ricorsiva di fattorizzazione di una densità di probabilità congiunta (attenzione a condizionare su eventi di probabilità strettamente positiva).
Per la regola ricorsiva si può anche scrivere $f_{(\underline{X},\theta)}(\underline{x},\theta) = f_{\underline{X}}(\underline{x}) f_{\theta|\underline{X}}(\theta|\underline{x})$. La densità marginale $f_{\underline{X}}(\underline{x})$ è detta densità predittiva a priori e $\pi(\theta|\underline{x}) = f_{\theta|\underline{X}}(\theta|\underline{x})$ densità a posteriori o finale di θ. Si osservi che

$$\pi(\theta|\underline{x}) = \begin{cases} \dfrac{\pi(\theta)\mathcal{L}(\theta,\underline{x})}{\int_\Theta f_{(\underline{X},\theta)}(\underline{x},\theta)\, d\theta} & \text{se } \theta \text{ è un parametro continuo} \\ \dfrac{\pi(\theta)\mathcal{L}(\theta,\underline{x})}{\sum_{\theta\in\Theta} f_{(\underline{X},\theta)}(\underline{x},\theta)\, d\theta} & \text{se } \theta \text{ è un parametro discreto} \end{cases}$$

La quantità $\int_\Theta f_{(\underline{X},\theta)}(\underline{x},\theta)\, d\theta = 1/C(\underline{x})$ (equivalentemente $\sum_{\theta\in\Theta} f_{(\underline{X},\theta)}(\underline{x},\theta)\, d\theta$) è detta densità predittiva, è una costante di normalizzazione, è la marginale delle osservazioni.
Riassumendo

$$\pi(\theta|\underline{x}) = C(\underline{x})\pi(\theta)L(\theta,\underline{x})$$

I modelli Bayesiani possono essere rappresentanti come modelli gerarchici a due livelli $\underline{X}\big|\theta \sim f_{\underline{X}}(\underline{x};\theta)$ e $\theta \sim \pi(\cdot)$.

Corollario 5.1 *1. Se T è statistica sufficiente per θ cioè $f_{(\underline{X};\theta)}(\underline{x};\theta) = f_{(\underline{X}|\theta)}(\underline{x}|\theta) = h(\underline{x})g(T(\underline{x}),\theta)$ per opportuni g e h allora $\pi(\theta|\underline{x}) = C(\underline{x})\pi(\theta)h(\underline{x})g(T(\underline{x}),\theta)$.*

2. Se $f_{\underline{X}}(\underline{x};\theta) = \prod_{i=1}^n f_{X_1}(x_i;\theta)$ allora $\pi(\theta|\underline{x}) = C(\underline{x})\prod_{i=1}^n \pi(\theta) f_{X_1}(x_i;\theta) \propto \prod_{i=1}^n p(\theta|x_i)$.

Per campione i.i.d. si intende i.i.d. condizionatamente a θ ovvero $f_{\underline{X}}(\underline{x}|\theta) = \prod_{i=1}^n f_{X_i}(x_i|\theta)$. La costante costante di proporzionalità nel punto 2 del Corollario 5.1 è funzione di tutto il campione osservato.

Esempio 5.2 (v. Gasparini) Si consideri $X_1, \ldots, X_n\big|\theta$ i.i.d. Bernoulli(θ) con $\theta \in \{0.4, 0.5, 0.6\}$. La densità a posteriori di θ è strettamente positiva per ogni θ e ogni valore della statistica sufficiente $S_n = \sum_{i=1}^n X_i \sim \mathrm{Bin}(n,\theta)$

$$\pi(\theta|\underline{x}) = \pi(\theta)\mathcal{L}(\theta,\underline{x})C(\underline{x}) = \frac{\theta^s(1-\theta)^{n-s}\pi(\theta)}{\sum_{\theta\in\{0.4,0.5,0.6\}}\theta^s(1-\theta)^{n-s}\pi(\theta)} \quad \text{per } s \in \{0,1,\ldots,n\}$$

Si considerino due diverse leggi a priori su θ e due esperimenti: nel primo per $n = 10$ si osserva $s_{10} = 5$, nel secondo si ha $s_{100} = 52$. La seguente tabella riassume i due casi

		$s_{10} = 5$	$s_{100} = 52$
θ	$\pi(\theta)$	$\pi(\theta\|\underline{x})$	$\pi(\theta\|\underline{x})$
0.4	1/3	0.31	0.04
0.5	1/3	0.38	0.74
0.6	1/3	0.31	0.22
0.4	0.5	0.46	0.06
0.5	0.4	0.45	0.87
0.6	0.1	0.09	0.06

Osservare che la a-posteriori non è mai nulla e interpretare le simmetrie ottenute e distrutte.

Esempio 5.3 Sia $X_1, \ldots, X_n\big|\theta$ i.i.d. Bernoulli(θ) con $\theta \sim \mathrm{beta}(a,b)$ con $a, b > 0$ noti, ovvero $\pi(\theta) = \dfrac{\theta^{a-1}(1-\theta)^{b-1}}{\mathrm{B}(a,b)}$ dove $\mathrm{B}(a,b) = \dfrac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ è la funzione beta $(E)_{a,b}(\theta) = \dfrac{a}{a+b}$ e $(Var)_{a,b}(\theta) = \dfrac{ab}{(a+b)^2(a+b+1)}$.
Si osservi che $\mathrm{Unif}[0,1] = \mathrm{B}(1,1)$.

Con $s = \sum x_i$ vale

$$\pi(\theta|\underline{x}) = C(\underline{x})\pi(\theta)\mathcal{L}(\theta, \underline{x}) \propto \theta^{a-1}(1-\theta)^{b-1}\ \theta^s(1-\theta)^{n-s} = \theta^{a+s-1}(1-\theta)^{b+n-s-1}$$

ovvero la legge a posteriori è beta$(a+s, b+n-s)$ e quindi la costante di normalizzazione è l'inverso della funzione di parametri $a+s$ e $b+n-s$. Si dice che il campionamento bernoulliano e la a priori Beta sono coniugate.

Esempio 5.4 Per

$$\begin{aligned} X_1, \ldots, X_n \big| \theta &\overset{i.i.d.}{\sim} \mathcal{N}(\mu, \tau^{-1}) \text{ con } \tau = \frac{1}{\sigma^2} \text{ parametro di precisione} \\ \mu &\sim \mathcal{N}(\mu_0, \tau_0^{-1}) \end{aligned}$$

con τ, μ_0, τ_0 noti. Vale

$$\begin{aligned} \pi(\mu|\underline{x}) &\propto \pi(\mu)\mathcal{L}(\mu, \underline{x}) \\ &\propto \exp\left(-\frac{\tau_0}{2}(\mu-\mu_0)^2\right)\exp\left(-\frac{\tau}{2}\sum_{i=1}^n (x_i-\mu)^2\right) \\ &\propto \exp\left(-\frac{\tau_0}{2}(\mu-\mu_0)^2\right)\exp\left(-\frac{\tau}{2}\left[\sum_{i=1}^n (x_i-\bar{x})^2 + n(\mu-\bar{x})^2\right]\right) \\ &\propto \exp\left(-\frac{\tau_0}{2}(\mu-\mu_0)^2 - \frac{n\tau}{2}(\mu-\bar{x})^2\right) \text{ la parte non riportata va in } C(\underline{x}) \text{ perchè non dipende dal parametro} \\ &= \exp\left(-\frac{1}{2}\left[(\tau_0+n\tau)\left(\mu - \frac{\tau_0\mu_0 + n\tau\bar{x}}{\tau_0+n\tau}\right)^2 + \frac{n\tau\tau_0(\bar{x}-\mu_0)^2}{\tau_0+n\tau}\right]\right) \text{ completamento dei quadrati} \\ &\propto \exp\left(-\frac{1}{2}(\tau_0+n\tau)\left(\mu - \frac{\tau_0\mu_0+n\tau\bar{x}}{\tau_0+n\tau}\right)^2\right) \end{aligned}$$

ovvero $\mu|\underline{x} \sim \mathcal{N}\left(\frac{\tau_0\mu_0+n\tau\bar{x}}{\tau_0+n\tau}, \frac{1}{\tau_0+n\tau}\right)$. Si noti che il valor medio a posteriori di μ si può scrivere come

$$\frac{\tau_0\mu_0+n\tau\bar{x}}{\tau_0+n\tau} = \mu_0 + \frac{\bar{x}-\mu_0}{1+\frac{\tau_0}{n\tau}} = \frac{\tau_0}{\tau_0+n\tau}\mu_0 + \frac{n\tau}{\tau_0+n\tau}\bar{x}$$

Questa è una combinazione convessa di valore atteso a priori di μ e stima di massima verosimiglianza di μ e il suo limite per n che tende a più infinito è la stima di massima verosimiglianza di μ.
Per $\tilde{\mu} = \frac{\tau_0\mu_0+n\tau\bar{x}}{\tau_0+n\tau}$ and $\tilde{\sigma}^2 = \frac{1}{\tau_0+n\tau}$ si può scrivere $\mu|\underline{x} \sim \mathcal{N}\left(\tilde{\mu}, \tilde{\sigma}^2\right)$. Si vuole determinare α, β such that la probablità a posteriori che μ appartenga all'intervallo $[\alpha, \beta]$ sia 0.95, ovvero l'intervallo di confidenza bayesiano al 95% per μ. Si fa

$$\begin{aligned} 0.95 &= \mathrm{P}\left(\mu \in [\alpha, \beta] \big| \bar{x}\right) \\ &= \mathrm{P}\left(\frac{\mu-\tilde{\mu}}{\tilde{\sigma}} \in \left[\frac{\alpha-\tilde{\mu}}{\tilde{\sigma}}, \frac{\beta-\tilde{\mu}}{\tilde{\sigma}}\right] \big| \bar{x}\right) \\ &= \mathrm{P}\left(Z \in \left[\frac{\alpha-\tilde{\mu}}{\tilde{\sigma}}, \frac{\beta-\tilde{\mu}}{\tilde{\sigma}}\right]\right) \end{aligned}$$

si può scegliere $\tilde{\mu} \pm 1.96\tilde{\sigma} = \frac{\tau_0}{\tau_0+n\tau}\mu_0 + \frac{n\tau}{\tau_0+n\tau}\bar{x} \pm 1.96\frac{1}{\tau_0+n\tau}$. Poiché la varianza a posteriori è minore della varianza a priori l'ampiezza dell'intervallo di confidenza bayesiano per μ è minore di quella dell'intervallo di confidenza a priori.

Esempio 5.5 Per

$$X_1,\dots,X_n|\theta \overset{i.i.d.}{\sim} \mathcal{N}(\mu,\tau^{-1}) \text{ con } \tau=\frac{1}{\sigma^2} \text{ parametro di precisione}$$
$$\mu\sim\mathcal{N}(\mu_0,(\tau\tau_0)^{-1})$$
$$\tau\Gamma(\alpha,\lambda)$$

con $\alpha,\lambda,\mu_0,\tau_0$ noti e la densità di $X\sim\Gamma(\alpha,\lambda)$ è $\pi(x)=\dfrac{\lambda^\alpha x^{\alpha-1}e^{-\lambda x}(x>0)}{\Gamma(\alpha)}$ e $\mathrm{E}(X)=\alpha/\lambda$ e $\mathrm{E}(X)=\alpha/\lambda^2$. L'ipotesi sul campione è dunque

$$\underline{X}|(\mu,\tau)\sim\mathcal{N}_n\left((\mu,\dots,\mu)^t,\mathrm{diag}(\tau^{-1})\right)$$

e la funzione di verosimiglianza è

$$\begin{aligned}\mathcal{L}(\mu,\tau;\underline{x}) &= \frac{1}{(\sqrt{2\pi})^n}\frac{1}{(\det\Sigma)^{1/2}}\exp\frac{-(\underline{x}-\underline{\mu})^t\Sigma^{-1}(\underline{x}-\underline{\mu})}{2}\\ &\propto \tau^{n/2}\exp\left(-\frac{\tau}{2}(\underline{x}-\underline{\mu})^t(\underline{x}-\underline{\mu})\right)\\ &= \tau^{n/2}\exp\left(-\frac{\tau}{2}\left[\sum_{i=1}^n(x_i-\bar{x})^2+n(\mu-\bar{x})^2\right]\right)\end{aligned}$$

e la densità congiunta dei parametri (a priori) è

$$\pi(\mu,\tau)=\pi(\tau)\pi(\mu|\tau)\propto\left(\frac{1}{\sqrt{2\pi(\tau\tau_0)^{-1}}}\exp\left(-\frac{\tau\tau_0}{2}(\mu-\mu_0)^2\right)\right)\left(\tau^{\alpha-1}\exp(-\lambda\tau)\frac{\lambda^\alpha}{\Gamma(\alpha)}(\tau>0)\right)$$

La densità congiunta a posteriori dei parametri è

$$\begin{aligned}\pi(\mu,\tau\,|\,\underline{x}) &= \pi(\mu,\tau)\mathcal{L}(\mu,\tau;\underline{x})\propto\tau^{\frac{n}{2}+\frac{1}{2}+\alpha-1}\exp\left(-\lambda\tau-\frac{\tau\tau_0}{2}(\mu-\mu_0)^2-\frac{\tau}{2}\left[\sum(x_i-\bar{x})^2+n(\mu-\bar{x})^2\right]\right)\\ &= \tau^{\frac{n}{2}+\alpha-1}\exp\left(-\tau\left(\lambda+\frac{\sum(x_i-\bar{x})^2}{2}+\frac{n\tau_0(\mu-\bar{x})^2}{2(\tau_0+n)}\right)\right)\sqrt{\tau}\exp\left(-\frac{\tau(\tau_0+n)}{2}\left(\mu-\frac{\tau_0\mu_0+n\bar{x}}{\tau_0+n}\right)^2\right)\end{aligned}$$

La prima parte non dipende da μ ed è il nucleo di una densità Gamma da cui

$$\begin{aligned}\tau|\underline{x} &\sim \Gamma\left(\alpha+\frac{n}{2},\lambda+\frac{\sum(x_i-\bar{x})^2}{2}+\frac{n\tau_0(\mu-\bar{x})^2}{2(\tau_0+n)}\right)\\ \mu|\tau\underline{x} &\sim \mathcal{N}\left(\frac{\tau_0\mu_0+n\bar{x}}{\tau_0+n},\frac{1}{\tau(\tau_0+n)}\right)\end{aligned}$$

ovvero

$$\begin{aligned}\tau\sim\Gamma &\rightsquigarrow \tau|\underline{x}\sim\Gamma\\ \mu|\tau\sim\mathcal{N} &\rightsquigarrow \mu|\tau\underline{x}\sim\mathcal{N}\end{aligned}$$

5.1 a priori

improprie Per un modello gerarchico $\underline{X}|\mu\sim\mathcal{N}_n\left(\underline{\mu},\mathrm{diag}(\tau^{-1})\right)$ e $\mu\sim\mathcal{N}_1(\mu_0,\tau_0^{-1})$ con μ_0,τ_0,τ noti, la legge a posteriori del parametro μ è $\mu|\underline{x}\sim\mathcal{N}\left(\frac{\tau_0\mu_0+n\tau\bar{x}}{\tau_0+n\tau},\frac{1}{\tau_0+n\tau}\right)$ come dimostrato nell'Esempio 5.4. Per $\tau_0=0$ la distribuzione a posteriori di μ è ben definita e è $N\left(\bar{x},\frac{1}{\tau_0+n\tau}\right)$, mentre la legge a priori non è ben definita poiché al limite per τ_0 che tende a zero avrebbe varianza infinita.

Si può pensare ad una successione di leggi a priori proprie convergenti all'impropria: sia $\{\tau_{0_i}\}$ una successione che converge a zero per i the tende a più infinito e $\mu\sim\mathcal{N}\left(\mu_0,\tau_0^{-1}\right)$. Si può dimostrare che $\mu|\underline{x}\sim\mathcal{N}\left(\dfrac{\tau_{0_i}+n\tau\bar{x}}{\tau_{0_i}+n\tau},\dfrac{1}{\tau_{0_i}+n\tau}\right)$ converge in legge a $\mathcal{N}\left(\bar{x},\frac{1}{n\tau}\right)$.

Non sempre una legge a priori impropria è definibile e non sempre ha limite *stabile*.

piatte Sia $\theta \in \Theta$ un parametro e si suppone la legge a priori $\pi(\theta)$ sia costante. Se lo spazio dei parametri è infinito la a priori piatta è impropria. Un fatto a cui fare attenzione è che le a priori piatte non sono invarianti per trasformazione.

Esempio 5.6 Siano $X \sim \text{Bernoulli}(p)$ con $p \in [0,1]$ e la trasformazione $\psi(p) = \log \frac{p}{1-p}$. La a priori è piatta se $\pi(p) = 1$). Per $v \in \mathbb{R}$ vale $\mathrm{P}_\pi(\psi \leq v) = \mathrm{P}_\pi\left(\frac{p}{1-p} \leq e^v\right) = \mathrm{P}_\pi\left(p \leq \frac{e^v}{1+e^v}\right) = \int_0^{\frac{e^v}{1+e^v}} 1\, dp = \frac{e^v}{1+e^v}$ e derivando in v si ottiene la densità a priori espressa in v: $\frac{d\,\mathrm{P}_\pi}{dv} = \frac{e^v}{(1+e^v)^2}$ che non è piatta.

di Jeffrey Se $\theta \in \mathbb{R}$ si definisce $\pi(\theta) \propto \mathrm{I}(\theta)^{1/2}$ e se $\theta \in \mathbb{R}^k$ si definisce $\pi(\theta) \propto \sqrt{\det(\mathrm{I}(\theta))}$

Esempio 5.7 Se $X \sim \text{Bernoulli}(p)$ allora $\mathrm{I}(p) = \frac{1}{p(1-p)}$ e $\pi(p) \propto \frac{1}{\sqrt{p(1-p)}} \sim \text{Beta}(1/2, 1/2)$ che è "abbastanza piatta".

5.2 Confronto con metodi frequentisti [DA SISTEMARE]

Per $\theta \in \Theta \subseteq \mathbb{R}$ si considera $\pi(\theta|\underline{x})$ e la si "riassume" con e.g. $\mathrm{E}(\theta|\underline{X})$.

Lemma 5.8 *Per Y v.a. a valori in $\mathbb{R}$ che ammette varianza finita si ha* $\mathrm{E}(Y) = \operatorname{argmin}_{a\in\mathbb{R}} \mathrm{E}((a-Y)^2)$

Proof. $\mathrm{E}((a-Y)^2) = \mathrm{E}((a-\mathrm{E}(Y))^2) + \mathrm{E}((\mathrm{E}(Y)-Y)^2) + 2\; 0 = \mathrm{E}((a-\mathrm{E}(Y))^2) + \mathrm{Var}(Y)$. ∎
Si ricordi $\mathrm{MSE}(T,\theta) = \mathrm{E}((T-\theta)^2) = \mathrm{Bias}_\theta(T)^2 + \mathrm{Var}(T)$ che raggiunge il minimo se $\mathrm{Bias}_\theta(T) = 0$ cioè se $\mathrm{E}_\theta(T) = \theta$.

Theorem 5.9 *Se $\theta \in \Theta \subseteq \mathbb{R}$ e $\theta \sim \pi(\theta|\underline{x})$, allora* $\mathrm{E}(Y) = \operatorname{argmin}_{a\in\mathbb{R}} \mathrm{E}((\theta-Y)^2)$ *c.f. funzione di perdita quadratica.*

Proof. Nel lemma si considera $T = \theta$ con legge $\pi(\theta|\underline{x})$. ∎

Esercizio 5.10 Dimostrare il teorema analogo per $\theta \in \Theta \subseteq \mathbb{R}$.

Esempio 5.11 (cont. Esempio 5.3) Si è visto che $\theta|\underline{X} \sim \text{Beta}(a+s, n+b-s)$ e quindi si ha

$$\mathrm{E}(\theta|\underline{X}) =^{10} \frac{a+s}{a+n+b} = \frac{a+b}{a+n+b}\frac{a}{a+b} + \frac{n}{a+n+b}\frac{s}{n} = \frac{a+b}{a+n+b}\mathrm{E}(\pi(\theta)) + \frac{n}{a+n+b}\bar{x}$$

Quindi il valore atteso a posteriori di θ è una combinazione convessa (media pesata) del valore atteso a priori e della stima di massima verosimiglianza ed il limite per taglia campionaria infinita è la stima di massima verosimiglianza, ovvero asintoticamente MLE e valore atteso Bayesiano coincidono.

Esempio 5.12 (cont. Esempio 5.4) Si è visto che $\mu|\underline{x} \sim \mathcal{N}\left(\frac{\tau_0\mu_0 + n\tau\bar{x}}{\tau_0+n\tau}, \frac{1}{\tau_0+n\tau}\right)$ e quindi si ha

$$\mathrm{E}\left(\mu\underline{x}\right) = \frac{\tau_0}{\tau_0+n\tau}\mu_0 + \frac{n\tau}{\tau_0+n\tau}\bar{x}$$

nuovamente il valore atteso a posteriori è una media pesata di valore atteso a priori e stima di massima verosimiglianza: $\lim_{n\mapsto+\infty} \mathrm{E}(\mu|\underline{X}=\underline{x}) = \bar{x}$.

E' un fatto generale, che non dimostriamo, che $\mathrm{E}(\theta|\underline{X})$ converge in legge allo stimatore di massima verosimiglianza di θ.

Esempio 5.13 (da All of Statistics di L. Wasserman Esempio 11.7) In un esperimento vi sono n_1 pazienti "controllo" di cui X_1 sopravvivono con probablità p_1 e n_2 pazienti "trattamento" di cui X_2 sopravvivono con probablità p_2. Si vuole stimare il contrasto $g(p_1,p_2) = p_2 - p_1$. Supponiamo $X_1 \sim \text{Binomial}(n_1, p_1)$ e $X_2 \sim \text{Binomial}(n_2, p_2)$ con spazio dei parametri $(p_1,p_2) \in [0,1]^2$ e $\pi(p_1,p_2) = 1$ a priori costante. Quindi si ha

$$\pi(p_1,p_2|x_1,x_2) = \pi(p_1,p_2)\mathcal{L}(p_1;x_1)\mathcal{L}(p_2;x_2) \propto 1\; p_1^{x_1}(1-p_1)^{n_1-x_1}p_2^{x_2}(1-p_2)^{n_2-x_2}$$

[10] Per $X \sim \text{Beta}(a,b)$ si ha $\mathrm{E}(X) = a/(a+b)$.

dove si suppone che X_1 e X_2 siano indipendenti. Quindi a posteriori p_1 e p_2 sono indipendenti e

$$p_1|x_1 \sim \text{Beta}(x_1+1, n_1-x_1+1)$$
$$p_2|x_2 \sim \text{Beta}(x_2+1, n_2-x_2+1)$$

Il frequentista stima g con $\frac{x_2}{n_2}-\frac{x_1}{n_1}$, il Bayesiano stima di g come $\mathrm{E}(p_1|x_1)-\mathrm{E}(p_2|x_2) = \frac{x_2+1}{n_2+2}-\frac{x_1+1}{n_1+2}$ oppure stimula B realizzazioni da $p_1|x_1$, say $p_{1,1},\dots,p_{1,B}$ e B realizzazioni da $p_2|x_2$, say $p_{2,1},\dots,p_{2,B}$ considera $\tau_b = p_{2,p}-p_{1,b}$ per $b=1,\dots,B$ e sceglierà e.g. $\bar{\tau}_b$ come stima di g.

Esempio 5.14 (da All of Statistics di L. Wasserman Esempio 6.14 (Bayesiano batte frequentista)) Ancora IC Bayesiani. Ricordare che gli intervalli di confidenza non sono "affermazioni" probabilistiche sul parametro. Esempio Berger and Wolpart 1984.

Approccio frequentista Siano X_1 e X_2 uniformi a valori in $\pm$ indipendenti con densità di probablità e per $\theta \in \mathbb{R}$ siano $Y_i = X_i + \theta$ per $i=1,2$. Solo Y_1 e Y_2 sono osservate. Il seguente è un intervallo stocastico (di confidenza) per θ

$$C(Y_1,Y_2) = \begin{cases} \{Y_1-1\} & \text{se } Y_1 = Y_2 \text{ sse } X_1 = X_2 \\ \{s\dfrac{Y_1+Y_2}{2}\} & \text{se } Y_1 \neq Y_2 \text{ sse } X_1 \neq X_2 \end{cases}$$

Si ha $\mathrm{P}_\theta(\theta \in C) = 3/4$ e quindi la realizzazione di $C(y_1,y_2)$ è un IC al 75% per θ.

Si noti inoltre che se $X_1 = X_2 = -1$ allora $C = \{\theta - 2\}$ e non contiene θ, e che se $X_1 = X_2 = 1$ o $X_1 \neq X_2$ allora $C = \{\theta\}$. Si osservano $Y_1 = 15$ e $Y_2 = 17$ e quindi $C_{osservato} = \{16\}$ ma a questo punto siamo certi che $\theta = 16$ ovviamente si può dire che $\{16\}$ è IC per θ al 75% ma l'affermazione probabilistica potrebbe essere $\mathrm{P}_\theta(\theta \in C \big| Y_1 = 15, Y_2 = 17) = 1$.

Approccio Bayesiano Sia $\theta \in \mathbb{Z}$ e $\pi(\theta) > 0$ una qualunque apriori.

Si osserva $y = (y_1,y_2) = (15,17)$ che implica $X_1 = -1$, $X_2 = 1$ (evento di probabilità 1/4) e $\theta = 16$ e quindi la funzione di verosimiglianza vale $\mathcal{L}(\theta|y) = \begin{cases} 1/4 & \text{se } \theta = 16 \\ 0 & \text{altrimenti} \end{cases}$.

Applicando il teorema di Bayes si ha

$$\pi(\theta|Y_1 = 15, Y_2 = 17) = \begin{cases} \pi(\theta)\,1/4 & \text{se } \theta = 16 \\ 0 & \text{altrimenti} \end{cases} = \begin{cases} 1 & \text{se } \theta = 16 \\ 0 & \text{altrimenti} \end{cases} \qquad \text{affinché integri ad 1}$$

Quindi rispetto alla legge a posteriori di θ si ha $\pi_{\theta|y}(\theta \in C) = \pi_\theta(\theta \in C \big| Y_1 = 15, Y_2 = 17) = 1$

Esempio 5.15 (da All of Statistics di L. Wasserman Esempio 11.9.) Questo è un esempio semplificato da Robins and Ritov (1997), Statistics in Medicine. I metodi Bayesiani dipendono dalla funzione di verosimiglianza, MA in alta dimensione (e anche in problemi non parametrici) la verosimiglianza può portare ad inferenze poco accurate.
I dati sono n copie i.i.d. di triplette $(X_1,R_1,Y_1),\dots,(X_n,R_n,Y_n)$. Sia B finito e molto alto (e.g. $B = 100^{100}$ e $n << B$. Sia $\theta = (\theta_1,\dots,\theta_B)$ con $0 \le \theta_j \le 1$ un vettore di parametri non noti e sia $\xi = (\xi_1,\dots,\xi_B)$ numeri *noti* tali che $0 < \delta \le \xi_j \le 1-\delta < 1$ per ogni $j = 1,\dots,B$ con δ noto.
Ciascun dato (X_i,R_i,Y_i) è ottenuto come segue

1. simulare $X_i \sim \text{Unif}(\{1,\dots,B\})$
2. simulare $R_i \big| X_i \sim \text{Bernoulli}(\xi_{X_i})$ (R_i indicatore di i nel campione)
3. se $R_i = 1$ allora $Y_i \big| R_i, X_i \sim \text{Bernoulli}(\theta_{X_i})$, se $R_i = 0$ non simulare Y_i.

Potrebbe essere un modello per alcuni problemi con dati mancanti: $R_i = 0$ indica che l'osservazione i-sima è mancante. L'obiettivo è stimare $\psi = \mathrm{P}(Y_i = 1) = \sum_{j=1}^B \mathrm{P}(Y_i = 1 \big| X_i = j)\,\mathrm{P}(X_i = j) = \frac{1}{B}\sum_{j=1}^B \theta_j$ perché per ogni i,j vale $\mathrm{P}(X_i = j) = 1/B$ e $\mathrm{P}(Y_i \big| X_i = j) = \mathrm{P}(Y_i i = 1 \big| R_i = 1, X_i = j) = \mathrm{P}(Y_i = 1 X X_i = j)$ (se $R_i = 0$ non vado ad osservare Y_i.

approccio Bayesiano la funzione di verosimiglianza per una singola osservazione è

$$\mathcal{L}(\theta \big| x_i, r_i, y_i) = f(x_i)f(r_i|x_i)f(y_i|x_i)^{r_i} = \frac{1}{B}\xi_{x_i}^{r_i}(1-\xi_{x_i})^{1-r_i}\left(\theta_{x_i}^{y_i}(1-\theta_{x_i})^{1-y_i}\right)^{r_i}$$

e per l'n-campione è

$$\mathcal{L}(\theta) = \prod_{i=1}^{n}\mathcal{L}(\theta \big| x_i, r_i, y_i) = \frac{1}{B^n}\prod_i \xi_{x_i}^{r_i}(1-\xi_{x_i})^{1-r_i}\left(\theta_{x_i}^{y_i}(1-\theta_{x_i})^{1-y_i}\right)^{r_i} \propto \prod_i \theta_{x_i}^{r_i y_i}(1-\theta_{x_i})^{r_i(1-y_i)}$$

trascurando fattori non dipendenti da θ. A meno di costante additiva la funzione di log-verosimiglianza è

$$l(\theta) = \sum_{i=1}^{n} y_i r_i \log\theta_{x_i} + \sum_{i=1}^{n}(1-y_i)r_i\log(1-\theta_{x_i}) = \sum_{j=1}^{B} n_j\log\theta_j + \sum_{j=1}^{B} m_j\log(1-\theta_j)$$

dove per $j = 1, \ldots, B$

$$n_j = \#\{i \in \{1,\ldots,B\} : Y_i = 1, R_i = 1, X_i = j\} \quad \text{e} \quad m_j = \#\{i \in \{1,\ldots,B\} : Y_i = 0, R_i = 1, X_i = j\}$$

Si noti che poiché $B >> n$ molti n_j e m_j sono nulli. Inoltre

1. lo stimatore di massima verosimiglianza per tanti θ_j non è definito
2. per molti θ_j la distribuzione a posteriori è uguale a quella a priori poiché θ_j non compare in $\mathcal{L}$, ovvero $\pi(\theta \big| Dati) \equiv \pi(\theta)$
3. da cui anche $\pi(\psi \big| Dati) \equiv \pi(\psi)$

Conclusione: l'analisi Bayesiana è poco informativa.

approccio frequentista Si stima ψ con $\hat{\psi} = \frac{1}{n}\sum_{i=1}^{n}\frac{R_i Y_i}{\xi_{X_i}}$ e vale $\mathrm{E}(\hat{\psi}) = \psi$ e $\mathrm{Var}(\hat{\psi}) \le \frac{1}{n\delta^2}$ da cui l'MSE tende a zero (se n cresce) con ordine $1/n$. Lo stimatore di Horwitz-Thompson non può essere ottenuto con metodi basati sulla verosimiglianza perché coinvolte gli ξ_{X_i}! .

Conclusione: se la verosimiglianza è poco informativa (piatta, problemi in alta dimensione,...) i metodi Bayesiani possono risultare poco utili.

5.3 Proprietà asintotiche degli stimatori Bayesiani

5.4 Aspetti computazionali

5.5 Strenghts and weaknesses of Bayesian approach

6 Appendici

6.1 Ripasso di probabilità

Sia X variabile aleatoria discreta a valori reali con densità di probabilità f_X.

1. La funzione di ripartizione di f_X è $F_X(x) = \mathrm{P}(X \leq x) = \sum_{y \leq x} f_X(y)$.
2. Il valore atteso è $\mathrm{E}(X) = \sum_x x f_X(x)$ se la sommatoria converge.
3. La varianza è $\mathrm{Var}(X) = \mathrm{E}\left((X - \mathrm{E}(X))^2\right) = \sum_x (x - \mathrm{E}(X))^2 f_X(x)$ se le sommatorie sono convergenti. Se $\mathrm{Var}(X)$ esiste finita, vale $\mathrm{Var}(X) = \mathrm{E}(X^2) - \mathrm{E}(X)^2 = \sum_x x^2 f_X(x) - \left(\sum_x f_X(x)\right)^2$.

La convarianza tra le v.a. X e Y con densità congiunta $f_{X,Y}$ è definita da $\mathrm{Cov}(X) = \mathrm{E}\left((X - \mathrm{E}(X))(Y - \mathrm{E}(Y))\right) = \sum_{x,y} (x - \mathrm{E}(X))(y - \mathrm{E}(Y)) f_{X,Y}(x, y)$.
Leggi discrete notevoli includono

1. Bernoulli(p) con $p \in [0, 1]$, $f_X(x) = p^x (1 - p)^{1-x}$ con $x = 0, 1$, $\mathrm{E}(X) = p$, $\mathrm{Var}(X) = p(1 - p)$.
2. Binomiale(n, p) con $p \in [0, 1]$ e n intero positivo, $f_X(x) = \binom{n}{x} p^x (1-p)^{1-x}$ con $x = 0, 1, \ldots, n$; $\mathrm{E}(X) = np$, $\mathrm{Var}(X) = np(1 - p)$.

   Se $X_1, \ldots, X_n$ i.i.d. Bernoulli(p) allora $\sum_{i=1}^n X_i \sim$ Binomiale(n, p).
3. Poisson(λ) con $\lambda > 0$, $f_X(x) = e^{-\lambda} \dfrac{\lambda^x}{x!}$; $\mathrm{E}(X) = \lambda$, $\mathrm{Var}(X) = \lambda$.
4. Geometrica(p) con $f_X(x) = p(1 - p)^{x-1}$ con $x = 1, 2, \ldots$; $\mathrm{E}(X) = \dfrac{1}{p}$, $\mathrm{Var}(X) = \dfrac{1 - p}{p^2}$.

Esercizio 6.1 Verificare i risultati precedenti.

Sia X una v.a. continua a valori reali che ammette densità di probabilità f_X. Si ha $\mathrm{P}(a \leq X \leq b) = \int_a^b f_X(x)\,dx$ per $a < b$ reali, per la funzione di ripartizione si ha $F_X(x) = \mathrm{P}(X \leq x) = \int_{-\infty}^x f_X(y)\,dy$ con $x \in \mathbb{R}$, per il valor medio si ha $\mathrm{E}(X) = \int_{\mathbb{R}} x f_X(x)\,dx$ e per la varianza $\mathrm{Var}(X) = \int_{\mathbb{R}} x^2 f_X(x)\,dx - \mathrm{E}(X)^2$ se tutti gli integrali sono finiti. Per due variabili aleatorie X e Y di densità congiunta $f_{X,Y}$ si ha $\mathrm{Cov}(X, Y) = \mathrm{E}\left((X - \mathrm{E}(X))(Y - \mathrm{E}(Y))\right) = \int_{\mathbb{R}^2} (x - \mathrm{E}(X))(y - \mathrm{E}(Y)) f_{X,Y}(x, y)\,dx\,dy$.
Leggi continue notevoli includono:

1. Uniforme($[a, b]$) con $a < b$, $f_X(x) = \dfrac{1}{b - a}$ con $x \in [a, b]$, $\mathrm{E}(X) = \dfrac{a + b}{2}$, $\mathrm{Var}(X) = \dfrac{(b - a)^2}{12}$.
2. Esponenziale(λ) con $\lambda \in \mathbb{R}$, $f_X(x) = \lambda \exp(-x\lambda)$ con $x \in \mathbb{R}$, $\mathrm{E}(X) = \dfrac{1}{\lambda}$, $\mathrm{Var}(X) = \dfrac{1}{\lambda^2}$.

   Se $X_1, \ldots, X_n$ i.i.d. Esponenziale(λ) allora $\sum_{i=1}^n X_i \sim \Gamma(n, \lambda)$.
3. Normale(μ, σ^2) con μ reale e $\sigma^2 > 0$, $f_X(x) = \dfrac{1}{\sqrt{2\pi\sigma^2}} \exp(-\dfrac{(x - \mu)^2}{2\sigma^2})$; $\mathrm{E}(X) = \mu$, $\mathrm{Var}(X) = \sigma^2$.

   Se $X_1 \sim$ Normale(μ_1, σ_1^2) e $X_2 \sim$ Normale(μ_2, σ_2^2) indipendenti e $a, b \in \mathbb{R}$ allora $aX_1 + bX_2 \sim$ Normale($a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2$);

   se $X_1, \ldots, X_n$ i.i.d. Normale($0, 1$) allora $\sum_{i=1}^n X_i^2 \sim \chi_n^2$;

   se $X \sim$ Normale($0, 1$) e $U \sim \chi_n^2$ indipendenti allora $V = \dfrac{X}{\sqrt{\frac{U}{n}}} \sim \mathrm{t}_n$.

Vale $\mathrm{E}(X + Y) = \mathrm{E}(X) + \mathrm{E}(Y)$ e $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X, Y)$, inoltre se X e Y sono indipendenti $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$.

6.2 Convergenze

Sia $\{X_n\}_n$ una successione di vettori aleatori a valori in $\mathbb{R}^v$ e definiti sullo spazio di probabilità $(\Omega, \mathcal{A}, \mathrm{P})$ e X un vettore aleatorio sugli stessi spazi. Siano F_n e F le funzioni di ripartizione di X_n e X rispettivamente.

1. X_n converge a X quasi certamente (almost surely) se $\mathrm{P}\left(\{\omega \in \Omega : \lim_{n \mapsto +\infty} X_n(\omega) = X(\omega)\}\right) = 1$
2. X_n converge a X in L^p (p^{th} mean) se $\lim_{n \mapsto +\infty} \mathrm{E}\left(\|X_n - X\|^p\right) = 0$ per $p \geq 1$ ($\| \cdot \|$ denota la norma euclidea in $\mathbb{R}^v$)
3. X_n converge a X in probabilità se $\lim_{n \mapsto +\infty} \mathrm{P}\left(\{\omega : |X_n(\omega) - X(\omega)| < \epsilon\}\right) = 1$ per ogni $\epsilon > 0$
4. X_n converge a X in distribuzione se $\lim_{n \mapsto +\infty} F_n(x) = F(x)$ per ogni x in cui F è continua.

Theorem 6.2 $X_n \xrightarrow{L_p} X$ *se e solo se* $\lim_{n \mapsto +\infty} \mathrm{E}\left(\dfrac{\|X_n - X\|}{1 + \|X_n - X\|}\right) = 0.$

Per la dimostrazione vi veda Jacod Protter Theorem 17.1 pagina 143. Un risultato analogo vale se a $f(x) = \frac{x}{x+1}$ si sostituisce una funzione f continua, limitata, strettamente crescente in $[0, +\infty)$ e con $f(0) = 0$.

1. La convergenza in L^p implica la convergenza in probabilità.
2. La convergenza quasi certa implica la convergenza in probabilità.
3. La convergenza in probabilità implica la convergenza in distribuzione.
4. Se la variabile limite è una costante c, cioè $\mathrm{P}(X = c) = 1$, allora la convergenza in distribuzione implica la convergenza in probabilità.

Per X_i, $i = 1, 2, \ldots$, i.i.d. con $\mathrm{E}(X_i) = \mu$ e $\bar{X_n} = \dfrac{\sum_{i=1}^n X_i}{n}$, si ha

1. Legge debole dei grandi numeri: $\bar{X_n} \longrightarrow \mu$ in probabilità, cioè $\lim_{n \mapsto +\infty} \mathrm{P}\left(|\bar{X_n} - \mu| < \epsilon\right) = 1$.
2. Legge forte dei grandi numeri: $\bar{X_n} \longrightarrow \mu$ quasi certamente, cioè $\mathrm{P}\left(\lim_{n \mapsto +\infty} \bar{X_n} = \mu\right) = 1$.
3. Teorema del limite centrale: Se inoltre $\mathrm{Var}(X_i) = \sigma^2$, allora $\dfrac{\bar{X_n} - \mu}{\sqrt{n\sigma^2}} \longrightarrow \mathcal{N}(0, 1)$ in distribuzione.

6.3 Funzione caratteristica

La funzione caratteristica di una distribuzione di probabilità F_X di un vettore aleatorio X n-dimensionale è definita come

$$\Phi_X(t) = \mathrm{E}(e^{i\langle t, X\rangle}) \qquad t \in \mathbb{R}^n$$

con i unità immaginaria. La funzione caratteristica esiste per ogni F_X e caratterizza F_X. Se i momenti di F_X esistono finiti, la funzione caratteristica è utile per calcolarli.

Theorem 6.3 *Sia* $\{X_k\}_{k \in \mathbb{Z}_{>0}}$ *una successione di vettori aleatori sullo stesso spazio di probabilità. Se in un intorno di zero esiste* $\lim_{k \to +\infty} \Phi_{X_k}(t) = \Phi_X(t)$ *e* Φ_X *è funzione caratteristica, allora nei punti di continuità di* F_X *vale* $\lim_{k \to} F_{X_k}(x) = F_X(x)$.

Questo teorema è usato nella dimostrazione del teorema (forte) del limite centrale.
Per X e Y variabili aleatorie indipendenti $\Phi_{X+Y}(t) = \Phi_X(t)\Phi_Y(t)$. Se le distribuzioni di probabilità di X e Y ammettono funzione di densità allora $f_{X+Y}(x) = \int f_X(y) f_Y(x - y)\, dy$ (convoluzione delle densità).

6.4 Densità e speranze condizionate

***** aggiungere lista delle proprietà delle speranze conditionate ***** Si veda il Capitolo 9.7 in David Williams, Probability with Martingales, Cambridge Mathematical Textbooks (1991).

Esempio 6.4 Siano X_1, X_2 indipendenti e distribuite secondo una legge Bernoulli(p) con $p \in]0,1[$. Calcolare la legge di $S = X_1 + X_2$. Ci sono almeno tre modi per affrontare questo esercizio.

1. Da studi precedenti sappiamo $S \sim \text{Binomial}(2,p)$.
2. Il prodotto delle funzioni caratteristiche $\Phi_{X_1}(t)\Phi_{X_2}(t) = (1-p+pe^{it})^2$ è la funzione caratteristica di una binomiale.
3. Direttamente: S assume valori $0,1,2$; $S=0$ se e solo se $X_1 = X_2 = 0$. Questo evento ha probabilià $(1-p)^2$ per l'indipendenza di X_1 e X_2; $S=1$ se e solo se $X_1 = 0, X_2 = 1$ o $X_1 = 1, X_2 = 0$, questi eventi hanno probabilità $p(1-p)$ ciascuno, sono disgiunti e quindi $S=1$ ha probabilità $2p(1-p)$; $S=2$ avviene con probabilità $1-(1-p)^2 - 2p(1-p) = p^2$.

Determinare la densità condizionata di X_1 dato S. La densità congiunta di (X_1, S) è data nella seguente tabella

	0	1	2	
0	$(1-p)^2$	$p(1-p)$	$0\|1-p$	
1	0	$(1-p)p$	p^2	p
	$(1-p)^2$	$2p(1-p)$	p^2	

Quindi

$$f_{X_1|S=0}(1) = (1-p)^2/(1-p)^2 = 1 \text{ e quindi } f_{X_1|S=0}(0) = 1 - f_{X_1|S=0}(1) = 0$$
$$f_{X_1|S=1}(1) = (1-p)p/2(1-p)p = 1/2 \text{ e quindi } f_{X_1|S=1}(0) = 1/2$$
$$f_{X_1|S=2}(1) = 0/p^2 = 0 \text{ e quindi } f_{X_1|S=1}(0) = 1$$

Esempio 6.5 Siano $X_1, \ldots, X_n$ i.i.d. Bernoulli(p). Calcolare la legge condizionata di X_1 dato $S = \sum_{i=1}^n X_i$. La legge congiunta di (X_1, S) è determinata come

$$\begin{aligned} f_{X_1,S}(x_1, s) &= \mathrm{P}(X_1 = x_1, S = s) = \mathrm{P}(X_1 = x_1, X_2 + \cdots + X_n = s - x_1) \\ &= \begin{cases} \mathrm{P}(X_1 = x_1)\,\mathrm{P}(X_2 + \cdots + X_n = s - x_1) & \text{se } s \geq x_1 \\ 0 & \text{se } s < x_1 \end{cases} \end{aligned}$$

Ora $X_2 + \cdots + X_n \sim \text{Binomial}(n-1, p)$ e quindi se $s \geq x_1$

$$f_{X_1,S}(x_1, s) = p^{x_1}(1-p)^{1-x_1}\binom{n-1}{s-x_1}p^{s-x_1}(1-p)^{n-1-s+x_1} = \binom{n-1}{s-x_1}p^s(1-p)^{n-s}$$

Si deduce che

$$f_{X_1|S}(x_1) = \frac{\binom{n-1}{s-x_1}p^s(1-p)^{n-s}}{\binom{n}{s}p^s(1-p)^{n-s}} = \begin{cases} \dfrac{s}{n} & \text{se } x_1 = 1 \\ \dfrac{n-s}{s} & \text{se } x_1 = 0 \end{cases}$$

Esempio 6.6 Nelle ipotesi dell'esercizio precedente calcolare la legge congiunta di $(X_1, \ldots, X_n)$ condizionata a S.

$$f_{(X_1,\ldots,X_n,S)}(x_1, \ldots, x_n, s) = \begin{cases} 0 & \text{se } \sum_{i=1}^n x_i \neq s \\ \prod_{i=1}^n p^{x_i}(1-p)^{1-x_i} = p^s(1-p)^{n-s} & \text{se } \sum_{i=1}^n x_i = s \end{cases}$$

quindi

$$f_{(X_1,\ldots,X_n|S=s)}(x_1, \ldots, x_n) = \begin{cases} 0 & \text{se } \sum_{i=1}^n x_i \neq s \\ \dfrac{p^s(1-p)^{n-s}}{\binom{n}{s}p^s(1-p)^{n-s}} & \text{se } \sum_{i=1}^n x_i = s \end{cases}$$

6.4.1 Indipendenza due-a-due e indipendenza stocastica

Sia $(\Omega, \mathcal{A}, \mathrm{P})$ uno spazio di probabilità e $A, B, C \in \mathcal{A}$ tre eventi tali che $\mathrm{P}(A) = \mathrm{P}(B) = \mathrm{P}(C) = \frac{1}{2}$, $\mathrm{P}(A \cap B) = \mathrm{P}(B \cap C) = \mathrm{P}(C \cap A) = \frac{1}{4}$ e $\mathrm{P}(A \cap B \cap C) = \frac{1}{6} \neq \frac{1}{2^3}$. Essi sono eventi indipendenti due-a-due ma non tre-a-tre (verificarlo e.g. con un diagramma di Venn).
Gli eventi $A, B, C \in \mathcal{A}$ tali che $\mathrm{P}(A) = \frac{3+3+2}{16} = \frac{1}{2} = \mathrm{P}(B) = \mathrm{P}(C)$, $\mathrm{P}(A \cap B) = \frac{5}{16} \neq \mathrm{P}(A)\,\mathrm{P}(B)$ e $\mathrm{P}(A \cap B \cap C) = \frac{2}{16} = \frac{1}{8} = \mathrm{P}(A)\,\mathrm{P}(B)\,\mathrm{P}(C)$ sono un esempio del fatto che l'indipendenza tre-a-tre non implica l'indipendenza due-a-due (verificarlo e.g. con un diagramma di Venn).
Quanto sopra è un'istanza del paradosso di Simpson. Si osservi invece che se il vettore aleatorio (X, Y, Z) ammette densità congiunta $f_{X,Y,Z}$ integrabile, continua, definita su uno spazio prodotto e tale che $f_{X,Y,Z} = f_X f_Y f_Z$ allora $f_{X,Y} = f_X f_Y$.

6.5 Distribuzione normale multivariata

file 7gauss.pdf

- Normale standard: $Z \sim \mathcal{N}_k(0, I)$ se $f_Z(z_1, \dots, z_k) = \frac{1}{\sqrt{2\pi}^k} \exp(-\frac{1}{2} \sum_{i=1}^{k} z_i^2)$. In particolare, per $B \subset \mathbb{R}^k$

$$\mathrm{Prob}(Z \in B) = \int_B f_Z(z_1, \dots, z_k)\, dz_1 \dots dz_k$$

- Trasformazioni affini: per $\mu \in \mathbb{R}^n$ e $A \in \mathcal{A}^{n \times k}(\mathbb{R})$ il vettore aleatorio $X = \mu + AZ \in \mathbb{R}^n$ segue una distribuzione normale di media μ e matrice di varianza-covarianza AA^t: $\mathcal{N}_n(\mu, AA^t)$ e si pone $\Sigma = AA^t$. Se $|\Sigma| \neq 0$ (dove $|A|$ è il determinante di A), la densità di X è $f_X(x) = \frac{1}{(2\pi|\Sigma|)^{1/2}} \exp(-\frac{1}{2}(x-\mu)^t \Sigma^{-1} (x-\mu))$. Se esiste, $K = \Sigma^{-1}$ è detta matrice di concentrazione.

  Esercizio 6.7 Determinare le matrici A che selezionano vettori marginali di X e determinare la legge di probabilità di questi vettori.

- Partizionare X non degenere in $(X_1, X_2) \sim \mathcal{N}_n\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}^t & \Sigma_{22} \end{pmatrix}\right)$. Allora

$$X_1 | (X_2 = x_2) \sim \mathcal{N}_{n_1}\left(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right)$$

  Esercizio 6.8 – Siano $X \sim \mathcal{N}(0,1)$, $W = 1$ con probabilità $1/2$ e $W = -1$ con probabilità $1/2$ e $Y = XW$. Verificare che X e Y hanno la stessa legge, non sono correlate, ma sono dipendenti.
  - Rifare l'esercizio precedente con $Y = -X$ se $|X| < c$ e $Y = X$ altrimenti, con $c > 0$.
  - Verificare che X_2 e $X_1 - \Sigma_{12}\Sigma_{22}^{-1}(X_2 - \mu_2)$ sono indipendenti.
  - Sia $X \sim \mathcal{N}(0,1)$ e $Y = -X$. Verificare che (X, Y) non è un vettore Gaussiano di $\mathbb{R}^2$ non-degenere, [marginali gaussiane ma vettore non gaussiano].
  - Indichiamo, senza dimostrarlo, che per $v, w \in \{1, \dots, n\}$ $\Sigma_{vw} = 0$ implica l'indipendenza tra X_v e X_w, ovvero e $(\Sigma^{-1})_{vw} = 0$ implica l'indipendenza di X_v e X_w condizionatamente alle altre variabili, ovvero $X_v \perp\!\!\!\perp X_w \,|\, (X_{-vw})$ dove $-vw = \{1, \dots, n\} \setminus \{v, w\}$.

- Interpretazione geometrica: le curve di livello della densità di $X \sim \mathcal{N}_n(\mu, \Sigma)$ con Σ di rango pieno, sono elissi centrate nella media e per le quali le direzioni degli assi principali sono date dagli autovettori di Σ. In particolare se $\Sigma = U\Lambda U^t$ dove Λ è la matrice diagonale degli autovalori di Σ e le colonne di U sono autovettori unitari di Σ, allora $X \sim \mu + U\mathcal{N}_n(0, \Lambda)$ ovvero ruotare un vettore normale con la matrice di rotazione U e traslarlo di μ.

- Momenti di ordine superiore: sia $X \sim \mathcal{N}_n(\mu, \Sigma)$. Per $r_1, \ldots, r_n$ interi non negativi e $r_1 + \ldots + r_n = k$, sia $\mu_{r_1,\ldots,r_n}(X) = \mathrm{E}(X_1^{r_1} \ldots X_n^{r_n})$. Se k è dispari $\mu_{r_1,\ldots,r_n}(X) = 0$, se $k = 2\lambda$ $\mu_{r_1,\ldots,r_n}(X) = \sum \sigma_{ij}...\sigma_{lm}$ dove la somma è su tutti i sottoinsiemi di $\{1, \ldots, 2\lambda\}$ contenenti λ coppie non ordinate. [riscrivere] E.g.

$$\begin{aligned} E\left[X_i^4\right] &= 3\sigma_{ii}^2 \\ E\left[X_i^3 X_j\right] &= 3\sigma_{ii}\sigma_{ij} \\ E\left[X_i^2 X_j^2\right] &= \sigma_{ii}\sigma_{jj} + 2\left(\sigma_{ij}\right)^2 \\ E\left[X_i^2 X_j X_k\right] &= \sigma_{ii}\sigma_{jk} + 2\sigma_{ij}\sigma_{ik} \\ E\left[X_i X_j X_k X_n\right] &= \sigma_{ij}\sigma_{kn} + \sigma_{ik}\sigma_{jn} + \sigma_{in}\sigma_{jk}. \end{aligned}$$

- Siano $h = \Sigma^{-1}\mu$ e $a = -\dfrac{n}{2}\log 2\pi + \dfrac{1}{2}\log\det\Sigma^{-1} - \dfrac{1}{2}\mu^t\Sigma^{-1}\mu$. Verificare che

$$\begin{aligned} f_X(x) &= \frac{1}{(2\pi|\Sigma|)^{n/2}}\exp(-\frac{1}{2}(x-\mu)^t\Sigma^{-1}(x-\mu)) \\ &= \exp\left(a + h^t x - \frac{1}{2}x^t\Sigma^{-1}x\right) = \exp\left(a + \sum_i^n h_i x_i - \frac{1}{2}\sum_{i,j}^n \left(\Sigma^{-1}\right)_{ij} x_i x_j\right) \end{aligned}$$

- Entropia: sia f la funzione di densità $\mathcal{N}_n(\mu, \Sigma)$. Allora $h(f) = -\int_{\mathbb{R}^n} f(x)\log f(x)\,dx = \frac{1}{2}(n + n\ln 2\pi + \ln|\Sigma|) = \frac{1}{2}\ln((2\pi e)^n|\Sigma|)$.

- **Distribuzione congiunta di $\bar{X}$ e S^2 in un campione normale.** Siano $X_1, \ldots, X_n$ i.i.d. $\mathcal{N}(\mu, \sigma^2)$. Il vettore degli scarti $(X_1 - \mu, \ldots, X_n - \mu)^t$ e la media campionaria $\bar{X} - \mu$ sono indipendenti (si dimostra per esempio usando le funzione caratteristiche), ciò implica l'indipendenza di $\bar{X}$ e S^2 che è funzione degli scarti. Una dimostrazione alternativa è come segue: vale $\mathbb{X} = (X_1, \ldots, X_n) \sim \mathcal{N}(\boldsymbol{\mu}, \sigma^2\,\mathrm{I}_n)$ dove $\boldsymbol{\mu}$ è il vettore di lunghezza n le cui componenti sono uguali a μ e I_n è la matrice identità $n \times n$. Si consideri la matrice

$$A = \begin{pmatrix} 1/\sqrt{n} & 1/\sqrt{n} & 1/\sqrt{n} & 1/\sqrt{n} & \ldots & 1/\sqrt{n} \\ 1/\sqrt{2-1} & -1/\sqrt{2-1} & 0 & 0 & \ldots & 0 \\ 1/\sqrt{3-2} & -1/\sqrt{3-2} & -1/\sqrt{3-2} & 0 & \ldots & 0 \\ 1/\sqrt{4-3} & -1/\sqrt{4-3} & -1/\sqrt{4-3} & -1/\sqrt{4-3} & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1/\sqrt{n(n-1)} & 1/\sqrt{n(n-1)} & 1/\sqrt{n(n-1)} & 1/\sqrt{n(n-1)} & 1/\sqrt{n(n-1)} & -(n-1)/\sqrt{n(n-1)} \end{pmatrix}$$

e si osservi $W = A\mathbb{X} \sim \mathcal{N}(t, \sigma^2\sigma^2\,\mathrm{I}_n)$ dove t è il vettore n-dimensionale $(\mu\sqrt{n}, 0, 0, \ldots)$. Si consideri la somma

$$\begin{aligned} W_2^2 + W_3^2 + \ldots + W_n^2 = W^tW - W_1^2 &= \mathbb{X}^tA^tA\mathbb{X} - \left(\frac{\sum X_i}{\sqrt{n}}\right)^2 \\ &= \sum_{i=1}^n X_i^2 - n\bar{X}^2 = \sum_{i=1}^n (X_i - \bar{X})^2 = (n-1)S^2 \end{aligned}$$

questo dimostra che S^2 è somma di $W_2, \ldots, W_1$ ed è quindi indipendente da W_1. Poiché $W_1 = \sqrt{n}\bar{X}$, risulta che S^2 e $\bar{X}$ sono indipendenti.

6.6 Un altro contresempio al teorema di Basu

Esercizio 6.9 Nota al teorema di Basu: l'ipotesi di completezza è essenziale oltre all'Esempio 2.44 si consideri $Unif(]\theta, \theta + 1[)$.

6.7 Stimatori intervallari

Sia $\theta \in \Theta \in \mathbb{R}$ e siano $L(\underline{X}), U(\underline{X})$ due statistiche a valori reali tali che per ogni $\underline{x} \in \mathcal{X}$ $L(\underline{x}) \leq U(\underline{x})$. L'intervallo stocastico $[L(\underline{X}), U(\underline{X})]$ è detto *stimatore intervallare* per θ. Se $\theta \in \Theta \subseteq \mathbb{R}^p$, $p > 1$ si parla di *regioni di confidenza.*

Esempio 6.10 Siano $X_1, \dots, X_n$ i.i.d. $\mathcal{N}(\mu, \sigma^2)$ con σ^2 noto e $\mu \in \mathbb{R}$. Allora $\bar{X} \sim \mathcal{N}(\mu, \sigma^2/n)$ e $Z \sim \dfrac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}}$ e per $\alpha \in]0,1[$ esiste $z_\alpha \in \mathbb{R}_{>0}$ tale che $\alpha = \mathrm{P}\left(|Z_\alpha| \leq z_\alpha\right)$ da cui

$$\alpha = \mathrm{P}\left(|\frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}}| \leq z_\alpha\right) = \mathrm{P}\left(\bar{X} - z_\alpha \sigma/\sqrt{n} \leq \mu \leq \bar{X} + z_\alpha \sigma/\sqrt{n}\right)$$

6.8 Semipositività delle matrici di varianza-covarianza

Nel Teorema 3.19 si è utilizzato il fatto che matrici di varianza-covarianza sono semidefinite positive, infatti per un vettore aleatorio $\underline{X} \in \mathbb{R}^n$ si ha

$$\mathrm{VarCov}(\underline{X}) = \mathrm{E}\left((\underline{X} - \mathrm{E}(\underline{X}))\,(\underline{X} - \mathrm{E}(\underline{X}))^t\right) = \mathrm{E}\left(YY^t\right) = \mathrm{VarCov}(Y)$$

per $Y = \underline{X} - \mathrm{E}(\underline{X})$. Ora sia $\underline{x} \in \mathbb{R}^n$

$$\underline{x}^t\,\mathrm{VarCov}(Y)\underline{x} = \mathrm{E}\left(\underline{x}^t Y Y^t \underline{x}\right) = \mathrm{E}\left(\left(Y^t\underline{x}\right)^t \left(Y^t \underline{x}\right)\right) = \mathrm{E}\left(Z^t Z\right) = \mathrm{E}\left(\sum_{i=1}^n Z_i^2\right) \geq 0 \qquad \text{evidentemente}$$

per $Z = Y^t \underline{x}$.

6.9 Stimatori consistenti

Definizione 6.11 *Sia* $\underline{X}_n = (X_1, \dots, X_n)$ *un campione statistico e* $T_n(\underline{X}_n)$ *uno stimatore di* $\theta \in \Theta \subseteq \mathbb{R}$. *La (famiglia di) stimatori* T_n, $n \in \mathbb{Z}_{>0}$ è consistente per θ *se per ogni* $\epsilon, \eta > 0$ *esiste* $N_{\epsilon,\eta} \in \mathbb{Z}_{>0}$ *tale che per ogni* $n > N_{\epsilon,\eta}$ *vale* $\mathrm{P}_\theta(|T_n - \theta| < \epsilon) > 1 - \eta$.

Esercizio 6.12 Si scriva la definizione di stimatore consistente di un parametro multidimensionale.

Osservazioni 6.13

1. La famiglia di stimatori $\{T_n\}_n$ è consistente se converge in probabilità a θ.
2. Dalla diseguaglianza di Čebyšëv segue che

$$\mathrm{P}_\theta(|T_n - \theta| \geq \epsilon) \leq \frac{\mathrm{E}_\theta\left((\hat{\theta}_n - \theta)^2\right)}{\epsilon^2} = \frac{\mathrm{MSE}(\theta, \hat{\theta}_n)}{\epsilon^2}$$

   Se $\lim_{n \mapsto +\infty} MSE = 0$ allora $\{\hat{\theta}_n\}_n$ è consistente per θ. Se entrambi $\mathrm{Var}(\hat{\theta}_n)$ e $\mathrm{Bias}(\hat{\theta}_n)$ convergono a zero per n che tende a più infinito, allora $\{\hat{\theta}_n\}_n$ è consistente per θ.

3. Stimatori di massima verosimiglianza sono consistenti (si veda ....). In particolare in un modello di classe esponenziale $\{\sum_{i=1}^n T(\underline{X}_i)\}_{n=1}^{+\infty}$ è consistente per $\nabla_\theta \psi$.

Esempio 6.14

1. Per una campione casuale $X_1, \dots, X_n$ da una densità di Cauchy $f(x) = \frac{1}{\pi(1+(x-\theta)^2)}$ con $\theta \in \mathbb{R}$, la media campionaria non è stimatore consistente di θ perché non esiste il suo valore atteso infatti $\bar{X}_n$ ha la stessa legge di X_1. (Lo si dimostri ricordando che la funzione caratteristica di X_1 è $\Phi_{X_1}(t) = \exp(-|t|)$).
2. Se esiste il valore atteso di una campione casuale allora $\bar{X}_n$ è stimatore consistente di X_1. Infatti per il teorema del limite centrale vale il seguente risultato: siano $X_1, \dots, X_n$ un campione i.i.d di legge $f(\theta)$ con $\theta \in \Theta$ e T_n una successione di statistiche tali che $\mathrm{E}_\theta(\bar{X}_n) = \theta + o(1/n)$ e $\mathrm{Var}(\bar{X}_n)$ allora $\dfrac{T_n - \theta}{\sqrt{T_n}}$ converge a zero in probabilità e $\{T_n\}_n$ è consistente per θ.

7 Codici R

7.1 Generating mixture of random distributions

A good reference is
https://stats.stackexchange.com/questions/70855/generating-random-variables-from-a-mixture-of-normal-distributions

```
Method 1
library(mixtools)
wait = faithful$waiting
mixmdl = normalmixEM(wait)
plot(mixmdl,which=2)
lines(density(wait), lty=2, lwd=2)
```

```
Method 2
N <- 100000
components <- sample(1:3,prob=c(0.3,0.5,0.2),size=N,replace=TRUE)
mus <- c(0,10,3)
sds <- sqrt(c(1,1,0.1))
samples <- rnorm(N)*sds[components]+mus[components]
hist(samples)
```

Method 3 is based on the following algorithm

1. Generate a continuous random variable U with a uniform distribution in $(0, 1)$
2. If $U \in \left[\sum_{i=1}^{k} p_i, \sum_{i=1}^{k+1} p_i\right)$, where p_k corresponds to the probability of the kth component of the mixture model, then generate from the distribution of the k-th component
3. Repeat steps 1. and 2. until you have the desired amount of samples from the mixture distribution

The code below applies the general algorithm given above, for sampling from a mixture of Gaussian distributions

```
#The number of samples from the mixture distribution
N = 100000
#Sample N random uniforms U
U =runif(N)
#Variable to store the samples from the mixture distribution
rand.samples = rep(NA,N)
#Sampling from the mixture
for(i in 1:N){
  if(U[i]<.3){
    rand.samples[i] = rnorm(1,0,1)
  }else if(U[i]<.8){
    rand.samples[i] = rnorm(1,10,1)
  }else{
    rand.samples[i] = rnorm(1,3,.1)
  }
}

#Density plot of the random samples
plot(density(rand.samples),main="Density Estimate of the Mixture Model")

#Plotting the true density as a sanity check
x = seq(-20,20,.1)
truth = .3*dnorm(x,0,1) + .5*dnorm(x,10,1) + .2*dnorm(x,3,.1)
plot(density(rand.samples),main="Density Estimate of the Mixture Model",ylim=c(0,.2),lwd=2)
```

```
lines(x,truth,col="red",lwd=2)
legend("topleft",c("True Density","Estimated Density"),col=c("red","black"),lwd=2)
```

Method 4 picks one distribution (from k possibilities) with some probability, and then generates pseudo-random variates from that distribution.

```
  set.seed(8)             # this makes the example reproducible
N     = 1000              # this is how many data you want
probs = c(.3,.8)          # these are *cumulative* probabilities; since they
#   necessarily sum to 1, the last would be redundant
dists = runif(N)          # here I'm generating random variates from a uniform
#   to select the relevant distribution

# this is where the actual data are generated, it's just some if->then
#   statements, followed by the normal distributions you were interested in
data = vector(length=N)
for(i in 1:N){
  if(dists[i]<probs[1]){
    data[i] = rnorm(1, mean=0, sd=1)
  } else if(dists[i]<probs[2]){
    data[i] = rnorm(1, mean=10, sd=1)
  } else {
    data[i] = rnorm(1, mean=3, sd=.1)
  }
}

# here are a couple of ways of looking at the results
summary(data)
#    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
# -3.2820  0.8443  3.1910  5.5350 10.0700 13.1600

plot(density(data))
```

8 Nota

Uso questi appunti per parte del corso di Statistica Matematica del corso di laurea in SMID e della laurea Magistrale/Specialistica in Matematica. Gli argomenti sono standard per un primo corso di Statistica Matematica rivolto a studenti dal terzo anno in poi.
Il corso si è sviluppato dal corso di Verosimiglianza, tenuto da Fabio Rapallo 2005-06, che ringrazio per aver condiviso appunti ed esercizi. Gli esercizi d'esame sono frutto di collaborazione tra l'autore e Maria Piera Rogantin, titolare della seconda parte del corso e con la quale è costante lo scambio di know-how sui temi del corso.
Ringrazio Manuele Leonelli per l'attenta lettura e gli studenti del corso che dal 2007 con le loro domande, correzioni ed attenzione mi hanno indicato l'opportunità di scrivere questi appunti e dato suggerimenti per migliorarli. Gli appunti sono in forma provvisoria, non sono esenti da lacune e imprecisioni. Ogni suggerimento di miglioramento è benvenuto.

9 Bibliografia

Testi di riferimento:
G. Casella e R.L. Berger, Statistical inference, Wadsworth 62-2002-02 62-2002-09
D. A. Freedman, Statistical Models, Theory and Practice, Cambridge, 62-2009-05
L. Pace e A. Salvan, Teoria della statistica, CEDAM 62-1996-01
M. Gasparini, Modelli probabilistici e statistici, CLUT 60-2006-08
D. Dacunha-Castelle e M. Duflo, Probabilites et Statistiques, Masson 60-1982-18/19/26 e 60-1983-22/23/24
A.C. Davison. Statistical Models, Cambridge University Press, Cambridge, 2003
L.A. Wasserman. All of Statistics: A Concise Course in Statistical Inference, Springer 2005

Letture consigliate:
David J. Hand, A very short introduction to Statistics, Oxford 62-2008-05
L. Wasserman. All of Statistics, Springer
J. Jacod and P. Protter, Probability Essentials, Springer 60-2004-09
S.L. Lauritzen, Graphical models, Oxford University press 62-1996-14
D. Williams, Probability with Martingales, Cambridge Mathematical Textbooks, 1991

10 Esercizi

Gli esercizi in questo paragrafo sono un'ntegrazione e variazione di un eserciziario per il corso di Statistica e Verosimiglianza compilato da Fabio Rapallo, with thanks.

10.1 Densità e speranze condizionate

Esercizio 1. Siano X_1 e X_2 variabili aleatorie indipendenti con distribuzione geometrica $G(p)$, $p \in (0,1)$. Sia $S = X_1 + X_2$. Calcolare le densità condizionate $f_{X_1|S}$ e $f_{S|X_1}$.

Esercizio 2. Siano $X_1, ..., X_n$ variabili aleatorie indipendenti con distribuzione di Bernoulli $Bern(p)$, $p \in (0,1)$, e siano $U = \min\{X_1, ..., X_n\}$ e $V = \max\{X_1, ..., X_n\}$.

1. Calcolare la densità congiunta del vettore (U, V) e le densità marginali;
2. calcolare le densità condizionate $f_{U|V}$ e $f_{V|U}$.

Esercizio 3. Siano X_1 e X_2 variabili aleatorie indipendenti con distribuzione binomiale, $X_1 \sim Bin(n_1, p)$, $X_2 \sim Bin(n_2, p)$, con $n_1, n_2 > 0$ e $p \in (0,1)$. Sia $S = X_1 + X_2$.

1. Calcolare la densità condizionata $f_{X_1|S}$;
2. calcolare la speranza condizionata $E(X_1|S)$.

Esercizio 4. Siano X_1 e X_2 variabili aleatorie indipendenti con distribuzione normale, $X_1, X_2 \sim N(\mu, \sigma^2)$, con $\mu \in \mathbb{R}$ e $\sigma^2 > 0$. Sia $S = X_1 + X_2$. Calcolare la densità condizionata $f_{X_1|S}$.

Esercizio 5. Siano $X_1, ..., X_n$ variabili aleatorie indipendenti con distribuzione esponenziale $E(\lambda)$, $\lambda > 0$, e siano $U = \min\{X_1, ..., X_n\}$ e $V = \max\{X_1, ..., X_n\}$.

1. Calcolare la densità condizionata $f_{X_1|U}$;
2. calcolare la densità condizionata $f_{X_1|V}$.

Esercizio 6. Siano X_1 e X_2 variabili aleatorie indipendenti con distribuzione uniforme discreta $\mathcal{U}\{1, \ldots, k\}$, con $k > 0$. Sia $S = X_1 + X_2$.

1. Preso $k = 4$, calcolare la densità condizionata $f_{X_1|S}$;
2. calcolare la densità condizionata $f_{X_1|S}$ e la speranza condizionata $E(X_1|S)$ nel caso generale.

Esercizio 7. Siano $X_1 \sim E(\lambda_1)$ e $X_2 \sim E(\lambda_2)$ due variabili aleatorie indipendenti, con $\lambda_1 > \lambda_2 > 0$. Sia $S = X_1 + X_2$.

1. Calcolare la densità congiunta del vettore (X_1, S) e le densità marginali;
2. calcolare la densità condizionata $f_{X_1|S}$;
3. calcolare la speranza condizionata $E(X_1|S)$.

Esercizio 8. Sia (X_1, X_2) un vettore aleatorio bivariato la cui distribuzione congiunta è una normale multivariata $N(\mu, \Sigma)$, con $\mu = (0,0)$ e $\Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$.

1. Calcolare la densità condizionata $f_{X_1|X_2}$;
2. calcolare la speranza condizionata $E(X_1|X_2)$.

Esercizio 9. Siano X_1 e X_2 variabili aleatorie indipendenti con distribuzione uniforme continua $U[0,b]$, $b > 0$, e sia $S = X_1 + X_2$.

1. Calcolare la densità condizionata $f_{X_1|S}$;
2. calcolare la speranza condizionata $E(X_1|S)$.

Esercizio 10. Sia (X, Y) un vettore aleatorio bivariato con densità congiunta: $f(x, y) = k(x + y)$, $x \in (0, 1)$, $y \in (0, 1)$.

1. Determinare il valore di k affinché $f(x, y)$ sia una densità di probabilità;
2. calcolare le densità marginali di X e di Y ;
3. calcolare la densità condizionata $f_{X_1|X_2}$;
4. calcolare la speranza condizionata $E(X_1|X_2)$.

10.2 Modelli statistici e classe esponenziale

Esercizio 1. Sia $X_1, \ldots, X_n$ un campione della legge geometrica $G(p)$, $p \in (0, 1)$.

1. Dire se il modello appartiene alla classe esponenziale e, in caso affermativo, individuare il parametro naturale e la funzione dei cumulanti;
2. indicare la statistica sufficiente e completa.

Esercizio 2. Sia $X_1, \ldots, X_n$ un campione della legge uniforme discreta $U\{1, ..., k\}$, $k > 0$.

1. Dire se il modello appartiene alla classe esponenziale e, in caso affermativo, individuare il parametro naturale e la funzione dei cumulanti;
2. indicare la statistica sufficiente.

Esercizio 3. Sia $X_1, \ldots, X_n$ un campione della legge binomiale $Bin(r, p)$, $r > 0$ e $p \in (0, 1)$.

1. Supponendo r noto, dire se il modello appartiene alla classe esponenziale e, in caso affermativo, individuare il parametro naturale e la funzione dei cumulanti. Indicare in ogni caso la statistica sufficiente;
2. rispondere alle stesse domande supponendo invece p noto;
3. rispondere alle stesse domande supponendo entrambi i parametri non noti.

Esercizio 4. Sia $X_1, \ldots, X_n$ un campione della legge sull'insieme $\{-1, 0, 1\}$ con densità:

$$f_\theta(\pm 1) = \frac{\theta}{2} \quad f_\theta(0) = 1 - \theta$$

1. Dire se il modello appartiene alla classe esponenziale e, in caso affermativo, individuare il parametro naturale e la funzione dei cumulanti;
2. indicare la statistica sufficiente.

Esercizio 5. Sia $X_1, \ldots, X_n$ un campione della legge gamma $\Gamma(a, \lambda)$, $a > 0$ e $\lambda > 0$.

1. Supponendo a noto, dire se il modello appartiene alla classe esponenziale e, in caso affermativo, individuare il parametro naturale e la funzione dei cumulanti. Indicare in ogni caso la statistica sufficiente;
2. rispondere alle stesse domande supponendo invece λ noto;
3. rispondere alle stesse domande supponendo entrambi i parametri non noti.

Esercizio 6. Sia $X_1, \ldots, X_n$ un campione della legge di Pareto $Par(a, \theta)$, $a > 0$ e $\theta > 0$.

1. Supponendo a noto, dire se il modello appartiene alla classe esponenziale e, in caso affermativo, individuare il parametro naturale e la funzione dei cumulanti. Indicare in ogni caso la statistica sufficiente;
2. rispondere alle stesse domande supponendo invece θ noto.

Esercizio 7. Sia $X_1, \ldots, X_n$ un campione della legge beta $B(\alpha, \beta)$, $\alpha, \beta > 0$.

1. Supponendo β noto e uguale a 1, dire se il modello appartiene alla classe esponenziale e, in caso affermativo, individuare il parametro naturale e la funzione dei cumulanti;
2. indicare, sempre nel caso $\beta = 1$, la statistica sufficiente.

Esercizio 8. Sia $X_1, \ldots, X_n$ un campione della legge uniforme continua sull'intervallo $[\theta - 1, \theta + 1]$, con $\theta \in \mathbb{R}$.

1. Determinare una statistica sufficiente e bi-dimensionale;
2. mostrare che tale statistica non è completa.

10.3 Stimatori e verosimiglianza

Esercizio 1. Sia $X_1, \ldots, X_n$ un campione della legge uniforme continua $\mathcal{U}[0, \theta]$, $\theta > 0$. Dati i seguenti stimatori di θ:

$$T_1 = \frac{n}{n+1} \max X_i; \quad T_2 = 2\bar{X}_n; \quad T_3 = X_1 + X_n; \quad T_4 = X_{(1)} + X_{(n)}$$

dire se sono non distorti ed individuare quello di minimo rischio (quadratico medio).

Esercizio 2. Sia $X_1, \ldots, X_n$ un campione della legge normale $\mathcal{N}(\mu, 1)$, $\mu \in \mathbb{R}$. Si vuole stimare μ.

1. Calcolare distorsione e rischio dello stimatore $T_1 = \bar{X}_n^2$;
2. calcolare distorsione e rischio dello stimatore $T_2 = X_1 X_2$.

Esercizio 3. Sia $X_1, \ldots, X_n$ un campione della legge di Poisson $P(\lambda)$, $\lambda > 0$.

1. Stimare λ con X_n e dire se è non distorto;
2. confrontare lo stimatore X_n con lo stimatore $(X_1 + X_n)/2$;
3. stimare e^λ con e^{X_n} , verificare che è distorto e correggerlo.

Esercizio 4. Sia $X_1, \ldots, X_n$ un campione della legge di Bernoulli Bern(p), $p \in (0, 1)$. Dire se lo stimatore di p $T_n = \frac{X_1^2 + \ldots + X_n^2}{n}$ è non distorto e di minimo rischio.

Esercizio 5. Sia $X_1, \ldots, X_n$ un campione della legge geometrica $G(p)$, $p \in (0, 1)$. Dire se lo stimatore di p

1. $T_n = \frac{1_{X_1=1} + \ldots + 1_{X_n=1}}{n}$ è non distorto e calcolarne il rischio.
2. Dire se lo stimatore $T_n = X_1 X_n$ è non distorto per $g(p) = 1/p^2$ e calcolarne il rischio.

Esercizio 6. Sia $X_1, \ldots, X_n$ un campione della legge di Poisson $P(\lambda), \lambda > 0$. Si supponga $n \geq 3$. Dire se lo stimatore $T_n = (X_1 + X_n) X_2$ è non distorto per $g(\lambda) = 2\lambda^2$ e calcolarne il rischio.

Esercizio 7. Sia $X_1, \ldots, X_n$ un campione della legge esponenziale $E(\lambda)$, $\lambda > 0$. Si supponga $n \geq 3$. Dire se la statistica $T_n \frac{X_1 + X_n}{X_2}$ è stimatore non distorto di qualche funzione di λ.

Esercizio 8. Sia $X_1, \ldots, X_n$ un campione di numerosità 50 della legge binomiale Bin(20,p), $p \in (0, 1)$. Se la media empirica delle 50 osservazioni è $x_n = 3.5$, è più verosimile $p = 0.1$ oppure $p = 0.2$?

Esercizio 9. Sia $X_1, \ldots, X_n$ un campione della legge geometrica G(p), $p \in (0, 1)$.

1. Determinare lo stimatore di massima verosimiglianza di p;
2. dire quale funzione del parametro p è stimata in modo non distorto dalla statistica sufficiente $S_n = X_1 + \ldots + X_n$;
3. calcolare il rischio di S_n come stimatore di $g(p)$.

Esercizio 10. Sia $X_1, \ldots, X_n$ un campione della legge sull'insieme $\{-1, 0, 1\}$ con densità:

$$f_\theta(\pm 1) = \frac{\theta}{2} \quad f_\theta(0) = 1 - \theta$$

1. Determinare lo stimatore di massima verosimiglianza di θ, verificare se è distorto ed eventualmente derivarne uno stimatore non distorto;
2. determinare lo stimatore di massima verosimiglianza di $(1-\theta)^2$ e derivarne lo stimatore UMVUE.

Esercizio 11. Sia $X_1, \ldots, X_n$ un campione della legge $\Gamma(a, \beta)$, $a > 0$ e $\beta > 0$.

1. Supponendo a noto, determinare lo stimatore UMVUE di β a partire dallo stimatore di massima verosimiglianza;
2. supponendo sempre a noto, determinare lo stimatore di massima verosimiglianza di $e^{-\lambda}$.

Esercizio 12. Sia $X_1, \ldots, X_n$ un campione della legge di $Pareto(a, \theta)$, $a, \theta > 0$: $\theta a^\theta x^{-(\theta+1)}$ per $x > a$.

1. Supponendo a noto, determinare lo stimatore di massima verosimiglianza di θ;
2. sempre supponendo a noto, determinare lo stimatore UMVUE di θ;
3. supponendo θ noto, calcolare lo stimatore di massima verosimiglianza di a e calcolarne il rischio.

Esercizio 13. Sia $X_1, \ldots, X_n$ un campione della legge log-normale $LN(\mu, \sigma)$, $\mu, \sigma \in \mathbb{R}$, $\sigma > 0$. Si ricordi che la densità della legge log-normale è: $f_{\mu,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} x^{-1} e^{-(\log x - \mu)^2/(2\sigma^2)}$ per $x > 0$ e 0 altrove.

1. Supponendo $\sigma = 1$, determinare lo stimatore di massima verosimiglianza di μ e calcolarne il rischio;
2. supponendo $\mu = 0$, determinare lo stimatore di massima verosimiglianza di σ nel caso di un campione di numerosità $n = 1$.

Esercizio 14. Sia $X_1, \ldots, X_n$ un campione della legge su $\{0, 1, 2\}$ con densità

x	0	1	2
$P(X = x)$	α	β	$1 - \alpha - \beta$

1. Determinare le condizioni più generali possibili affinché la densità sia ben definita;
2. dire se il modello appartiene alla classe esponenziale;
3. individuare la statistica sufficiente;
4. calcolare gli stimatori di massima verosimiglianza per (α, β).

Esercizio 15. Sia u una tabella di contingenza $r \times c$, $r, c \in \mathbf{Z}_{\geq 0}$. Sia $\{n_{ij}\}_{i,j}$ la matrice dei conteggi osservati con $n_{ij} > 0$ e sia $n = \sum_{ij} n_{ij}$. Sotto l'ipotesi di conteggi multinomiali, dimostrare che lo stimatore di massima verosimiglianza delle probabilità p_{ij} è $\hat{p}_{ij} = n_{ij}/n$. [Hint: utilizzare i moltiplicatori di Lagrange]

Esercizio 16. Sia u una tabella di contingenza $r \times c$, $r, c \in \mathbf{Z}_{\geq 0}$ e si ipotizzi il modello di indipendenza $p_{ij} = p_{i+}p_{+j}$ ($i = 1, \ldots, r$ e $j = 1, \ldots, c$). Sia $\{n_{ij}\}_{i,j}$ la matrice dei conteggi osservati e sia $n = \sum_{ij} n_{ij}$.

1. Scrivere le funzioni di verosimiglianza e di log-verosimiglianza.
2. Verificare che
$$\hat{p}_{i+} = \frac{u_{i+}}{n} \text{ per } i = 1, \ldots, r \quad \text{e} \quad \hat{p}_{+j} = \frac{u_{+j}}{n} \text{ per } j = 1, \ldots, c$$
sono stime di massima verosimiglianza.

Esercizio 17. Un giocatore lancia una moneta $k = 4$ volte e annota il numero di teste che può essere $0, 1, 2, 3, 4$. Il gioco è ripetuto per $n = 242$ volte e si osservano $(n_0, \ldots, n_4)$ dove n_i è il numero di giochi nel quale escono i teste ($i = 0, \ldots, 4$). Sia $n = \sum_{i=0}^4 n_i$. Sia p_i la probabilità di osservare i teste in un gioco. Scrivere la funzione di verosimiglianza.

Si sospetta che il giocatore imbrogli e che usi due monete, che indichiamo con A e B. Si sospetta che ad ogni gioco scelga A con probabilità $\pi \in (0, 1)$. Inoltre, la probabilità che esca testa in A è α e in B è β con $0 < \alpha, \beta < 1$. Scrivere la relazione tra i p_i e π, α, β.

Verificare che non esiste un unico stimatore di massima verosimiglianza.

10.4 Informazione secondo Fisher e disuguaglianza di Cramér-Rao

Esercizio 1. Sia $X_1, \ldots, X_n$ un campione della legge di Pascal $Pas(r, p)$ con $r \leq 1$ noto e $p \in (0, 1)$.

1. Calcolare l'informazione di Fisher per il parametro p;
2. verificare se lo stimatore di massima verosimiglianza è efficiente.

Esercizio 2. Sia $X_1, \ldots, X_n$ un campione della legge $\Gamma(\alpha, \lambda)$ con $\alpha > 0$ noto e $\lambda > 0$.

1. Calcolare l'informazione di Fisher per il parametro λ;
2. verificare se lo stimatore di massima verosimiglianza di λ è efficiente;
3. determinare il limite inferiore di Cramér-Rao per $g(\lambda) = 1/\lambda$.

Esercizio 3. Calcolare l'informazione di Fisher per il parametro (μ, σ) per un campione i.i.d. log-normale.

Esercizio 4. Sia $X_1, \ldots, X_n$ un campione della legge normale $N(0, \sigma^2)$, $\sigma^2 > 0$.

1. Determinare lo stimatore di massima verosimiglianza del parametro;
2. calcolare l'informazione di Fisher;
3. verificare se lo stimatore di massima verosimiglianza è efficiente.

Esercizio 5. Sia $X_1, \ldots, X_n$ un campione della legge su $\mathbb{R}$ avente densità: $f_\theta(x) = \frac{x+1}{\theta(\theta+1)} e^{-x/\theta}$ per $x > 0$ e zero altrove, $\theta > 0$.

1. Dire se il modello appartiene alla classe esponenziale ed individuare la statistica sufficiente T;
2. dire quale funzione di θ è stimata in modo non distorto da T;
3. dire se T è uno stimatore efficiente di questa funzione.

Esercizio 6. Considerare un campione di taglia uno della legge uniforme $\{1, \ldots, k+1\}$ con $\theta_i \in (0, 1)$ la probabilità dell'evento $i = 1, \ldots, k+1$ e porre $\theta_{k+1} = 1 - \theta_{i=1}^{k} \theta_i$.

1. Calcolare le funzioni di verosimoglianza e di log-verosimiglianza;
2. calcolare l'informazione di Fisher.
3. dire se T è uno stimatore efficiente di questa funzione.

10.5 Stima con il metodo dei momenti

Esercizio 1. Si consideri un campione i.i.d. di taglia n estratto dalla legge con densità: $f(x; \theta) = \theta^{-1} x^{\theta^{-1}-1}$ per $x \in (0, 1)$ e $\theta > 0$. Determinare lo stimatore di θ con il metodo dei momenti, verificare se è non distorto ed eventualmente correggerlo.

Esercizio 2. Si consideri un campione di taglia n della legge esponenziale traslata $\mathcal{E}(a, \lambda)$ avente densità: $f(x; a, \lambda) = \lambda e^{-\lambda(x-a)}$ per $x > a$ e 0 altrove e $a \in \mathbb{R}$ e $\lambda > 0$. Determinare gli stimatori di a e λ con il metodo dei momenti, verificare se sono non distorti e calcolarne il rischio.

Esercizio 3. Si consideri un campione di taglia n della legge di Weibull $W(\alpha, \lambda)$, avente densità: $f(x; \alpha, \lambda) = \alpha \lambda x^{\alpha-1} e^{-\lambda x^\alpha}$ se $x > 0$ e 0 altrove, con $\alpha, \lambda > 0$.

1. Fissato $\alpha = 2$, determinare lo stimatore di λ con il metodo dei momenti e verificare se è distorto.
2. scrivere la verosimiglianza ed individuare la statistica sufficiente;
3. dire se lo stimatore ottenuto con il metodo dei momenti (ed eventualmente corretto) è UMVUE.

10.6 Modelli grafici [solo alcuni anni]

Esercizio 1. Sia $\mathcal{G} = \mathcal{G}(V, E)$ un grafo non diretto comprendente i soli criteri rilevanti. Scrivere la proprietà di fattorizzazione rispetto agli insiemi completi di V, le proprietà di Markov e le loro relazioni.

Esercizio 2. In uno studio sulla relazione tra grandezza dell'auto e gravità degli incidenti, questi sono stati classificati secondo il tipo di incidente e se l'autista è stato espulso dall'auto (Fienberg, 1981). La seguente tabella raccoglie i dati. Vi sono quattro variabili binarie

W=grandezza dell'auto A=tipo di incidente E=se l'autista è stato espulso S=gravità dell'incidente

		Accident type			
		Collision		Rollover	
		Severity		Severity	
Car Weight	Driver ejected	Not severe	Severe	Not severe	Severe
Small	No	350	150	60	112
	Yes	26	23	19	80
Standard	No	1878	1022	148	404
	Yes	111	161	22	265

1. Formulare un modello grafico $\mathcal{G}$ che implichi la relazione di indipendenza condizionata $A \perp\!\!\!\perp (S, E, W)$.
2. Scrivere la lista di separatori e cliques.
3. Scrivere un'espressione generale per lo stimatore di massima verosimiglianza di $p(A = a, S = s, W = w, E = e)$ per il modello in $\mathcal{P}_E(\mathcal{G})$.
4. Valutare lo stimatore di massima verosimiglianza nella cella (Collision, Severe, Small, Ejected).

Esercizio 3. I seguenti dati (Schoener, 1968) si riferiscono a dove preferiscono sostare due specie di lucertole e sono classificati secondo le seguenti tre variabili binarie

A:	Specie	A=1	Anoli
		A=0	Distichus
B:	Diametro del trespolo	B=1	≤ 4
		B=0	> 4
C:	Altezza del trespolo	C=1	> 4.75
		C=0	≤ 4.75

La relativa tabella dei conteggi, per un totale di 409 unità, è

Specie	Diametro	Altezza	
		C=1	C=0
A=1	B=1	32	86
	B=0	11	35
A=0	B=1	61	73
	B=0	41	70

1. Scrivere il reticolo dei modelli grafici non diretti associabili a A,B,C e le loro dimensioni (=numero di parametri).
2. Calcolare la stima di massima verosimiglianza nell'ipotesi che B e C siano indipendenti condizionatamente ad A.
3. Interpretare e commentare il punto precedente.

Esercizio 4. Indicare l'esempio di tre variabili aleatorie A, B, C per cui $A \perp\!\!\!\perp$ B non implica $A \perp\!\!\!\perp B|C$.

Esercizio 5. Usare la seguente tabella per verificare la falsità della seguente relazione

$$A \perp\!\!\!\perp B|C \text{ e } A \perp\!\!\!\perp B|C^c \text{ implica } A \perp\!\!\!\perp B$$

dove C^c è l'insieme complementare di C e la probailità di C è 1/2

Probabilità condizionate						Marginale		
dato C	B	B^c	dato C^c	B	B^c		B	B^c
A	0.1	0.1	A	0.1	0.4	A	0.10	0.25
A^c	0.4	0.4	A^c	0.1	0.4	A^c	0.25	0.40

10.7 Statistica Bayesiana [solo alcuni anni]

Esercizio 1. Considerare un campione $X_1, \ldots, X_n$ di legge

$$X_1|\theta \sim \exp(\theta T(x) - \psi(\theta) - c(x))$$

e una legge a priori su θ appartenente alla famiglia

$$\theta \sim \exp(\theta\mu - \phi(\mu) - d(\theta))^{11}$$

con μ parametro e ϕ, d note funzioni.

1. Dimostrare che sono distribuzioni coniugate.
2. Verificare i dettagli per un campione $Bernoulli(\theta)$.

Esercizio 2. Considerare un campione $X_1, \ldots, X_n$ di legge esponenziale λ, con $\lambda > 0$ e su λ porre la distribuzione a priori $\Gamma(a, \mu)$ con $a = 2$ e $\mu = 3$. La densità per $\Gamma(a, \mu)$ è

$$\frac{\mu^a y^{a-1} e^{-\mu y}}{\Gamma(\alpha)} \qquad y > 0, \quad a,\, \mu > 0$$

ed il valore medio è a/μ.

1. Verificare che sono distribuzioni coniugate.
2. Lo stimatore di massima verosimiglianza per λ è $\frac{n}{\sum_{i=1}^n x_i}$. Confrontarlo con il valore atteso dello stimatore a posteriori di λ. Se necessario, ipotizzare di avere $n = 5$ osservazione con media campionaria pari a $\sum x_i/n = 2/15$.

Esercizio 3. Siano $X_1, \ldots, X_n$ variabili aleatorie indipendenti ed identicamente distribuite secondo una legge di Poisson(λ), con $\lambda > 0$, e sia $S_n = \sum_{i=1}^n X_i$.

1. Verificare che S_n è statistica sufficiente per λ e che S_n ha legge Poisson($n\lambda$).

In un'ottica bayesiana siano $S_n|\Lambda \sim$ Poisson(Λ) e $\Lambda \sim \Gamma(a, l)$ con $a, l > 0$ noti, dove

$$f_\Lambda(\lambda; a, l) = \frac{l^a\, \lambda^{a-1}\, e^{-l\lambda}}{\Gamma(a)\, l^a} \qquad (\lambda > 0)$$

2. Individuare la legge a-posteriori di Y e verificare che dipende dai dati solo tramite la statistica sufficiente.
3. Generalizzare il risultato precedente dimostrando il seguente fatto. Per il modello statistico bayesiano

$$\underline{\mathrm{X}} = (X_1, \ldots, X_n)|\theta \sim f_{\underline{\mathrm{X}}}(x_1, \ldots, x_n; \theta)$$
$$\theta \sim \pi(\theta)$$

con S statistica sufficiente per θ, la legge a posteriori di θ dipende dai dati solo tramite S.

Hint: la funzione generatrice dei momenti di X_1 è $M_{X_1}(s) = e^{-\lambda(1-e^s)}$.

[11] Notare l'abuso di notazione per il quale θ indica la variabile aleatoria ed anche la variabile dipendente della densità.

11 Esercizi d'esame

Esercizio
Sia $X_1, \ldots, X_n$ un campione indipendente da

$$f(x; a) = \exp(-(x-a))\,(a \leq x < +\infty)$$

con $a \in \mathbb{R}$.

1. E' un modello di classe esponenziale ? Giustificare la risposta.
2. E' uno modello di posizione e/o scala? Giustificare la risposta.
3. Determinare una statistica unidimensionale sufficiente per a.
4. Determinare una statistica ancillare per a.
5. Determinare un gruppo di trasformazioni rispetto al quale il modello sia invariante.
6. Determinare uno stimatore di a invariante per il gruppo al punto precedente.

Traccia di soluzione:

1. Non è modello regolare quindi non può essere di classe esponenziale.
2. E' un modello di posizione. Infatti * per $a = 0$ riconosciamo in $f(\cdot; 0)$ una densità esponenziale di parametro uno * vogliamo verificare che la densità $f(\cdot; a)$ è ottenuta traslando $f(\cdot; 0)$.

   Sia $X \sim Esp(1)$ e $X = Y + a$ per $a \in \mathbb{R}$. Allora per $x \in \mathbb{R}$ vale

   $$Prob(X \leq x) = Prob(Y \leq x - a) = (1 - e^{-(x-a)})(x - a \geq 0) = (1 - e^{-(x-a)})(x \geq a) = F(x; a)$$

   dove F è la funzione di ripartizione da $f(\cdot; a)$.

   Equivalentemente, lavorando sulle densità

   $$f_X(x) = f_Y(x-a)\frac{d\,(x-a)}{dx} = \exp(-(x-a))\,(0 \leq x - a < +\infty)$$

3. La funzione di verosimiglianza per il campione è

   $$\mathcal{L}(a; \underline{x}) = \left(\prod_{i=1}^{n} exp(-x_i)\right) exp(na)(a \leq x_{(1)})$$

   da cui $T = X_{(1)}$ è statistica sufficiente per a.

4. Abbiamo visto a lezione che il rango è una statistica ancillare per modelli di posizione.

   Oppure se T è sufficiente, per esempio $T = X_{(1)}$, la legge di $\underline{X}|T$ non dipende da a. Una funzione unidimensionale di $\underline{X}|T$, per esempio $X_1|T$ o $\sum X_i|T$, è ancillare e unidimensionale per a.

5. Abbiamo visto a lezione che modelli di locazione sono invarianti per gruppi di traslazione, per esempio $\mathcal{G} = \{g_c(x_x, \ldots, x_n) = (x_1 + c, \ldots, x_n + c) :\ \text{con } c \in \mathbb{R}\}$ e la trasformazione indotta sullo spazio dei parametri da $g_c \in \mathcal{G}$ è: $\bar{g}_c : a \mapsto a + c$.

   Oppure dato $c \in \mathbb{R}$ $\mathcal{G}_2 = \{g_k(x_x, \ldots, x_n) = (x_1 + kc, \ldots, x_n + kc) :\ \text{con } k \in \mathbb{Z}\}$ e per $g_k \in \mathcal{G}_2$ si ha $\bar{g}_k : a \mapsto a + kc$.

6. W è stimatore invariante per a rispetto al gruppo $\mathcal{G}$ se $\bar{g}(W(\underline{X})) = W(g(\underline{X}))$ per ogni $g \in \mathcal{G}$. Per $W(\underline{X}) = X_{(1)}$ si ha

$$\begin{aligned} W(g_c(\underline{X})) &= \mathit{minimo}\{X_1 + c, \ldots, X_n + c\} = X_{(1)} + c \\ &= \mathit{minimo}\{X_1, \ldots, X_n\} + c = W(\underline{X}) + c = \bar{g}_c(W(\underline{X})) \end{aligned}$$

Analogamente si verifica che $\bar{X} = \sum_{i=1}^{n} X_i/n$ è invariante per a, mentre $W_2(\underline{X}) = \sum_i X_i$ non è invariante per a perché

$$W(g_c(\underline{X})) = W(X_1 + c, \ldots, X_n + c) = \sum_i X_i + nc \neq \bar{g}_c(W(\underline{X})) = g_c(\sum_i X_i) = \sum_i X_i + c$$

Esercizio
Per $a, k > 0$ sia $X_1, \dots, X_n$ un campione indipendente e identicamente distribuito secondo la legge

$$f_X(x) = ke^{-k(x-a)} \qquad \text{con } x \geq a$$

1. E' il modello di classe esponenziale?
2. E' modello di posizione e/o scala?
3. Determinare statistiche sufficienti e minimali per (a, k), per a e per k.
4. Determinare, se esiste, lo stimatore di massima verosimiglianza di a noto k.
5. Determinare, se esiste, lo stimatore di massima verosimiglianza di k noto a.
6. Determinare, se esiste, lo stimatore di massima verosimiglianza della coppia (k, a).

Traccia di soluzione:

1. Il modello non è regolare, infatti il supporto di f_X è $[a, +\infty[$ e dipende dal parametro a. Ne segue che il modello non è di classe esponenziale.
2. $Y = ...$
3. $\boxed{f_{\underline{X}}(\underline{x}) = k^n e^{-k\sum x_i} e^{nka}(x_{(1)} \geq a)}$ da cui le statistiche sufficienti sono

$$\begin{array}{lr}
(X_{(1)}, \sum_{i=1}^n X_i) \text{ per } (a,k) & \text{e.g. con } h(\underline{x}) = 1 \text{ e } g(T(\underline{x}), k, a) = \mathcal{L}(k, a; \underline{x}) \\
X_{(1)} \text{ per } a & \text{e.g. con } h(\underline{x}) = k^n e^{-k\sum x_i} \text{ e } g(T(\underline{x}), k, a) = e^{nka}(x_{(1)} \geq a) \\
\sum X_i = S_n \text{ per } k & \text{e.g. con } h(\underline{x}) = (x_{(1)} \geq a) \text{ e } g(T(\underline{x}), k, a) = e^{nka} k^n e^{-k\sum x_i}
\end{array}$$

Per la minimialità

$$\mathcal{L}(k, a; \underline{x}) = \begin{cases} k^n e^{-k\sum x_i} e^{nka}(x_{(1)} \geq a) & \text{se } x_{(1)} \geq a \\ 0 & \text{se } x_{(1)} < a \end{cases}$$

Per $\underline{x}$ e $\underline{y}$ valori del campione, numeratore e denominatore di $\mathcal{L}(k, a; \underline{x})/\mathcal{L}(k, a; \underline{y})$ sono entrambi strettamente positivi per lo stesso valore di a se e solo se $x_{(1)} = y_{(1)}$ **. Inoltre se $x_{(1)} = y_{(1)}$, il rapporto vale $e^{-k(\sum x_i - \sum y_i)}$ che non dipende da k sse $\sum x_i = \sum y_i$. Questo implica che le statistiche sufficienti sono anche minimali.

Espandiamo su **. Il rapporto

$$\frac{\mathcal{L}(k, a; \underline{x})}{\mathcal{L}(k, a; \underline{y})} = e^{-k(\sum x_i - \sum y_i)} \frac{(x_{(1)} \geq a)}{(y_{(1)} \geq a)}$$

non dipende da a sse il rapporto $\dfrac{(x_{(1)} \geq a)}{(y_{(1)} \geq a)}$ non dipende da a.

Ora

$$(x_{(1)} \geq a) = \begin{cases} 1 & \text{se } x_{(1)} \geq a \\ 0 & \text{se } x_{(1)} < a \end{cases}$$
$$(y_{(1)} \geq a) = \begin{cases} 1 & \text{se } y_{(1)} \geq a \\ 0 & \text{se } y_{(1)} < a \end{cases}$$

$(x_{(1)} \geq a) = (y_{(1)} \geq a)$ per ogni a sse $x_{(1)} = y_{(1)}$. In questo caso il rapporto non dipende da a; potrebbe essere non ben definito $1/0$, $0/0$ ma ciò non dipende da a.

Vediamo in altro modo e consideriamo $x_{(1)} \neq y_{(1)}$. Il rapporto

$$\frac{(x_{(1)} \geq a)}{(y_{(1)} \geq a)} = \begin{cases} 1/1 & \text{se } x_{(1)} \geq a \text{ e } y_{(1)} \geq a \\ 1/0 & \text{se } x_{(1)} \geq a \text{ e } y_{(1)} < a \\ 0/1 & \text{se } x_{(1)} < a \text{ e } y_{(1)} \geq a \\ 0/0 & \text{se } x_{(1)} < a \text{ e } y_{(1)} < a \end{cases}$$

da cui si vede che il dominio del rapporto dipende da a.

4.

$$\mathcal{L}(a; k\underline{x}) \propto e^{nak}(x_{(1)} \geq a) = \begin{cases} 0 & \text{se } x_{(1)} \geq a \\ \text{strettamente crescente in } a & \text{altrimenti} \end{cases}$$

Lo stimatore di massima verosimiglianza è $\hat{a} = X_{(1)}$.

5. La funzione da massimizzare in $k > 0$ è $k^n e^{k(na-s_n)} = e^{k(na-s_n)+n\log(k)}$ dove $na - s_n \leq 0$ con probabilità uno. E' funzione strettamente positiva, continua e derivabile. In zero si estende per continuità con zero, il limite per $k \mapsto +\infty$ è zero. E' sufficiente studiare $k(na - s_n) + n\log(k)$. Esiste un unico punto critico definito da $\frac{d}{dk}k(na - s_n) + n\log(k) = 0$ inoltre $\frac{d^2}{dk^2}k(na - s_n) + n\log(k) < 0$ ovunque. Si conclude che $\hat{k} = \frac{n}{S_n - na}$ è stimatore di massima verosimiglianza di k noto a.

6. Consideriamo $k > 0$ e $a \leq x_{(1)}$

$$\mathcal{L}(k, a; \underline{x}) = k^n e^{-k\sum x_i + kna} \leq k^n e^{-k\sum x_i + knx_{(1)}} \leq \hat{\hat{k}}^n e^{-\hat{\hat{k}}\sum x_i + \hat{\hat{k}}nx_{(1)}}$$

con $\hat{\hat{k}} = \frac{n}{\sum x_i - nx_{(1)}}$. La prima diseguaglianza segue dal fatto che la funzione è strettamente crescente in a, per la seconda si veda il punto precedente. Quindi $\left(\frac{n}{\sum X_i - nX_{(1)}}, X_{(1)}\right)$ è di massima verosimiglianza per (k, a).

Esercizio

Sia $X_1, \dots, X_n$ un insieme di variabili aleatorie indipendenti e con densità

$$f_{X_i}(x_i) = \begin{cases} \dfrac{1}{2i\theta} & \text{per } x_i \in]-i(\theta-1), i(\theta+1)[\\ 0 & \text{altrimenti} \end{cases}$$

con $\theta > 0$ e $i = 1, \dots, n$.

1. Determinare una trasformazione rispetto alla quale il modello statistico sia identicamente distribuito. Continuare l'esercizio con questa forma del modello.
2. E' un modello di classe esponenziale?
3. Scrivere la funzione di verosimiglianza ed individuare una statistica sufficiente.
4. Determinare una statistica sufficiente e unidimensionale per θ.
   Hint: $1_{\{\theta>a\}}1_{\{\theta>b\}} = 1_{\{\theta>\max\{a,b\}\}}$
5. Fare il grafico della funzione di verosimiglianza.
6. Discutere l'esistenza di uno stimatore di massima verosimiglianza.
7. Discutere l'esistenza dello stimatore dei momenti di θ per il modello originale.

Traccia di soluzione:

1. Per $Y_i = X_i/i \in]1-\theta, 1+\theta[$, si ha $f_{Y_i}(y) = \dfrac{1}{2\theta}$ per $y \in]1-\theta, 1+\theta[$.
2. No, il supporto dipende dal parametro.
3. $\mathcal{L}(\theta, y_1, \dots, y_n) = \prod_{i=1}^n \dfrac{(1-\theta < y_i < 1+\theta)}{2\theta} \propto \dfrac{1}{\theta^n}(\theta > 1 - \min_i y_i)(\theta > \max_i y_i - 1)$
   $T = (1 - \min_i Y_i, \max Y_i - 1)$
4. $\max\{1 - \min_i Y_i, \max Y_i - 1\}$
5. Sia $a = \max\{1 - \min_i Y_i, \max Y_i - 1\}$. La funzione da considerare è $\dfrac{(\theta > a)}{\theta^n}$ che vale zero per $\theta \leq a$, assume il massimo in $\theta = a$ e quindi descresce a zero all'aumentare della taglia campionaria.
6. Esiste unico stimatore di massima verosimiglianza e vale $\hat{\theta} = a$.
7. Il campione non è identicamente distribuito.

Esercizio
Siano $\mu \in \mathbb{R}$ e per i numero intero positivo $X_i \sim \text{Unif}\left(]\mu - 1 + \frac{1}{i}, \mu + 1 + \frac{1}{i}[\right)$ indipendenti.

1. Dare una rappresentazione grafica di X_1, X_2, X_3.
2. Calcolare la funzione di verosimiglianza per $X_1, \ldots, X_n$ e studiarla.
3. Determinare una statistica sufficiente per μ.
4. Individuare tutte le stime di massima verosimiglianza.
5. Cosa si può dire sul comportamento asintotico della funzione di verosimiglianza.

Traccia di soluzione:

1.
2. Dal grafico si dovrebbe capire che le densità di probabilità sono non nulle contemporaneamente se $x_1, \ldots, x_n \in]\mu, \mu + 1 + 1/n[$, quindi se $x_{(1)}, x_{(n)} \in]\mu, \mu + (n+1)/n[$ e quindi $x_{(n)} - x_{(1)} < (n+1)/n$.

   La funzione di verosimiglianza vale $1/2^n$ se $x_{(n)} - \frac{n+1}{n} < \mu < x_{(1)}$ e zero altrimenti. Se il range di $X_1, \ldots, X_n$, cioè $X_{(n)} - X_{(1)}$, è maggiore di $\frac{n+1}{n}$, allora è identicamente nulla.
3. $X_{(n)}$ e $X_{(1)}$
4. Non esistono se $x_{(n)} - x_{(1)} > \frac{n+1}{n}$ altrimenti un qualunque valore in $]x_{(n)} - \frac{n+1}{n}, x_{(1)})$.
5. L'intervallo $[x_{(n)} - \frac{n+1}{n}, x_{(1)}]$ tende a avere ampiezza uno, mentre la costante $1/2^n$ converge a zero. Quindi la funzione di verosimiglianza tende ad annullarsi su $\mathbb{R}$.

Esercizio
Siano $\phi > \theta > 0$ e $X_1, \dots, X_n$ un campione i.i.d. Unif $([\theta, \phi])$.

1. Calcolare la funzione di verosimiglianza.
2. Calcolare la probabilità dell'evento $\{\min X_i \le \theta\}$.
3. Discutere l'esistenza dello stimatore di massima verosimiglianza di ϕ.
   Hint: θ è fissato. Distinguere i casi $\min x_i > \theta$ e $\min x_i \le \theta$
4. Calcolare, se esiste, uno stimatore di massima verosimiglianza della coppia (θ, ϕ).
5. Calcolare uno stimatore dei momenti di (θ, ϕ).

Traccia di soluzione:

1. $\mathcal{L}(\theta, \phi; x_1, \dots, x_n) = f(x_1, \dots, x_n) = \prod_{i=1}^n \dfrac{1}{\phi - \theta}(\theta < x_i < \phi) = \dfrac{1}{(\phi-\theta)^n}(\theta < x_{(1)})(x_{(n)} < \phi)$
2. zero
3. $\mathcal{L}(\phi; \theta, x_1, \dots, x_n) = \begin{cases} \dfrac{1}{(\phi-\theta)^n}(x_{(n)} < \phi) & \text{se } x_{(1)} > \theta \\ 0 & \text{se } x_{(1)} \le \theta \end{cases}$

   In virtù del punto precedente è sufficiente calcolare il massimo in ϕ di $\dfrac{1}{(\phi-\theta)^n}(x_{(n)} < \phi)$ che è raggiunto in $\widehat{\phi}_{MV} = X_{(n)}$.
4. Si chiede di determinare i massimi della funzione bidimensionale

   $$\mathcal{L}(\theta, \phi; x_1, \dots, x_n) = \frac{1}{(\phi-\theta)^n}(\theta < x_{(1)})(x_{(n)} < \phi)(\phi > \theta > 0)$$

   E' utile indicare sul piano (θ, ϕ) le regioni in cui la funzione è nulla. Risulta che la funzione è non nulla nella regione rettangolare con $\theta \in]0, x_{(1)}[$ e $\phi > x_{(n)}$ ove assume valore $\dfrac{1}{(\phi-\theta)^n}$. Questa funzione è massima laddove $\phi - \theta$ è minima, specificamente per θ che assume il valore massimo che può assumere e ϕ il minimo (considerare le curve di livello $\phi - \theta = a$ al variare di a). Lo stimatore di massima verosimiglianza di (ϕ, θ) esiste unico ed è $\widehat{(\phi, \theta)}_{MV} = (X_{(n)}, X_{(1)})$.
5. I momenti teorici primo e secondo in questo caso sono $\mu_1 = \mathrm{E}(X_1) = \dfrac{\theta + \phi}{2}$ e $\mu_2 = \mathrm{E}\left(X_1^2\right) = \dfrac{\theta^2 + \phi\theta + \phi^2}{3}$. Mettendo a sistema si ottiene

   $$\theta = \mu_1 - \sqrt{3(\mu_2 - \mu_1^2)} \qquad \phi = \mu_1 + \sqrt{3(\mu_2 - \mu_1^2)} \tag{3}$$

   Uno stimatore dei momenti di (θ, ϕ) è ottenuto sostituendo nelle equazioni (3) i momenti empirici a quelli teorici. Quindi uno stimatore dei momenti di (θ, ϕ) è

   $$\left(\widehat{\mu}_1 - \sqrt{3(\widehat{\mu}_2 - \widehat{\mu}_1^2)}, \widehat{\mu}_1 + \sqrt{3(\widehat{\mu}_2 - \widehat{\mu}_1^2)}\right)$$

   dove $\widehat{\mu}_1 = \dfrac{\sum_{i=1}^n x_i}{n}$ è la media empirica e $\widehat{\mu}_2 = \dfrac{\sum_{i=1}^n x_i^2}{n}$.
   Si noti che l'argomento delle radici quadrate coinvolte è non negativo per la diseguaglianza di Jensen.

Esercizio
Sia $X_1, \dots, X_n$ un campione i.i.d. dalla densità

$$f_{X_1}(x) = \frac{a}{\theta^a} x^{a-1} \, (0 < x < \theta)$$

1. Individuare lo spazio dei parametri e lo spazio campionario.
2. È modello regolare?
3. È modello di scala? Se sì, rispetto a quale parametro?
4. Noto θ, verificare che la trasformazione $Z_i = \log(X_i/\theta)$ induce un modello di scala. Individuare lo spazio dei parametri e riconoscere la legge di Z_i. Riconoscere la legge di Z_i.
5. Determinare una statistica ancillare per θ.
6. Considerare il modello statistico $Y_i = \langle x_i, \beta \rangle + \varepsilon$ con $\mathbb{E}(\varepsilon) = 0$ e con $\beta, x_i \in \mathbb{R}^n$.
   (a) È di classe esponenziale? Discutere.
   (b) È parametrico? Discutere.
   (c) È di scala? Discutere.

Traccia di soluzione:

1. Occorre $\theta > 0$ e $a > 0$ affinché la densità sia positiva, inoltre l'integrale $\int_0^\theta \frac{a}{\theta^a} x^{a-1}\, dx = \frac{a}{\theta^a} \frac{x^a}{a} |_0^\theta$ è definito ed uguale a uno per ogni $a, \theta > 0$. Quindi $\Theta = \mathbb{R}^2_{>0}$ e $\mathcal{X} = (0, \theta) \subset R_{>0}$. La funzione di ripartizione è $F_X(x) = x^a/\theta^a$ se $x \in]0, \theta[$ e 1 per $x \geq \theta$.
2. No, infatti ha supporto $[0, \theta]$ che dipende dal parametro.
3. E' modello di scala in θ infatti posto $Y = X/\theta$ si ha (a) $Y \in]0, 1[$ e (b) per $y \in]0, 1[$, $\mathrm{P}(Y \leq y) = \mathrm{P}(X \leq y\theta) = \frac{(\theta y)^a}{\theta^a} (0 < \theta y < \theta) = y^a (0 < y < 1)$. Nota generale: per vedere se un modello è di posizione e/o scala porre $Y = cX + b$, calcolare $\mathrm{P}(Y \leq y)$ e cercare se esistono c, b per cui il modello è di scala o posizione. Rinconscere in questa procedura la 'standardizzazione' di variabili aletorie Gaussiane.
4. Vale $Z = \log(X/\theta) \in] - \infty, 0[$. Posti $Y = aZ = a \log(X/\theta)$ e $y \in] - \infty, 0[$, si ha

   $$\mathrm{P}(Y \leq y) = \mathrm{P}(\log(X/\theta) \leq y/a) = \mathrm{P}(X \leq \theta \exp(y/a)) = \frac{\left(\theta e^{y/a}\right)^a}{\theta^a} = e^y$$

   Lo spazio dei parametri rimane $a > 0$. La variabile aleatoria $-Y$ segue una legge esponenziale di parametro uno, quindi $-Z_i$ è esponenziale di parametro a.
5. Poichè il modello è di scala in θ ogni funzione di $X_1/X_n, \dots, X_{n-1}/X_n$ è ancillare per θ. Per esempio si potrebbe scegliere $\frac{\bar{X}}{X_n}$.
6. Le risposte a queste domande dipendono dalla legge di probabilità o dal modello statistico posto su ε. Per esempio, se la legge di probabilità di ε è nota allora le uniche quantità non note del modello sono i β. La quantità $\langle \beta, x_i \rangle$ ha l'effetto di una traslazione del valor medio di ε, quindi β è un parametro (vettoriale) di posizione. Brevemente si può rispondere: (a) sì se lo è ε, (b) in genere è semi-parametrico, dove la componente non-parametrica è data dalla distribuzione non nota di ε (c) è di posizione rispetto ai parametri β se si può supporre che la legge di $\varepsilon = Y_i - \langle \beta, x_i \rangle$ non dipenda da β. E' di scala se ε lo è.

Esercizio
Si supponga che la variabile aleatoria X abbia densità

$$f_X(x) = \frac{\alpha}{\theta^\alpha} x^{\alpha-1} \, (0 < x < \theta)$$

con $\alpha, \theta > 0$.

1. Individuare una trasformazione Y di X e dei parametri per cui

$$f_Y(y) \propto y^\gamma (0 < y < 1)$$

   dove $\propto$ indica 'è proporzionale a'. Individuare la costante di proporzionalità e lo spazio in cui varia γ.

2. Determinare uno stimatore di massima verosimiglianza di γ date n copie i.i.d. di Y.

3. Discuterne l'unicità.

4. Indicate quali problemi si pongono relativamente alla bontà del modello se le osservazioni $y_1, \ldots, y_n$ sono tali per cui $\sum_{i=1}^n \log y_i$ approssima $n/2$. Indicate anche come li affrontereste.

5. Determinare uno stimatore dei momenti di γ.

6. Indicare quale problema si pone se $\bar{y} \sim 0$ e come affrontarlo.

Traccia di soluzione:

1. Si veda l'Esercizio del 15 Febbraio 2011. L'indicazione che lo spazio campionario di Y sia $0 < y < 1$ suggerisce $Y = X/\theta$. Inoltre

$$1 = \int_0^\theta \frac{\alpha}{\theta} \left(\frac{x}{\theta}\right)^{\alpha-1} dx = \int_0^1 \frac{\alpha}{\theta} y^{\alpha-1} \, \theta dy = y^\alpha|_0^1$$

   da cui $f_Y(y) = (\gamma + 1)y^\gamma (0 < y < 1)$ e $\gamma > -1$.

2. Vale $l(\gamma; y_1, \ldots, y_n) = n \log(\gamma + 1) + \gamma T(\underline{y}) + \mathit{costante}$ con $T(\underline{y}) = \sum_{i=1}^n \log(y_i)$ e $\underline{y} = (y_1, \ldots, y_n)$. Eguagliando a zero l'equazione di verosimiglianza (derivata della funzione di log-verosimiglianza) e notando che la derivata seconda di tale funzione è negativa, si deduce che lo stimatore di massima verosimiglianza nell'interno dello spazio parametrico è $\hat{\gamma}_{MV} = -1 - n/T(\underline{Y}) = -1 - \frac{1}{\bar{T}}$.

3. Il massimo locale $-1 - \frac{1}{\bar{T}}$ è anche massimo globale di l. Infatti l è sempre crescente prima di tale valore e sempre decrescente dopo. Inoltre, per γ che tende a $+\infty$ e per γ che tende a -1, l tende a $-infty$.

4. Per quel valore lo stimatore $\hat{\gamma}_{MV}$ di γ è prossimo a -1. Però per campione proveniente da quel modello dovrebbe essere $y_i \in]0, 1[$, quindi $\log y_i < 0$ e così la loro somma, mentre $n/2 > 0$. Ricontrollare modello e dati.

5. Il valore atteso di Y vale $\frac{\gamma+1}{\gamma+2}$, eguagliandolo al primo momento empirico $\bar{Y}$ si ha $\hat{\gamma}_M = \frac{2\bar{Y}-1}{1-\bar{Y}}$.

6. Per $0 < y_i < 1$ vale $\bar{y} \sim 0$ se tutti i valori campionati sono vicini allo zero. Inoltre la stima dei momenti risulta prossima a -1, che è un estremo dello spazio parametrico. Questi valori campionari sono plausibile per il modello dato. Si potrebbe verificare la bontà del modello per esempio raccogliendo un altro insieme di dati.

Esercizio
Sia Y una variabile aleatoria definita sugli interi maggiori di 0 tale che:

$$\mathbb{P}(Y = y) = \frac{1}{2} y(y+1)p^3(1-p)^{y-1} \qquad p \in (0,1),$$

1. Verificare che appartiene alla famiglia dei modelli esponenziali, indicare il parametro canonico con il suo dominio e la statistica sufficiente canonica.
2. Calcolare il valore atteso e la varianza della variabile aleatoria Y, scritti in funzione di p.
3. Si considerino n variabili aleatorie indipendenti con la stessa legge di Y.
   (a) Scrivere la log-verosimiglianza del modello per il campione.
   (b) Calcolare lo stimatore di massima verosimiglianza V del parametro p. Lo stimatore V è distorto? è asintoticamente non distorto?

Traccia di soluzione:

1. Riconoscere una trasformazione di legge geometrica. Per $n = 1$ $l(p; y) = costante + 3 \log p + (y-1)\log(1-p) = y\theta - \psi(\theta)$ con $\theta = \log(1-p) \in]-\infty, 0[$, $p = 1 - e^\theta$, $T(Y_1) = Y_1$ e $\psi(\theta) = \log(1-p) - 3\log(p) = \theta - 3\log(1-e^\theta)$.
2. $\mathrm{E}(Y) = \dfrac{d}{d\theta}\psi(\theta) = \dfrac{1+2e^\theta}{1-e^\theta} = \dfrac{3-2p}{p}$ e $\mathrm{Var}(Y) = \dfrac{d^2}{d\theta^2}\psi(\theta) = \dfrac{3e^\theta}{(1-e^\theta)^2} = \dfrac{3(1-p)}{p^2}$
3. (a) $l(p; y_1, \ldots, y_n) = \sum_i y_i\theta - n\psi(\theta)$ in particolare la statistica canonica è $n\bar{Y}$, il parametro canonico rimane θ e la funzione dei comulanti è $n\psi(\theta)$
   (b) Lo stimatore di massima verosimiglianza di $\dfrac{d}{d\theta}n\psi(\theta)$ è $\sum_i Y_i$ e quindi $\bar{Y}$ è stimatore di massima verosimiglianza di $\dfrac{d}{d\theta}\psi(\theta) = \dfrac{3-2p}{p}$. Per il teorema di invarianza degli stimatori di massima verosimiglianza si ha $\hat{p}_{MV} = \dfrac{3}{\bar{Y}+2}$. Dalla disuguaglianza di Jensen segue $\mathrm{E}(\hat{p}_{MV}) \neq \dfrac{3}{\mathrm{E}(\bar{Y})+2} = p$. Quindi $\hat{p}_{MV}$ è stimatore distorto di p. Gli stimatori di massima verosimiglianza sono asintoticamente corretti.

Esercizio RIFARE I CONTI
Sia Y una variabile aleatoria discreta con densità di probabilità

$$f_Y(y;\alpha) = y(1-\log\alpha)^2(\log\alpha)^{y-1} \qquad \alpha \in (1,\mathrm{e}) \quad y \in \mathbb{Z}^+$$

1. Verificare che appartiene alla famiglia dei modelli esponenziali, indicare il parametro canonico con il suo dominio e la statistica sufficiente canonica.
2. Calcolare valore atteso e varianza della variabile aleatoria Y, scritti in funzione di α.
3. Si considerino n variabili aleatorie indipendenti con la stessa legge di Y.
   (a) Scrivere la log-verosimiglianza del modello per il campione.
   (b) Calcolare lo stimatore di massima verosimiglianza V del parametro $\frac{\log\alpha}{1-\log\alpha}$ e scriverlo in funzione della media campionaria $\overline{Y}$. Lo stimatore V è distorto?
   (c) Calcolare la varianza dello stimatore V e dire se raggiunge il limite inferiore di Cramér-Rao.

Traccia di soluzione:

1. La log-verosimiglianza si può scrivere come $l(\alpha;y) = 2\log(1-\log\alpha) + (y-1)\log(\log\alpha)$ quindi parametro canonico è $\theta = \log(\log\alpha)$, dove $\log\alpha \in (0,1)$ e quindi $\theta \in (-\infty, 0)$. Se si considera come statistica sufficiente $Y-1$ allora $\psi(\theta) = -2\log(1-e^\theta)$. Se si considera come statistica sufficiente Y allora $\psi(\theta) = \log\dfrac{e^\theta}{(1-e^\theta)^2} = \theta - 2\log(1-e^\theta)$.
2. Il valore atteso e la varianza della statistica sufficiente, e quindi di Y, sono:

$$\mathbb{E}(Y) = \psi'(\theta) = \frac{1+e^\theta}{1-e^\theta} = \frac{1+\log\alpha}{1-\log\alpha} \text{ con } T = Y) \qquad \mathbb{V}(Y) = \psi''(\theta) = \frac{2e^\theta}{(1-e^\theta)^2} = \frac{2\log\alpha}{(1-\log\alpha)^2} \text{ con } T = Y-1$$

3. Si procede con $T = Y - 1$ (studiate il legame tra le varie nozioni e quantità considerate nell'esercizio nei due casi)
   (a) La log-verosimiglianza del modello per un n-campione è: $l(\theta; y_1, \ldots, y_n) = 2\ n\ \log(1-e^\theta) + \theta\sum y_i$
   (b) Il parametro $\frac{\log\alpha}{1-\log\alpha}$ si può scrivere come $\frac{\mathbb{E}(Y)}{2}$, quindi stimatore di massima verosimiglianza è $V = \frac{\overline{Y}}{2}$. È non distorto ($\mathbb{E}(V) = \frac{\mathbb{E}(\overline{Y})}{2}$) e ha varianza $\mathbb{V}(V) = \frac{\mathbb{V}(\overline{Y})}{4} = \frac{2\log\alpha}{4n(1-\log\alpha)^2}$. La varianza raggiunge il limite di CR in quanto il parametro da stimare è combinazione lineare della media della statistica sufficiente.

Riconoscere di quale legge è trasformazione la densità di probablità utilizzata nell'esercizio.

Esercizio
Sia Y una variabile aleatoria definita sugli interi positivi dispari $\{1, 3, 5 \dots\}$ tale che:

$$\mathbb{P}(Y = y) = \frac{p}{\sqrt{1-p}}(1-p)^{y/2} \qquad p \in (0,1),$$

1. Verificare che appartiene alla famiglia dei modelli esponenziali, indicare il parametro canonico con il suo dominio e la statistica sufficiente canonica.
2. Calcolare il valore atteso della variabile aleatoria Y, scritto in funzione di p.
3. Si considerino n variabili aleatorie indipendenti con la stessa legge di Y.
   (a) Scrivere la log-verosimiglianza del modello per il campione.
   (b) Calcolare lo stimatore di massima verosimiglianza V del parametro p. Lo stimatore V è distorto? è asintoticamente non distorto?

Traccia di soluzione:

1. La log-verosimiglianza si può scrivere come $l(p; y) = \log p - \frac{1}{2}\log(1-p) + \frac{1}{2}y\log(1-p) + \text{costante}$ quindi scegliendo come parametro $\theta = \frac{1}{2}\log(1-p)$ con $\theta \in (-\infty, 0)$ si ha $p = 1 - e^{2\theta}$ e la verosimiglianza in forma canonica per i modelli esponenziali è $l(\theta; y) = \log\left(1 - e^{2\theta}\right) - \theta + \theta y + \text{costante}$. La statistica sufficiente è Y. Inoltre $\psi(\theta) = \theta - \log\left(1 - e^{2\theta}\right)$.
2. Il valore atteso della statistica sufficiente, e quindi di Y, è:

$$\mathbb{E}(Y) = \frac{d\ \psi(\theta)}{d\ \theta} = 1 - \frac{-2e^{2\theta}}{1 - e^{2\theta}} = \frac{1 + e^{2\theta}}{1 - e^{2\theta}} = \frac{2-p}{p}$$

   Attenzione 1: derivare rispetto al parametro canonico. Attenzione 2: se la statistica canonica scelta non è Y ma $T(Y)$ (anche il parametro canonico è verosimilmente diverso), allora $\mathbb{E}(T(Y)) = \frac{d\ \psi(\theta)}{d\ \theta}$ e di qui occorre ricavare il valore atteso di Y. Per esempio se T è trasformazione lineare di Y, $T(Y) = \frac{Y}{5} - 6$ si ha $\mathbb{E}(T(Y)) = \frac{\mathbb{E}(Y)}{5} - 6 = \frac{d\ \psi(\theta)}{d\ \theta}$ e $\mathbb{E}(Y) = 5\frac{d\ \psi(\theta)}{d\ \theta} + 30$.
3. (a) La log-verosimiglianza del modello per un n-campione è:

$$l(\theta; y_1, \dots, y_n) = n\left(\log\left(1 - e^{2\theta}\right) - \theta\right) + \theta \sum y_i + \text{cost.}$$

   (b) Lo stimatore di massima verosimiglianza per $\mathbb{E}(Y)$ è $\overline{Y}$. Essendo $\mathbb{E}(Y) = \frac{2-p}{p}$, si ha $p = \frac{2}{E(Y)+1}$ e quindi: $V = \frac{2}{\overline{Y}+1}$. Un risultato analogo si poteva ottenere ponendo uguale a 0 la derivata prima della log-verosimiglianza in p e verificando che la soluzione corrisponde a un punto di massimo.
   Lo stimatore V è distorto perché $\overline{Y}$ è non distorto e, essendo $\frac{2}{\overline{Y}+1}$ una funzione convessa in $\overline{Y}$, si ha, per la disuguaglianza di Jensen, $\mathbb{E}(V) > \frac{2}{\mathbb{E}(\overline{Y})+1}$. È asintoticamente non distorto perché stimatore di massima verosimiglianza.

Riconoscere nella legge di Y la trasformazione di qualche nota densità di probabilità.

Esercizio
Sia Y una variabile aleatoria definita su $\left\{0, \frac{1}{k}, \frac{2}{k}, \ldots, \frac{k-1}{k}, 1\right\}$ tale che:

$$\mathbb{P}(Y = y) = \binom{k}{ky} \lambda^{ky} (1 + \lambda)^{-k} \qquad \lambda \in \mathbb{R}^+, \quad k \text{ noto }, k \in \mathbb{Z}^+$$

1. Verificare che appartiene alla famiglia dei modelli esponenziali, indicare il parametro canonico con il suo dominio e la statistica sufficiente canonica.

2. Calcolare il valore atteso della variabile aleatoria Y, scritto in funzione di λ.

3. Si considerino n variabili aleatorie indipendenti con la stessa legge di Y.

   (a) Scrivere la log-verosimiglianza del modello per il campione.

   (b) Indicare lo stimatore di massima verosimiglianza di $\mathbb{E}(Y)$ e calcolare lo stimatore di massima verosimiglianza V del parametro λ. Lo stimatore V è distorto? è asintoticamente non distorto?

   (c) Calcolare la varianza asintotica dello stimatore V e dire qual è la legge asintotica di V. [Suggerimento per il calcolo dell'informazione di Fisher utilizzare la log-verosimiglianza scritta in funzione di λ].

Traccia di soluzione:

1. La log-verosimiglianza si può scrivere come $l(\lambda; y) \propto ky \log \lambda - k \log(1 + \lambda)$ quindi parametro canonico è $\theta = k \ \log \lambda$ con $\theta \in \mathbb{R}$. La statistica sufficiente è Y. Inoltre $\psi(\theta) = k \log(1 + e^{\theta/k})$.

2. Il valore atteso della statistica sufficiente, e quindi di Y, è:

$$\mathbb{E}(Y) = \psi'(\theta) = k \frac{e^{\theta/k}/k}{1 + e^{\theta/k}} = \frac{e^{\theta/k}}{1 + e^{\theta/k}} = \frac{\lambda}{1 + \lambda}$$

3. (a) La log-verosimiglianza del modello per un n-campione è: $l(\theta; y_1, \ldots, y_n) \propto nk \ \log(1 + e^{\theta/k}) + \theta \sum y_i$

   (b) Lo stimatore di massima verosimiglianza per $\mathbb{E}(Y)$ è $\overline{Y}$. Essendo $\lambda = \frac{\mathbb{E}(Y)}{1-\mathbb{E}(Y)}$, si ha: $V = \frac{\overline{Y}}{1-\overline{Y}}$

   Lo stimatore è distorto perché $\overline{Y}$ è non distorto. È asintoticamente non distorto perché stimatore di massima verosimiglianza.

   (c) La varianza asintotica di V è l'inverso dell'informazione di Fisher calcolata in λ.
   Si ha: $\mathcal{I}_\lambda = - \ \mathbb{E}\left(\frac{d^2 l(\lambda; Y)}{d\lambda^2}\right)$

$$\frac{dl(\lambda)}{d\lambda} = k \sum Y_i \frac{1}{\lambda} - nk \frac{1}{1 + \lambda} \qquad \frac{dl(\lambda)^2}{d^2\lambda} = -k \sum Y_i \frac{1}{\lambda^2} + nk \frac{1}{(1 + \lambda)^2}$$

   Da cui:

$$\mathcal{I}_\lambda = - \ \mathbb{E}\left(\frac{d^2 l(\lambda; Y)}{d\lambda^2}\right) = \frac{nk}{\lambda(1 + \lambda)^2} \qquad \mathbb{V}(V) = \frac{\lambda(1 + \lambda)^2}{nk}$$

Esercizio
Sia Y una variabile aleatoria definita su $\{i, i+0.5, \dots \mid i \in \mathbb{Z}^+\}$ ovvero $\{0, 0.5, 1, 1.5, \dots\}$ tale che:

$$\mathbb{P}(Y = y) = \frac{1}{(2y)!} e^{2\gamma y} \exp(-e^{\gamma}) \qquad \gamma \in \mathbb{R},$$

1. Verificare che appartiene alla famiglia dei modelli esponenziali, indicare il parametro canonico con il suo dominio e la statistica sufficiente canonica.
2. Calcolare il valore atteso della variabile aleatoria Y, scritto in funzione di γ.
3. Si considerino n variabili aleatorie indipendenti con la stessa legge di Y.
   (a) Scrivere la log-verosimiglianza del modello per il campione.
   (b) Calcolare lo stimatore di massima verosimiglianza V del parametro γ. Lo stimatore V è distorto? è asintoticamente non distorto?
   (c) Calcolare la varianza asintotica dello stimatore V e dire qual è la legge asintotica di V. [Suggerimento per il calcolo dell'informazione di Fisher utilizzare la log-verosimiglianza scritta in funzione di γ].

Traccia di soluzione:

1. La log-verosimiglianza si può scrivere come $l(\gamma; y) \propto 2\gamma y - e^{\gamma}$ quindi scegliendo come parametro $\theta = 2\gamma$ con $\theta \in \mathbb{R}$ la verosimiglianza in forma canonica per i modelli è $l(\theta; y) \propto \theta y - e^{\theta/2}$ La statistica sufficiente è Y. Inoltre $\psi(\theta) = e^{\theta/2}$.
2. Il valore atteso della statistica sufficiente, e quindi di Y, è:

$$\mathbb{E}(Y) = \psi'(\theta) = \frac{1}{2} e^{\theta/2} = \frac{1}{2} e^{\gamma}$$

3. (a) La log-verosimiglianza del modello per un n-campione è: $l(\theta; y_1, \dots, y_n) \propto -ne^{\theta/2} + \theta \sum y_i$
   (b) Lo stimatore di massima verosimiglianza per $\mathbb{E}(Y)$ è $\overline{Y}$. Essendo $\gamma = \log(2E(Y))$, si ha: $V = \log(2\overline{Y})$. Un risultato analogo si poteva ponendo uguale a 0 la derivata prima della log-verosimiglianza in γ e verificando che la soluzione corrisponde a un punto di massimo.
   Lo stimatore V è distorto perché $\overline{Y}$ è non distorto e $\gamma = \log 2 + \log \mathbb{E}(\overline{Y}) \geq \log 2 + \mathbb{E}(\log \overline{Y}) = \mathbb{E}(2 \log \overline{Y}) = \mathbb{E}(V)$. È asintoticamente non distorto perché stimatore di massima verosimiglianza.
   (c) La varianza asintotica di V è l'inverso dell'informazione di Fisher calcolata in γ.
   Si ha: $\mathcal{I}_{\gamma} = -\,\mathbb{E}\left(\frac{d^2 l(\gamma; Y)}{d\gamma^2}\right)$

$$\frac{dl(\gamma)}{d\gamma} = 2\sum Y_i - ne^{\gamma}; \qquad \frac{dl(\gamma)^2}{d^2\gamma} = -ne^{\gamma}; \qquad \mathcal{I}_{\gamma} = -\,\mathbb{E}\left(\frac{d^2 l(\gamma; Y)}{d\gamma^2}\right) = ne^{\gamma}; \qquad \mathbb{V}(V) = \frac{1}{n} e^{-\gamma}$$

   La legge asintotica è normale.

Esercizio
Si ricordi il teorema per cui se $W = g(V)$ con V e W variabili aleatorie reali e g funzione monotona allora vale

$$f_W(w) = \begin{cases} f_V(g^{-1}(w)) \left| \dfrac{d}{dw} g^{-1}(w) \right| & \text{se } w = g(v) \text{ per } v \text{ tale che } f_V(v) > 0 \\ 0 & \text{altrimenti} \end{cases}$$

1. Verificare che il modello statistico $\mathcal{F} = \{f_Y(\cdot; a) : a > 0\}$, per la variabile aleatoria Y a valori reali, è una famiglia di scala, dove
$$f_Y(y; a) = ae^{ay}e^{-e^{ay}} \qquad \text{con } y \in \mathbb{R}$$

2. Siano Y_1 e Y_2 copie indipendenti di Y. Verificare che il rapporto $\dfrac{Y_1}{Y_2}$ è statistica ed è ancillare per a.

3. Verificare che il seguente modello statistico, espresso in termini di densità di probabilità, per la variabile aleatoria X a valori reali non negativi, non è di classe esponenziale
$$\mathcal{G} = \left\{ f(x; \gamma, \beta) = \frac{\gamma}{\beta} x^{\gamma - 1} \exp\left(-\frac{x^\gamma}{\beta} \right) : x \geq 0, \ \gamma, \beta > 0 \right\}$$

4. Individuare i valori dei parametri γ e β per cui $Y = \log X$.

5. Dai precedenti punti 3. e 4. si può dedurre che Y non è di classe esponziale?

Traccia di soluzione:

1. Sia Z la v.a. con densità $f_Z(z; 1) = e^z e^{-e^z}$. Verifichiamo che $Z = aY$. Per la funzione di ripartizione si ha $\mathrm{P}(Y \leq y) = \mathrm{P}(Z \leq ay)$ e per la densità $f_Y(y) = f_Z(ay) \dfrac{d(ay)}{dy} = e^{ay}e^{-e^{ay}} \ a$, come da teorema.

2. Secondo un esercizio lasciato da finire a lezione in un campione casuale X_i, $i = 1, \ldots, n$, e per un modello di scala ogni statistica funzione dei rapporti X_i/X_1, $i = 1, \ldots, n$, è ancillare per il parametro di scala. Specificamente in questo caso si ha
$$\frac{Y_1}{Y_2} = \frac{aZ_1}{aZ_2} = \frac{Z_1}{Z_2}$$
questo rapporto non dipende da parametri ed è funzione del solo campione, quindi è statistica. Inoltre la sua legge è funzione delle sole leggi di Z_1 e Z_2 che non dipendono dal parametro. Quindi la statistica $\dfrac{Y_1}{Y_2}$ è ancillare perché la sua legge non dipende dal parametro.

3. La funzione di log-verosimiglianza è
$$l(x; \gamma, \beta) \propto \log \gamma/\beta + \gamma \log(x) - x^\gamma/\beta$$
l'ultimo termine non è esprimibile come prodotto scalare di una statistica e del parametro.

4. Per il teorema citato con $y = \log X$ e $g(y) = \log(x)$ e $g^{-1}(x) = e^y$, si ha
$$f_Y(y) = f_X(e^y)e^y = \gamma/\beta(e^y)^{\gamma-1} \exp\left(-(e^y)^\gamma/\beta\right) e^y = \gamma/\beta e^{\gamma y} \exp\left(-(e^{\gamma y}/\beta\right)$$
da cui $\gamma = a$ e $\beta = 1$.

5. Non si può dire. Nel caso specifico la funzione di verosimiglianza non si può scrivere come un esponenziale per lo stesso motivo.

Esercizio
Siano $X_1, \ldots, X_n$ i.i.d. $\mathcal{N}\left(\theta, a\theta^2\right)$ con a costante nota positiva e $\theta > 0$.

1. Determinare una statistica sufficiente e minimale per θ bidimensionale.
2. È un modello di classe esponenziale?
3. Verificare che $T = (\bar{X}, S^2)$ è statistica sufficiente e minimale per θ.
4. Dimostrare che T non è completa per il modello dato calcolando $\mathrm{E}\left(\dfrac{n}{a+n}\bar{X}^2 - \dfrac{S^2}{a}\right)$.
5. Calcolare uno stimatore dei momenti di θ.

Traccia di soluzione:

1. Si procede come con una normale μ, σ e per $n = 1$ la log-verosimiglianza è
$$\langle \left(\frac{-1}{2a\theta^2}, \frac{1}{a\theta}\right), (x^2, x) \rangle - \log\theta$$
Per generico n si ha $V = (\sum_{i=1}^n X_i^2, \sum_{i=1}^n X_i)$ statistica sufficiente. È minimale infatti:
$$\begin{aligned} \log\mathcal{L}(\underline{x}) - \log\mathcal{L}(\underline{y}) &= \langle \left(\frac{-1}{2a\theta^2}, \frac{1}{a\theta}\right), \left(\sum_{i=1}^n x_i^2, \sum_{i=1}^n x_i\right) \rangle \\ &\quad - n\log\theta - \langle \left(\frac{-1}{2a\theta^2}, \frac{1}{a\theta}\right), \left(\sum_{i=1}^n y_i^2, \sum_{i=1}^n y_i\right) \rangle + n\log\theta \\ &= \langle \left(\frac{-1}{2a\theta}, \frac{1}{a\theta}\right), \left(\sum_{i=1}^n (x_i^2 - y_i^2), \sum_{i=1}^n (x_i - y_i)\right) \rangle \end{aligned}$$
Questo è un polinomio di secondo grado in $1/\theta$, quindi è uguale a zero se e solo se $\sum_{i=1}^n x_i^2 = \sum_{i=1}^n y_i^2$ and $\sum_{i=1}^n x_i = \sum_{i=1}^n y_i$. È statistica sufficiente e minimale.
2. Non è di classe esponenziale, per esempio, perché la statistica sufficiente minimale è bidimensionale mentre lo spazio dei parametri è unidimensionale.
3. T è funzione invertibile di V.
4. $\mathrm{E}\left(\dfrac{n}{a+n}\bar{X}^2 - \dfrac{S^2}{a}\right) = \dfrac{n}{a+n}\mathrm{E}\left(\bar{X}^2\right) - \dfrac{1}{a}\mathrm{E}\left(S^2\right) = \dfrac{n}{a+n}\left(\dfrac{a\theta^2}{n} + \theta^2\right) - \dfrac{1}{a}\left(a\theta^2\right) = 0$
5. Per esempio $\hat{\theta}_{mom} = \bar{X}$.

Esercizio
Sia Y una variabile aleatoria discreta che assume valori nei numeri naturali maggiori di 0 con

$$\mathbb{P}(Y=y)=e^{-\lambda/2}\ \lambda^{y-1}\ 2^{1-y}\ \frac{1}{(y-1)!}\qquad \lambda\in(0,+\infty)$$

1. Scrivere la verosimiglianza del modello; verificare che appartiene alla famiglia dei modelli esponenziali, indicare parametro canonico e statistica sufficiente.
2. Calcolare valore atteso e varianza della variabile aleatoria Y, scritti in funzione di λ.
3. Si considerino n variabili aleatorie indipendenti con la stessa legge di Y.
   (a) Scrivere la verosimiglianza del modello per il campione.
   (b) Calcolare lo stimatore di massima verosimiglianza del parametro λ. È distorto?
   (c) Calcolare il limite inferiore di Cramér-Rao per la varianza dello stimatore di massima verosimiglianza di λ.

Traccia di soluzione:

1. $e^{-\lambda/2}\exp((y-1)\log\lambda)=\exp(-\frac{\lambda}{2}+y\log\lambda-\log\lambda)=\exp(y\log\lambda-(\frac{\lambda}{2}+\log\lambda))$

   da cui $T(Y)=Y$, $\boxed{\theta=\log\lambda\in\mathbb{R}}$, $\lambda=\exp\theta$ e $\boxed{\psi(\theta)=\lambda/2+\log\lambda=\frac{e^{\theta}}{2}+\theta}$.

2. $\mathrm{E}_\theta(Y)=\dfrac{d\psi}{d\theta}=\dfrac{1}{2}\exp\theta+1=\lambda/2+1$ e $\mathrm{Var}_\theta(Y)=\dfrac{d^2\psi}{d\theta^2}=\dfrac{1}{2}\exp\theta=\lambda/2$.

3. (a) $L(\lambda,\underline{x})\propto\exp(\sum_i y_i\log\lambda)-n(\frac{\lambda}{2}+\log\lambda))$

   (b) Sappiamo che $\sum_i Y_i$ è di massima verosimiglianza per $\dfrac{d\psi(\theta)}{d\theta}$. Per il teorema di invarianza degli stimatori di massima verosimiglianza si ha $\sum Y_i=n(\dfrac{\hat\lambda_{MV}}{2}+1)$, da cui $\hat\lambda_{MV}=2\dfrac{\sum_i Y_i}{n}-2$.
   Inoltre $\mathrm{E}(\frac{2}{n}\sum_i Y_i-2)=2n\,\mathrm{E}(Y_1)/n-2=2(\lambda/2+1)-2=\lambda$ e quindi lo stimatore di massima verosimiglianza di λ è corretto.

   (c) Poichè $\hat\lambda_{MV}$ il limite inferiore di Cramér Rao coincide con l'inverso dell'informazione di Fisher, che vale
   $$\mathcal{I}_\lambda(\hat\lambda_{MV})=\mathrm{E}((\frac{dl}{d\lambda})^2)=-\mathrm{E}(\frac{d^2l}{d\lambda^2})$$
   Poichè $\dfrac{d^2l}{d\lambda^2}=\dfrac{1}{\lambda^2}(n-\sum_i Y_i)$ si ha $\mathcal{I}_\lambda(\hat\lambda_{MV})=\frac{1}{\lambda^2}(-n+\mathrm{E}(\sum_i Y_i))=\frac{n}{\lambda^2}(\mathrm{E}(Y_1)-1)=\frac{n}{\lambda^2}(\frac{\lambda}{2}+1-1)=$ $n/(2\lambda)$ e quindi il limite inferiore di Cramér-Rao per $2\dfrac{\sum_i Y_i}{n}-2$ è $2\lambda/n$.

Esercizio

Sia $X_1, \dots, X_n$ un insieme di variabili aleatorie indipendenti ed identicamente distribuite ed il modello statistico costruito sulla densità

$$f_{X_1}(x) = \frac{1-\alpha}{x^\alpha} \qquad \text{con } x \in]0,1[\text{ e } 0 < \alpha < 1$$

1. Determinare una statistica sufficiente V utilizzando il teorema di fattorizzazione di Neyman-Fisher.
2. Verificare che il modello statistico è di classe esponenziale. Individuare il parametro naturale θ, la statistica sufficiente T e la funzione dei cumulanti $\psi(\theta)$.
3. Dimostrare che V è statistica sufficiente minimale e completa.
4. Calcolare i valori attesi e le varianze di T rispetto alle leggi del modello statistico.
5. Determinare lo stimatore di massima verosimiglianza di α e di $n/(\alpha-1)$.
6. Determinare lo stimatore dei momenti di α.
7. Calcolare l'informazione di Fisher del modello statistico.
8. Calcolare il limite inferiore di Cramér-Rao per T.

Traccia di soluzione

1. $V(\underline{X}) = \prod_{i=1}^n X_i$ infatti
$$f_X(\underline{x}) = \prod_{i=1}^n f(x_i) = \frac{(1-\alpha)^n}{(\prod_{i=1}^n x_i)^\alpha}(0 < x_{(1)})(x_{(n)} < 1)$$
2. $\log \mathcal{L}(\alpha; \underline{x}) = n\log(1-\alpha) - \alpha \log \prod x_i$ da cui $\theta = \alpha$, $T(\underline{X}) = -\log \prod_i X_i = \sum_i \log X_i = -\log V(\underline{X})$ e $\psi(\theta) = -n\log(1-\theta)$.
3. T è statistica sufficiente minimale e completa per $\theta = \alpha$. V è funzione composta di T e di una funzione invertibile $(-\log / \exp -)$. Quindi V è sufficiente minimale e completa per θ.
4. $\mathrm{E}_\alpha(T) = \mathrm{E}_\theta(T) = \dfrac{\partial \psi(\theta)}{\partial \theta} = n/(1-\theta) = n/(1-\alpha)$.
5. $\mathrm{Var}_\alpha(T) = \dfrac{\partial^2 \psi(\theta)}{\partial \theta^2} = n/(1-\alpha)^2$.
6. $\dfrac{\partial l}{\partial \alpha} = -\dfrac{n}{1-\alpha} - \sum_i \log x_i = 0$ se e solo se $\hat{\alpha} = 1 + \dfrac{1}{\sum_i \log X_i}$ che è massimo infatti $\dfrac{\partial^2 l}{\partial \alpha^2} = -n/(1-\alpha)^2 < 0$.

   Per il teorema di invarianza degli stimatori di massima verosimiglianza si ha $\widehat{\dfrac{n}{\alpha-1}} = \dfrac{n}{\hat{\alpha}-1} = \sum_i \log X_i = -T(\underline{X})$

Esercizio Siano $Y_1, \dots, Y_n$ copie di una variabile aleatoria con densità

$$f(y; \alpha) = \alpha \log(\alpha) \alpha^{-y}$$

con $y \in [1, +\infty[$ e $\alpha \in (1, +\infty)$.

1. Verificare che il modello statistico è di classe esponenziale; precisare lo spazio del parametro naturale e la statistica sufficiente T. [2]
2. Calcolare il valore atteso della statistica sufficiente e la sua varianza. [3]
3. Dire quale funzione di α è stimata in modo non distorto da T. [2]
4. Calcolare, se esiste, lo stimatore di massima verosimiglianza per α. [3]

Traccia di soluzione

$$l(\alpha; y_1, \dots, y_n) = \prod_{i=1}^{n} l_i(\alpha; y_i) \qquad \text{con } y_i \in [1, +\infty[\text{ e } \alpha \in]1, +\infty[$$

1. $l_i(\alpha; y_i) = \log(\alpha \log(\alpha)) - y_i \log \alpha$ da cui $\theta = \log \alpha$ e $\alpha = e^\theta$ con $\alpha \in]1, +\infty[$ e $\theta \in]0, +\infty[$. Inoltre $T(Y_1, \dots, T_n) = -\sum_{i=1}^n Y_i$.
2. $\psi(\theta) = -\log(\alpha \log(\alpha)) = -\log(e^\theta \theta) = -\theta - \log \theta$ per $n = 1$ da cui $\mathrm{E}(T(\underline{Y})) = \frac{d}{d\theta} \psi(\theta) = -1 - 1/\theta$ e $\mathrm{Var}(T(\underline{Y})) = \frac{d^2}{d\theta^2} \psi(\theta) = 1/\theta^2$.
3. $\mathrm{E}(T(\underline{Y})) = \frac{d}{d\theta} \psi(\theta) = -1 - 1/\log \alpha = -\frac{\log \alpha + 1}{\log \alpha}$
4. $\hat{Y} = -\sum_i Y_i$ è MLE per $\dfrac{\partial \psi}{\partial \theta} = -n \left(\dfrac{1 + \theta}{\theta} \right) = -n \left(1 + \dfrac{1}{\log \alpha} \right)$. Per il teorema di invarianza degli MLE si ha $\hat{Y} = -n \left(1 + \dfrac{1}{\log \alpha} \right)$ se e solo se $\left(\frac{\hat{Y}}{-n} - 1 \right)^{-1} = \log \alpha$ se e solo se $\hat{\alpha} = \exp \left(\dfrac{-n}{n + \hat{Y}} \right) = \exp \left(\dfrac{1}{\bar{Y} - 1} \right)$.

11.1 Altri esercizi d'esame

Esercizio
Siano $\theta > 0$ e $X_1, \ldots, X_n$ un campione i.i.d. Unif $([1, 1+\theta])$.

1. Calcolare la funzione di verosimiglianza. [3]
2. Determinare una statistica sufficiente per θ. [3]
3. Determinare lo stimatore di massima verosimiglianza di θ. [3]
4. Determinare lo stimatore di massima verosimiglianza di $1/\theta$. [3]
5. Determinare una statistica sufficiente per $1/\theta$. [3]
6. Calcolare lo stimatore dei momenti di θ. [3]
7. Calcolare lo stimatore dei momenti di $1/\theta$. [3]
8. Per $n = 5$ e $x_1 = 1.2, x_2 = 1, x_3 = 1.7, x_4 = 1.3, x_5 = 1.55$ calcolare le stime di verosimiglianza e dei momenti di θ. Quale delle due preferite e perché. [3]

Esercizio
Sia Y una variabile aleatoria con densità rispetto alla misura di Lebesgue

$$\frac{1}{\sqrt{2\pi}}\, x^{-3/2}\ \exp\left(-\frac{(x-\mu)^2}{2\mu^2 x}\right) \quad \text{con } x \in (0, \infty)$$

dipendente da un parametro $\mu \in (0, \infty)$.

1. Scrivere la verosimiglianza di Y in forma di modello esponenziale, indicare parametro canonico e statistica sufficiente. [2]
2. Calcolare valore atteso e varianza della statistica sufficiente. [3]
3. Si consideri un n-campione di Y.
   (a) Calcolare valore atteso e varianza della statistica sufficiente. [1]
   (b) Calcolare lo stimatore di massima verosimiglianza del parametro. [3]

Esercizio
Sia $X_1, \ldots, X_n$ un campione indipendente dalla densità

$$f_X(x) = \begin{cases} \exp(-(x-\theta)) & \text{per } x \geq \theta \\ 0 & \text{altrimenti.} \end{cases}$$

1. Verificare che f_X è una densità di probabilità e disegnarne il grafico.
2. E' un modello di classe esponenziale?
3. Determinare una statistica sufficiente e unidimensionale per θ.
4. Fare il grafico della funzione di verosimiglianza e della log-verosimiglianza, se possibile.
5. Calcolare uno stimatore di massima verosimiglianza per θ; se possibile.
6. Discutere l'unicità dello stimatore di massima verosimiglianza per θ.
7. Calcolare uno stimatore dei momenti di θ.

Esercizio
Sia $X_1, \ldots, X_n$ un campione indipendente $Uniforme(]\theta, \theta+1[)$ con $\theta \in R$.

1. E' un modello di classe esponenziale?
2. Calcolare la funzione di verosimiglianza.
3. Determinare una statistica sufficiente per θ.
4. Fare il grafico della funzione di verosimiglianza.
   Hint: Fare attenzione ai valori assunti dalla statistica sufficiente.
5. Calcolare uno stimatore di massima verosimiglianza per θ, se esiste.
6. Discutere l'unicità dello stimatore di massima verosimiglianza per θ.
7. Facoltativo. Dimostrare che la statistica $R = X_{(n)} - X_{(1)}$ è ancillare per il nostro campione, e più in generale per le famiglie di scala.
   Hint: Verificare prima che per le variabili ausiliarie $Z_i = X_i - \theta$, $Prob(Z_i \leq x)$ non dipende da θ, per ogni x reale. Quindi per valutare $Prob(R \leq r)$ scrivere R in funzione delle Z_i, $i = 1, \ldots, n$.

Esercizio
Determinare una statistica sufficiente bidimensionale per un campione casuale di legge proporzionale a $\exp(-x\theta)(0 < x < \theta)$.

Esercizio
Si consideri un campione casuale di legge proporzionale a $\exp(-\frac{x-\alpha}{\beta})(0 < \alpha < x < \beta)$.

1. Determinare una statistica sufficiente bidimensionale.
2. Noto β, determinare uno stimatore di massima verosimiglianza per α.
3. Noto α, determinare uno stimatore di massima verosimiglianza per β e calcolarne il valore atteso. Hint: la somma di esponenziali i.i.d. segue una legge gamma.s

Esercizio
Sia $X_1, \ldots, X_n$ un campione di legge $Uniforme(]\frac{1}{\theta}, \theta[)$ con θ numero reale e $\theta > 1$.

1. E' un modello di classe esponenziale?
2. Calcolare la funzione di verosimiglianza.
3. Calcolare uno stimatore di massima verosimiglianza per θ, se esiste.
4. Determinare condizioni sul campione osservato $\{x_1, \ldots, x_n\}$ per cui lo stimatore di massima verosimiglianza non può essere calcolato.
5. Per $n = 1$ determinare un valore del campione osservato per cui lo stimatore di massima verosimiglianza non può essere calcolato.
6. Determinare due statistiche sufficienti (multidimensionali) per θ.

Esercizio
Sia $\theta \in]-\infty, +\infty[$ e siano $X_1, \ldots, X_n$ variabili aleatorie indipendenti di densità $f_{X_i}(x) = e^{i\theta - x}(x \geq i\theta)$

1. Indicare se è un modello di classe esponenziale e giustificare la risposta.
2. $min(X_i/i)$ è statistica sufficiente per θ?
3. Calcolare uno stimatore di massima verosimiglianza per θ, se esiste, e discuterne l'unicità.

4. Calcolare se esiste lo stimatore dei momenti di θ.

5. Calcolare una stima di θ per il seguente campione di taglia nove $x_1 = 3$, $x_2 = 6.4$, $x_3 = 12$, $x_4 = 10$, $x_5 = 15$, $x_6 = 15$, $x_7 = 21$, $x_8 = 24$, $x_9 = 27$.

6. Proporre un'altra stima per θ e motivare la proposta.

Esercizio
Per $\alpha, \beta > 0$ e $x \in [\alpha, +\infty[$ sia

$$f_X(x; \alpha, \beta) = \frac{\beta \alpha^\beta}{x^{\beta+1}}$$

la densità di una variabile aleatoria univariata X.

1. Il modello statistico $\{f_X(\cdot; \alpha, \beta) : \alpha > 0, \beta > 0\}$ è di classe esponenziale? Giustificare la risposta.

2. Calcolare i momenti di X e discuterne l'esistenza.

3. Per $\beta > 2$ calcolare uno stimatore dei momenti di β.

4. Per un campione di taglia uno, calcolare uno stimatore di massima verosimiglianza per α, noto β.

Esercizio
Sia $X_1, \ldots, X_n$ un campione indipendente da

$$f(x; \theta) = \theta / x^2 \qquad \text{per } \theta \leq x < \infty$$

1. Per quali valori di θ è una densità?

2. E' un modello di classe esponenziale? Giustificare la risposta.

3. Determinare una statistica sufficiente per θ.

4. Determinare lo stimatore di massima verosimiglianza per θ.

Esercizio
Sia $\theta \in \{0, 1\}$ e sia $X_1, \ldots, X_n$ un campione i.i.d. da una delle due seguenti leggi di probabilità: se $\theta = 0$

$$f(x; \theta) = \begin{cases} 1 & \text{se } 0 < x < 1 \\ 0 & \text{altrimenti} \end{cases}$$

mentre se $\theta = 1$

$$f(x; \theta) = \begin{cases} 1/(2\sqrt{x}) & \text{se } 0 < x < 1 \\ 0 & \text{altrimenti} \end{cases}$$

Determinare lo stimatore di massima verosimiglianza per θ.

Esercizio
Siano Y una variabile aleatoria discreta che assume valori in $\{100, 101, \ldots, 120\}$ tali che

$$\mathbb{P}(Y = y) = \binom{20}{y - 100} \left(\frac{\theta}{1 - \theta}\right)^y \frac{(1 - \theta)^{120}}{\theta^{100}} \qquad \theta \in (0, 1)$$

1. Scrivere la verosimiglianza del modello per ciascuna variabile aleatoria Y_i; verificare che appartiene alla famiglia dei modelli esponenziali, indicare parametro canonico e statistica sufficiente.

2. Calcolare valore atteso e varianza della variabile aleatoria Y.

3. Si consideri un n campione di Y.

   (a) Scrivere la verosimiglianza del modello per il campione.

(b) Calcolare lo stimatore di massima verosimiglianza del parametro θ. Quale funzione di θ, indicata con $g(\theta)$ è stimata in massima verosimiglianza dalla statistica sufficiente.

(c) Calcolare il limite inferiore di Cramér-Rao per la varianza dello stimatore di massima verosimiglianza di $g(\theta)$.

Esercizio
Sia $X_1, \ldots, X_n$ un campione estratto da una popolazione su cui è definita una variabile con densità

$$f_\beta(x) = \beta x^{\beta-1} \qquad x \in [0,1],\ \beta > 0$$

e zero altrove.

1. Scrivere la verosimiglianza per il campione e indicare se si tratta di una famiglia esponenziale.
2. Indicare una statistica sufficiente e minimale T.
3. Determinare uno stimatore di massima verosimiglianza di β e quello di β^2.
4. Indicare la funzione di β che è stimata in modo non distorto dalla statistica sufficiente T.
5. Scrivere l'informazione di Fisher.
6. Determinare uno stimatore β con il metodo dei momenti.
7. Posto $n = 100$ e $\sum_{i=1}^{100} \log x_i = -136$, costruire un test del rapporto di verosimiglianza di $H_0 : \beta = 1$ contro $H_1 : \beta \neq 1$ al livello $\alpha = 5\%$. [solo alcuni anni]
8. Sapendo che le variabili aleatorie $-\log(X_i)$ hanno distribuzione Esponenziale(β), indicare se lo stimatore di massima verosimiglianza è non distorto.

Esercizio
Sia $X_1, \ldots, X_n$ un campione estratto da una popolazione su cui è definita una variabile con densità Gamma(θ_1, θ_2)

$$f(x) = \frac{\theta_2^{\theta_1}}{\Gamma(\theta_2)} x^{\theta_1 - 1} \exp(-\theta_2 x)$$

con $\theta_1, \theta_2 > 0$ and $x > 0$. Scrivere la verosimiglianza per il campione e indicare se si tratta di una famiglia esponenziale.