**ESERCIZI DI INFERENZA STATISTICA** [1]

**Fulvio De Santis - Marco Perone Pacifico - Luca Tardella - Isabella Verdinelli**

Anno Accademico 2015-2016

## I PARTE: VARIABILI ALEATORIE, VEROSIMIGLIANZA E SUFFICIENZA

### Variabili aleatorie

**Esercizio 1.** Un rappresentante percorre frequentemente in automobile il tratto tra New York e Boston. Si ipotizza che il tempo di percorrenza sia una variabile aleatoria con distribuzione normale di valore atteso 4.3 ore e varianza pari a $0.2^2$ ore. Determinare la probabilità che un viaggio del rappresentante duri

*a*) più di 4.5 ore;

*b*) meno di 4 ore.

**Esercizio 2.** Un quiz è costituito da dieci domande a risposta multipla. Per ciascuna domanda sono previste 3 risposte, di cui solamente una è esatta. Per superare il test è necessario rispondere correttamente ad almeno sei domande. Supponendo di scegliere a caso le risposte a tutte le domande,

a) determinare la probabilità di superare il test;

b) stabilire se la probabilità di rispondere esattamente a tutte le domande del test è superiore alla probabilità di sbagliare tutte le risposte.

**Esercizio 3.** Si consideri un campione di 16 persone che soffrono di emicrania. Supponendo di utilizzare su questi pazienti un farmaco che, in base alla ricerca, si ritiene efficace nell' 80% dei casi, determinare:

a) la probabilità che abbia effetto su tutte le unità del campione;

b) la probabilità che il farmaco sia efficace su almeno 14 pazienti del campione;

c) il numero medio di pazienti del campione che ci si aspetta trovino giovamento dall'uso del farmaco.

d) Supponendo di somministrare il farmaco ad una popolazione di 1000 soggetti, determinare la probabilità che questo abbia effetto su almeno 750 individui.

**Esercizio 4.** Il tasso di guarigione garantito da un farmaco per una determinata malattia è pari al 70%.

*a*) Qual è la probabilità che, su 10 pazienti curati con il farmaco considerato, più di 8 guariscano?

*b*) Qual è il numero medio di pazienti per i quali ci si aspetta la guarigione?

**Esercizio 5.** Il 65% dei laureati di una facoltà viene assunto entro un anno dalla laurea. Supponendo di considerare $n = 9$ laureati della facoltà in esame, determinare:

a) il numero medio di laureati assunti entro un anno;

b) la probabilità che almeno cinque di questi trovino lavoro entro un anno;

c) la probabilità che al più due di questi trovi lavoro entro un anno.

**Esercizio 6.** Una variabile statistica $X$ ha distribuzione normale di valore atteso $\mu$ e varianza $\sigma^2$ incognite. Determinare il valore di $\mu$ e $\sigma$ sapendo che la probabilità che $X$ assuma valori minori di 245 è pari a 0.33, e che la probabilità che $X$ assuma valori superiori a 260 è pari a 0.48.

**Esercizio 7.** Il 2,35 per cento delle persone adulte di un collettivo è mancina. Determinare la probabilità che, su una scolaresca di 120 studenti, ne siano mancini

[1] *Gli esercizi contrassegnati dall'asterisco si riferiscono a prove di esame assegnate negli scorsi anni accademici.*

(a) 3;
(b) almeno 3;
(c) al massimo 3.

**Esercizio 8*.** In una indagine del 1994, il Census Bureau degli U.S.A. ha stabilito che il 70% dei cittadini americani aveva stipulato un contratto di assicurazione sanitaria privata. Sulla base di tale valutazione, qual era in quell'anno la probabilità che, su $n = 5$ cittadini scelti casualmente, al più 2 avessero il contratto?

**Esercizio 9*.** In una indagine del 1994, il Census Bureau degli U.S.A. ha stabilito che il 30% dei cittadini americani non aveva stipulato un contratto di assicurazione sanitaria privata. Sulla base di tale valutazione, qual era in quell'anno la probabilità che, su $n = 5$ cittadini scelti casualmente, almeno 3 persone non avessero il contratto?

**Esercizio 10*.** Da un'indagine demografica condotta dal *Census Bureau* statunitense, risulta che il 9.96 % dei cittadini americani di età superiore a 18 anni è di origine ispanica. Si consideri un campione casuale di $n = 1200$ cittadini americani e determinare

a) il numero medio di cittadini ispanici in un campione di tale ampiezza;

b) la probabilità che di un campione di tale ampiezza facciano parte meno di 100 cittadini ispanici.

## Campioni Casuali, Statistiche Campionarie e Proprietà

**Esercizio 1.** Sia $X_1, \ldots, X_n$ un campione casuale da una popolazione di Poisson di parametro incognito $\theta$. Determinare (in funzione di $\theta$) la probabilità di osservare il seguente campione di dimensione $n = 5$:

$$\mathbf{x}_n = (2, 3, 1, 5, 5).$$

Che valore assume la probabilità considerata se $\theta = 2$? E se invece $\theta = 3$?

**Esercizio 2.** Sia $X_1, \ldots, X_n$ un campione casuale da una popolazione di esponenziale di parametro incognito $\theta$. Determinare la funzione di densità congiunta (in funzione di $\theta$) in corrispondenza del seguente campione di dimensione $n = 5$:

$$\mathbf{x}_n = (1.12, 0.88, 0.13, 0.42, 0.36).$$

Che valore assume la densità congiunta se si assume $\theta = 0.5$? E se invece si assume $\theta = 1$?

**Esercizio 3.** La temperatura alla quale un termostato scatta ha distribuzione normale con varianza $\sigma^2$. Supponendo di effettuare $n = 5$ controlli di qualità, calcolare

a) $P(\frac{S_n^2}{\sigma^2} \leq 1.8)$;

b) $P(0.85 \leq \frac{S_n^2}{\sigma^2} \leq 1.15)$,

dove $S_n^2$ reppresenta la varianza campionaria corretta del campione relativo alle 5 prove effettuate.

**Esercizio 4.** Il tempo di vita di un certo componente elettrico è una v.a. con valore atteso $\mu = 100$ e deviazione standard $\sigma = 20$. Se si provano $n = 16$ componenti di questo tipo (tra loro indipendenti), quanto vale approssimativamente la probabilità che la media campionaria delle loro durate di vita sia

a) minore di 104;

b) compresa tra 98 e 104.

**Esercizio 5.** In una azienda, si suppone che il numero di ore di straordinario degli impiegati in un mese sia una v.a. con valore atteso $\mu = 5.75$ ore e deviazione standard $\sigma = 0.48$ ore. Se si considera un campione casuale di $n = 36$ impiegati, qual è la probabilità che le ore complessive del loro lavoro straordinario sia compreso tra 202 e 210 ore?

**Esercizio 6.** Sia $X$ una v.a. discreta con la seguente distribuzione:

$$P(X = 0) = 0.2, \quad P(X = 1) = 0.3, \quad P(X = 2) = 0.5.$$

Si determini, nel caso di un campione casuale di dimensione $n = 2$, la distribuzione della media campionaria, $\bar{X}$. Si calcoli quindi il valore atteso e la varianza della v.a. $\bar{X}$.

**Esercizio 7*.** Si suppone che il numero di telefonate che un operatore di un grande centralino riceve in un'ora del giorno sia una variabile aleatoria di Poisson di parametro $\theta = 12$. Considerato un campione casuale $X_1, \ldots, X_n$ di $n = 100$ operatori, determinare la probabilità (approssimazione) che il numero complessivo $Y = \sum_{i=1}^{100} X_i$ di telefonate a cui rispondono i 100 operatori in un'ora sia compreso tra 1150 e 1280 telefonate.
(Sugg.: ricordare che per una v.a. $X$ con distribuzione di Poisson di parametro $\theta$, si ha che $\mathrm{E}(X) = \mathrm{V}(X) = \theta$).

**Esercizio 8*.** Si suppone che il numero di clienti che si presentano a uno sportello di una grande banca in un giorno dell'anno sia una variabile aleatoria di Poisson di parametro $\theta = 30$. Considerato un campione casuale $X_1, \ldots, X_n$ di $n = 40$ sportelli, determinare la probabilità (approssimazione) che il numero complessivo $Y = \sum_{i=1}^{40} X_i$ di clienti serviti dai 40 sportelli in un giorno sia compreso tra 1180 e 1270.
(Sugg.: ricordare che per una v.a. $X$ con distribuzione di Poisson di parametro $\theta$, si ha che $\mathrm{E}(X) = \mathrm{V}(X) = \theta$).

**Esercizio 9*.** Si consideri un campione casuale di $n = 4$ osservazioni provenienti da una distribuzione normale di parametri $\mu = 8$ e $\sigma^2 = 8$. Date le tre statistiche campionarie:

$$T_1(\mathbf{X}_n) = \bar{X}_n, \qquad T_2(\mathbf{X}_n) = S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2, \qquad T_3(\mathbf{X}_n) = \frac{1}{4}\bar{X}_n + \frac{3}{4}S_n^2,$$

calcolarne il valore atteso e la varianza.

(**Sugg.** Ricordare che se una v.a. $Y \sim \chi^2_\nu$, si ha che $E[Y] = \nu$ e che $V[Y] = 2\nu$)

**Esercizio 10*** Sia $X_1, \ldots, X_n$ un campione casuale tale che, $\forall i = 1, \ldots, n$,

$$E[X_i] = \theta \qquad \text{e} \qquad V[X_i] = \frac{1}{2}\theta^2$$

a) Determinare valore atteso e varianza delle variabili aleatorie $\frac{1}{2}\bar{X}_n$ e $2X_1 + 2X_n - 3$.

b) Determinare l'approssimazione asintotica per la distribuzione di $\bar{X}_n$.

c) Supponendo che $\theta = 2$ e $n = 25$, determinare la probabilità (approssimazione) che la v.a. $\bar{X}_n$ assuma valori nell'intervallo $(2, 2.4)$.

**Esercizio 11*.** Sia $X_1, \ldots, X_n$ un campione casuale tale che, $\forall i = 1, \ldots, n$,

$$E[X_i] = \frac{\theta}{\theta + 1} \qquad \text{e} \qquad V[X_i] = \frac{\theta}{\theta^2(\theta + 2)}.$$

a) Determinare valore atteso e varianza delle variabili aleatorie $\bar{X}_n$ e $2\sum_{i=1}^{n} X_i - 3$.

b) Determinare l'approssimazione asintotica per la distribuzione di $\bar{X}_n$.

c) Supponendo che $\theta = 0.2$ e $n = 25$, determinare la probabilità (approssimazione) che la v.a. $\bar{X}_n$ assuma valori nell'intervallo $(1/2, 1)$.

**Esercizio 12*.** Sia $X_1, X_2, \ldots, X_n$ un campione casuale proveniente da una popolazione di Poisson di parametro $\theta = 0.1$.

a) Determinare la distribuzione esatta e l'approssimazione normale per la statistica campionaria $U_n = 2Y_n - 3$, dove $Y_n = \sum_{i=1}^n X_i$.

b) Determinare la probabilità che $U_n$ assuma valori positivi, considerando $n = 20$.

## Verosimiglianza e Sufficienza

**Esercizio 1.** Si consideri un campione casuale di $n$ osservazioni da una popolazione $\mathrm{N}(0, \theta)$ (qui $\theta$ rappresenta la varianza incognita della v.a.). Supponendo di avere osservato il seguente campione di 10 osservazioni:

$$\mathbf{x}_n = (2.52, 0.76, 1.55, 0.98, 4.03, 0.09, -2.27, 1.67, -0.54, -0.27),$$

a) scrivere la funzione di verosimiglianza, $L(\theta; \mathbf{x}_n)$, del parametro $\theta$;

b) determinare la stima di massima verosimiglianza di $\theta$.

c) In base al campione osservato, risulta maggiormente verosimile per $\theta$ il valore $\theta_1 = 3$ o $\theta_2 = 5$?

**Esercizio 2.** Si consideri un campione casuale di $n$ osservazioni da una popolazione $X$ con funzione di densità

$$f_X(x; \theta) = \frac{\theta\, \mathrm{e}^{\theta}}{x^{\theta+1}}, \qquad x > \mathrm{e}, \quad \theta > 0.$$

a) Scrivere la funzione di verosimiglianza, $L(\theta; \mathbf{x}_n)$, del parametro $\theta$.

b) Verificare analiticamente che lo stimatore di massima verosimiglianza di $\theta$ risulta essere:

$$\hat{\theta} = \frac{n}{\sum_{i=1}^n \ln x_i - n}.$$

c) Supponendo di avere osservato il seguente campione $\mathbf{x}_n$ di $n = 10$ osservazioni:

3.595 5.048 3.410 3.288 5.585 4.429 3.691 4.372 4.044 5.986,

determinare la stima di massima verosimiglianza di $\theta$.

**Esercizio 3.** Sia $X_1, \ldots, X_n$ un campione casuale da una popolazione con funzione di densità

$$f_X(x; \theta) = e^{-(x-\theta)}\, I(x)_{(\theta, +\infty)}, \qquad \theta > 0.$$

Determinare lo stimatore di massima verosimiglianza di $\theta$.

**Esercizio 4.** Si consideri un campione casuale da una distribuzione di Poisson di parametro incognito $\theta$. Supponendo di avere osservato un campione di dimensione $n = 10$, tale che la somma dei valori osservati sia pari a 28, stabilire quale tra i seguenti è il valore più verosimile per il parametro incognito $\theta$:

$$\theta_1 = 2, \quad \theta_2 = 3, \quad \theta_3 = 4.$$

**Esercizio 5*.** Si consideri un campione casuale di $n$ osservazioni da una popolazione $X$ con funzione di densità

$$f_X(x; \theta) = 2\theta\, x\, e^{-\theta x^2}\, I_{(0, +\infty)}(x), \qquad \theta > 0.$$

In corrispondenza di un generico campione osservato $\mathbf{x}_n = (x_1, \ldots, x_n)$,

a) scrivere la funzione di verosimiglianza, $L(\theta; \mathbf{x}_n)$, del parametro $\theta$;

b) individuare il nucleo della funzione $L(\theta; \mathbf{x}_n)$ e una statistica sufficiente per il modello;

c) verificare analiticamente che lo stima di massima verosimiglianza di $\theta$ risulta essere:

$$\hat{\theta}(\mathbf{x}_n) = \frac{n}{\sum_{i=1}^n x_i^2}.$$

**Esercizio 6*.** Si consideri un campione casuale di $n$ osservazioni da una popolazione $X$ con funzione di densità

$$f_X(x;\theta) = \frac{2}{\theta}\, x\, e^{-\frac{x^2}{\theta}}\, I_{(0,+\infty)}(x), \qquad \theta > 0.$$

In corrispondenza di un generico campione osservato $\mathbf{x}_n = (x_1, \ldots, x_n)$,

a) scrivere la funzione di verosimiglianza, $L(\theta; \mathbf{x}_n)$, del parametro $\theta$;

b) individuare il nucleo della funzione $L(\theta; \mathbf{x}_n)$ e una statistica sufficiente per il modello;

c) verificare analiticamente che lo stima di massima verosimiglianza di $\theta$ risulta essere:

$$\hat{\theta}(\mathbf{x}_n) = \frac{\sum_{i=1}^n x_i^2}{n}.$$

**Esercizio 7*.** Sia $X$ una variabile aleatoria con funzione di densità

$$f_X(x;\theta) = \frac{4x^2}{\theta^3\sqrt{\pi}}\, e^{-\frac{x^2}{\theta^2}}\, I_{(0,+\infty)}(x) \qquad \theta > 0.$$

a) Determinare una statistica sufficiente.

b) Determinare lo stimatore di massima verosimiglianza per $\theta$.

**Esercizio 8*.** Sia $\mathbf{x}_n = (3, 4, 2, 7, 4, 5, 8, 1, 0, 0)$ un campione casuale dalla popolazione con funzione di massa di probabilità

$$f_X(x;\theta) = e^{-\theta}\, \frac{\theta^x}{x!} \qquad x = 0, 1, 2, \ldots \qquad \theta > 0.$$

a) Calcolare la stima di massima verosimiglianza del parametro $\theta$;

b) Calcolare la probabilità dell'evento $\{X_1 = 0\}$ e, osservando che si tratta di una funzione del parametro incognito $\theta$, calcolare la stima di massima verosimiglianza di tale quantità.

**Esercizio 9*.** Sia $\mathbf{X}_n = (X_1, \ldots, X_n)$ un campione casuale dalla popolazione con funzione di densità

$$f_X(x;\theta) = \frac{2x}{\theta^2} \qquad 0 < x < \theta \qquad \theta > 0.$$

a) Calcolare $E_\theta(\bar{X})$, dove $\bar{X} = \sum_{i=1}^n X_i/n$ indica la media campionaria.

b) Determinare la funzione di verosimiglianza e lo stimatore di massima verosimiglianza del parametro $\theta$.

**Esercizio 10*.** Sia $X_1, \ldots, X_n$ un campione casuale da una popolazione $X$ con funzione di densità

$$f_X(x;\theta) = 2\,\theta^2\, x^{-3}\, I_{(\theta,\infty)}(x), \qquad \theta > 0.$$

a) Scrivere la funzione di verosimiglianza di $\theta$ per un campione osservato $\mathbf{x}_n$.

b) Determinare, se esiste, una statistica sufficiente per $\theta$.

c) Determinare la stima di massima verosimiglianza di $\theta$.

**Esercizio 11.** Sia $\mathbf{X}_n = (X_1, \ldots, X_n)$ un campione casuale dalla popolazione con distribuzione di probabilità

$$f_X(x; \theta, \alpha) = \frac{\Gamma(\alpha + x)}{\Gamma(x+1)\Gamma(\alpha)}\, \theta^{\alpha}(1-\theta)^x, \qquad x = 0, 1, 2, \ldots, \quad \theta \in (0,1), \quad \alpha > 0.$$

Assumendo che $\alpha$ sia una quantità nota, si determini, se esiste, una statistica sufficiente.

**Esercizio 12*.** Con riferimento all'esercizio precedente, si supponga che $\alpha = 1$ e si consideri il campione

$$\mathbf{x}_n = (3, 2, 7, 3, 5).$$

a) Determinare la funzione di verosimiglianza di $\theta$ per il campione considerato.

b) Determinare la stima di verosimiglianza per il parametro $\theta$;

c) Verificare se, alla luce dei dati osservati, risulta più verosimile per $\theta$ il valore $\theta_1 = 0.3$ oppure il valore $\theta_2 = 0.6$.

**Esercizio 13.** Sia $\mathbf{X}_n = (X_1, \ldots, X_n)$ un campione casuale dalla popolazione con funzione di densità

$$f_X(x; \theta) = \frac{2}{\theta}\left(1 - \frac{x}{\theta}\right) I_{[0,\theta]}(x), \qquad \theta > 0.$$

a) Determinare la funzione di verosimiglianza.

b) Verificare che $E_\theta(X) = \frac{\theta}{3}$.

**Esercizio 14.** In un processo di controllo di qualità emerge che, su 371 pezzi controllati, 18 sono difettosi. Trattando il campione considerato come una realizzazione di un campione casuale,

a) determinare la funzione di verosimiglianza;

b) calcolare il valore della stima di massima verosimiglianza di $p$, proporzione dei pezzi difettosi nella popolazione da cui proviene il campione.

**Esercizio 15*.** Si consideri un campione casuale di $n$ osservazioni da una popolazione $X$ con funzione di densità

$$f_X(x; \theta) = \frac{1}{\sqrt{2\pi}}\, \theta^{-1}\, \frac{1}{x}\, \exp\{-\frac{1}{2\theta^2}(\ln x - 3)^2\}\, I_{(0,+\infty)}(x), \qquad \theta > 0.$$

In corrispondenza di un generico campione osservato $\mathbf{x}_n = (x_1, \ldots, x_n)$,

a) determinare l'espressione la funzione di verosimiglianza, $L(\theta; \mathbf{x}_n)$, del parametro $\theta$;

b) individuare il nucleo della funzione $L(\theta; \mathbf{x}_n)$ e una statistica sufficiente per il modello;

c) determinare la stima di massima verosimiglianza di $\theta$;

d) calcolare il valore della stima di massima verosimiglianza per il seguente campione di dimensione $n = 3$: $\mathbf{x}_3 = (e^1, e^2, e^4)$.

**Esercizio 16*.** Sia $X_1, \dots, X_n$ un campione casuale dalla popolazione con funzione di densità:

$$f_X(x;\theta) = \theta\, x^{-(\theta+1)}, \qquad x > 1, \qquad \theta > 0,$$

e sia $\mathbf{x}_n$ un generico campione osservato.

a) Determinare la funzione di verosimiglianza di $\theta$ e indicarne il nucleo.

b) Verificare che $T_0(\mathbf{X}_n) = \prod_{i=1}^{n} X_i$ è una statistica sufficiente.

c) Determinare lo stimatore di massima verosimiglianza di $\theta$, $T_1(\mathbf{X}_n)$.

d) Per il campione di $n = 3$ osservazioni $\mathbf{x}_3 = (2, 2, 3)$, calcolare la stima di massima verosimiglianza di $\theta$ e verificare quale, tra i valori $\theta_0 = 2$ e $\theta_1 = 3$ risulta più verosimile.

e) Determinare la stima di massima verosimiglianza di $h(\theta) = \sqrt{(\theta)} + 1$.

**Esercizio 17*.** Sia $X_1, \dots, X_n$ un campione casuale dalla popolazione con distribuzione di probabilità:

$$f_X(x;\theta) = \left(\frac{\theta}{2}\right)^{|x|} (1-\theta)^{1-|x|}, \qquad x = -1, 0, 1 \qquad \theta \in (0, 1).$$

a) Determinare la funzione di verosimiglianza di $\theta$ e indicarne il nucleo.

b) Determinare una statistica sufficiente.

c) Determinare lo stimatore di massima verosimiglianza di $\theta$.

d) Per il campione di $n = 3$ osservazioni $\mathbf{x}_3 = (-1, -1, 0)$, calcolare la stima di massima verosimiglianza di $\theta$ e verificare quale, tra i valori $\theta_0 = 0.5$ e $\theta_1 = 0.6$ risulta più verosimile.

e) Calcolare la stima di massima verosimiglianza di $h(\theta) = \frac{1}{2}\sqrt{\theta}$.

**Esercizio 18.** Sia $X_1, \dots, X_n$ un campione casuale da una popolazione di Poisson di parametro $\theta$. Per un generico campione osservato $\mathbf{x}_n$, si determini

a) la funzione di verosimiglianza relativa, $\bar{L}(\theta; \mathbf{x}_n)$;

b) l'informazione osservata di Fisher, $I_n^{oss}$;

c) l'approssimazione normale di $\bar{L}(\theta; \mathbf{x}_n)$;

d) l'insieme di verosimiglianza di livello $q$, utilizzando l'approssimazione normale di cui al punto precedente.

**Esercizio 19*.** Si consideri un campione casuale $X_1, \dots, X_n$, dove $X_i$ indica il tempo che un impiegato di banca dedica a ciascun cliente, e si supponga che $X_i$ abbia distribuzione normale di parametri $\mu$ e $\sigma^2$, entrambi incogniti. Si consideri quindi un campione osservato di dimensione $n = 16$, per il quale si ha $\sum_{i=1}^{n} x_i = 49.6$ e $\sum_{i=1}^{n} (x_i - \bar{x}_n)^2 = 2.56$.

a) Verificare che le stime di massima verosimiglianza dei parametri incogniti sono pari a $\hat{\mu} = 3.10$ minuti e $\hat{\sigma} = 0.40$ minuti.

b) Utilizzando le stime riportate al punto a), determinare la probabilità che il tempo dedicato dall'impiegato a un singolo cliente sia superiore a 3 minuti.

c) Utilizzando le stime riportate al punto a), determinare la probabilità che il tempo complessivo dedicato a 10 clienti sia inferiore a 35 minuti.

**Esercizio 20*.** Si consideri un campione casuale di $n$ osservazioni da una popolazione $X$ con distribuzione di probabilità

$$f_X(x;\theta) = \binom{2}{x}\theta^x(1-\theta)^{2-x} \qquad x = \{0,1,2\}, \qquad 0 < \theta < 1.$$

a) Verificare che
$$E_\theta[X] = 2\theta \qquad V_\theta[X] = 2\theta(1-\theta).$$

b) Determinare $E_\theta[\bar{X}_n]$ e $V_\theta[\bar{X}_n]$.

c) Determinare l'espressione della funzione di verosimiglianza del parametro $\theta$, $L(\theta;\mathbf{x}_n)$, associata a un generico campione osservato, $\mathbf{x}_n$, individuare il nucleo della funzione di verosimiglianza e, se esiste, una statistica sufficiente unidimensionale.

**Esercizio 21*.** Con riferimento al precedente esercizio, si consideri un campione osservato di $n = 20$ osservazioni, tale che $x_{(1)} = x_{(2)} = \ldots = x_{(14)} = 0$, $x_{(15)} = x_{(16)} = 1$, $x_{(17)} = \ldots = x_{(20)} = 2$.

a) Determinare la stima di massima verosimiglianza di $\theta$.

b) Determinare il valore dell'informazione osservata, $I_n^{oss}$, e l'espressione dell'approssimazione normale, $\bar{L}_N(\theta;\mathbf{x}_n)$, per la funzione di verosimiglianza relativa.

c) Determinare l'insieme di verosimiglianza di livello $q = 0.85$, utilizzando l'approssimazione normale ottenuta al punto precedente.

d) Sulla base dei precedenti risultati, come è possibile stimare la quantità $E_\theta[\bar{X}_n]$?

**Esercizio 22*.** Si suppone che la durata di funzionamento ($X$, in decine ore) di una popolazione di macchinari prodotti da una fabbrica sia una v.a. di Weibull con funzione di densità

$$f_X(x;\theta) = 2\,\theta\,x\,e^{-\theta x^2}, \qquad x \geq 0, \qquad \theta > 0.$$

In un esperimento si è riscontrato che, per $n = 100$ pezzi esaminati, la somma del quadrato dei tempi di durata è pari a 25.5 (decine di ore). Sulla base del campione osservato,

a) Determinare la stima di massima verosimiglianza di $\theta$.

b) Determinare una stima per intervallo per $\theta$, utilizzando l'insieme di verosimiglianza approssimato di livello $q = 0.147$.

**Esercizio 23*.** Sia $X_1, \ldots X_n$ un campione casuale proveniente da una popolazione con distribuzione uniforme nell'intervallo $(\theta, 2\theta)$, con parametro incognito $\theta$ e funzione di densità

$$f_X(x;\theta) = \frac{1}{\theta} I_{(\theta,2\theta)}(x), \qquad \theta > 0.$$

a) Verificare che
$$E_\theta[X] = \frac{3}{2}\theta, \qquad V_\theta[X] = \frac{\theta^2}{12}.$$

b) Verificare (giustificando tutte le affermazioni) che l'approssimazione normale della distribuzione campionaria della v.a. media campionaria, $\bar{X}_n$, risulta essere $N(\frac{3}{2}\theta, \frac{\theta^2}{12n})$.

c) Sulla base del precedente punto, determinare la probabilità che la v.a. $\bar{X}_n$ assuma valori superiori a $\frac{19}{6}$, ponendo $\theta = 2$ e $n = 48$.

d) Verificare che, per il modello considerato, una statistica sufficiente è rappresentata dal vettore $\big(x_{(1)}, x_{(n)}\big)$, ovvero dalla coppia costituita dal minimo e massimo dei valori campionari.

**Esercizio 24*.** Sia $X_1, \ldots, X_n$ un campione casuale proveniente da una popolazione geometrica [2] di parametro incognito $\theta$, la cui funzione di massa di probabilità è:

$$f_X(x;\theta) = \theta(1-\theta)^{x-1}, \qquad x = 1, 2, \ldots; \qquad \theta \in (0,1).$$

a) Determinare la funzione di verosimiglianza di $\theta$ associata a un generico campione osservato e una statistica sufficiente per il modello.

b) Determinare la stima di massima verosimiglianza di $\theta$.

c) Verificare che la famiglia delle distribuzioni geometriche costituisce una famiglia esponenziale.

d) In un campione di $n = 5$ osservazioni, si è rilevato che $x_1 = 3,\ x_2 = 5,\ x_3 = 1,\ x_4 = 2,\ x_5 = 4$. Utilizzando la stima di massima verosimiglianza di $\theta$, determinare la probabilità che la v.a. $X$ assuma valori maggiori di 3.

**Esercizio 25*.** Si consideri un campione casuale di $n$ osservazioni da una popolazione $X$ con distribuzione di probabilità:

$$f_X(x;\theta) = \frac{1}{\theta^2}\, x\, \exp\left\{-\frac{x}{\theta}\right\}, \qquad \theta > 0, \qquad x > 0.$$

a) Scrivere il modello statistico probabilistico e verificare che $f_X(x;\theta)$ appartiene alla famiglia esponenziale.

b) Determinare la funzione di verosimiglianza $L_{\mathbf{x}_{oss}}(\theta)$.

c) Ottenere, se esiste, una statistica sufficiente per $\theta$.

d) Determinare lo stimatore di massima verosimiglianza per $\theta$.

e) Determinare l'espressione dell'informazione osservata di Fisher.

**Esercizio 26*.** Sia

$$\mathbf{x}_{oss} = (0.56, 0.47, 0.30, 0.60, 0.22, 0.41, 0.76, 0.38, 0.08, 0.29, 0.57, 0.97, 0.81, 0.87, 0.36, \\ 0.20, 1.27, 0.20, 1.38, 1.12, 0.46, 0.52, 1.17, 0.32, 0.21, 0.61, 0.61, 1.47, 0.64, 0.08)$$

Un campione di numerosità $n = 30$ estratto da una popolazione distribuita come la variabile aleatoria $X$ dell'esercizio precedente. Si ha che $\sum_{i=1}^n x_i = 17.91$, $(\sum_{i=1}^n x_i)^2 = 320.77$, $(\sum_{i=1}^n x_i)^3 = 5744.96$.
Si determini

a) Il valore della stima di massima verosimiglianza $\hat{\theta}_{mv}$.

b) Il valore dell'informazione osservata di Fisher.

c) L'intervallo di verosimiglianza approssimato di livello $q = 0.146$.

[2] La v.a. aleatoria geometrica viene utilizzata per descrivere il numero (aleatorio) minimo di prove, ciascuna con possibile esito di tipo *successo* o *insuccesso*, da effettuare per osservare un successo. Si noti infatti che la v.a. geometrica assume valori nei numeri naturali e che il valore minimo che può assumere è pari a 1.

d) Il valore approssimato della seguente probabilità, utilizzando il valore calcolato di $\hat{\theta}_{mv}$ e tenendo conto che la variabile aleatoria $X$ ha media $\mathbb{E}[X] = 2\theta$ e varianza $\mathbb{V}[X] = 2\theta^2$,

$$\mathbb{P}\left(\sum_{i=1}^{n} X_i < 10\right).$$

**Esercizio 27*.** Si consideri un campione casuale di $n$ osservazioni da una popolazione $X$ con distribuzione di probabilità:

$$f_X(x;\theta) = \frac{3\,x^2}{\theta^3}\ \exp\left\{-\frac{x^3}{\theta^3}\right\} \quad \text{con} \quad x > 0, \quad \theta > 0.$$

1. Scrivere il modello statistico probabilistico per il campione casuale $\mathbf{X} = (X_1, \dots X_n)$ e verificare se $f_X(x;\theta)$ appartiene alla famiglia esponenziale.

2. Scrivere la funzione di verosimiglianza $L_{\mathbf{x}_{oss}}(\theta)$ ed ottenere una statistica sufficiente per $\theta$. Spiegare se la statistica sufficiente ottenuta è anche minimale.

3. Determinare la stima di massima verosimiglianza $\hat{\theta}_{MLE}$ per $\theta$ e l'informazione osservata di Fisher $\mathcal{I}(\hat{\theta}_{MLE})$. Calcolare il loro valore per il campione osservato $\mathbf{x}_{oss} = (4, 6, 7)$.

**Esercizio 28*.** Si consideri una variabile aleatoria $X$ con distribuzione di probabilità

$$f_X(x;\theta) = 2\,\theta\,e^{-2\,\theta x} \qquad x > 0, \qquad \theta > 0$$

1. Scrivere il modello statistico probabilistico e verificare che $f_X(x;\theta)$ appartiene alla famiglia esponenziale.

2. Dato un campione casuale di $n$ osservazioni proveniente da $f_X(x;\theta)$, determinare gli stimatori di massima verosimiglianza per $\theta$ e per $\gamma = 2/\theta$. Ottenere le stime di massima verosimiglianza per $\theta$ e per $\gamma$ sapendo che in un campione osservato di dimensione $n = 150$ si è ottenuto $\sum_{i=1}^{n} x_i = 1500$.

3. Determinare l'espressione dell'informazione osservata di Fisher ed il valore ottenuto in corrispondenza del campione osservato al punto 3.

**Esercizio 29*.** Si consideri un campione casuale di $n$ osservazioni da una popolazione $X$ con legge di probabilità:

$$f_X(x;\theta) = \mathbb{P}\big(X = x;\theta\big) = (x+1)(1-\theta)^2\,\theta^{\,x} \qquad 0 \le \theta \le 1 \qquad x = 0, 1, \dots$$

1. Per un generico campione osservato $\mathbf{x}_{oss}$ si ottenga la stima di massima verosimiglianza $\hat{\theta}_{MLE}$ per il parametro $\theta$, verificando che il valore che azzera la derivata prima è effettivamente un massimo.

2. Si determini il valore di $\hat{\theta}_{MLE}$ quando $\mathbf{x}_{oss} = (2, 5, 8, 10, 0, 20, 2, 1)$.

3. Dato il campione osservato al punto 2. stabilire quale valore è preferibile tra $\theta_1 = \dfrac{7}{10}$ e $\theta_2 = \dfrac{8}{10}$.

4. Assumendo che $\theta = \hat{\theta}_{MLE}$, si calcoli la probabiltà dei seguenti eventi: $X = 0$, $X = 1$. Si calcoli infine la probabilità di osservare il campione di 3 elementi $\mathbf{X}_3 = (0, 1, 0)$.

**Esercizio 30*.** Si consideri un campione casuale di $n$ osservazioni da una popolazione $X$ con legge di probabilità:

$$f_X(x;\theta) = \mathbb{P}(X = x;\theta) = \frac{1}{2}(x+1)(x+2)\,\theta^3\,(1-\theta)^{\,x} \quad 0 \leq \theta \leq 1 \qquad x = 0, 1, \dots$$

1. Per un generico campione osservato $\mathbf{x}_{oss}$ si ottenga la stima di massima verosimiglianza $\hat{\theta}_{MLE}$ per il parametro $\theta$.
2. Si determini il valore di $\hat{\theta}_{MLE}$ quando $\mathbf{x}_{oss} = (3, 1, 0, 1, 2, 2)$.
3. Dato il campione osservato al punto 2. stabilire quale valore è preferibile tra $\theta_1 = \dfrac{23}{30}$ e $\theta_2 = \dfrac{17}{30}$.
4. Assumendo che $\theta = \hat{\theta}_{MLE}$, si calcoli la probabiltà dei seguenti eventi: $X = 0$, $X = 1$. Si calcoli infine la probabilità di osservare il campione di 3 elementi $\mathbf{x} = (0, 1, 0)$.

**Esercizio 31*.** Si consideri un campione casuale proveniente da una popolazione con legge di probabilità:

$$f_X(x;\theta) = \theta(1+x)^{-(1+\theta)}; \qquad x > 0, \qquad \theta > 0,$$

1. Scrivere il modello statistico per il campione casuale $\mathbf{X} = (X_1, \dots, X_n)$.
2. Verificare se il modello appartiene alla famiglia esponenziale ed identificare, se esiste, una statistica sufficiente.
3. Determinare la funzione di verosimiglianza, il suo nucleo e l'espressione della stima di massima verosimiglianza $\hat{\theta}_{MLE}$ per un generico campione osservato $\mathbf{x}_{oss}$.
4. Ottenere l'espressione dell'informazione osservata di Fisher $\mathcal{I}_{\mathbf{x}_{oss}}(\hat{\theta}_{MLE})$.
5. In corrispondenza ad un campione osservato di numerosità $n = 50$ e $\sum \log(1 + x_i) = 106.6$ calcolare il valore di $\hat{\theta}_{MLE}$, di $\mathcal{I}_{\mathbf{x}_{oss}}(\hat{\theta}_{MLE})$ e scrivere l'approssimazione normale per la funzione di verosimiglianza.

**Esercizio 32*.** Si consideri un campione casuale proveniente da una popolazione con legge di probabilità:

$$f_X(x;\theta) = \frac{2\,x}{\theta^2}\; e^{-\dfrac{x^2}{\theta^2}} \qquad\qquad x > 0, \quad \theta > 0$$

1. Scrivere il modello statistico per il campione casuale $\mathbf{X} = (X_1, \dots, X_n)$.
2. Verificare se il modello appartiene alla famiglia esponenziale ed identificare, se esiste, una statistica sufficiente.
3. Determinare la funzione di verosimiglianza, il suo nucleo e l'espressione della stima di massima verosimiglianza $\hat{\theta}_{MLE}$ per un generico campione osservato $\mathbf{x}_{oss}$.
4. Ottenere l'espressione dell'informazione osservata di Fisher $\mathcal{I}_{\mathbf{x}_{oss}}(\hat{\theta}_{MLE})$.
5. In corrispondenza ad un campione osservato di numerosità $n = 100$ e $\sqrt{\sum_{i=1}^{n} x_i^2} = 29.01$ calcolare il valore di $\hat{\theta}_{MLE}$, di $\mathcal{I}_{\mathbf{x}_{oss}}(\hat{\theta}_{MLE})$ e scrivere l'approssimazione normale per la funzione di verosimiglianza.

## SOLUZIONI

### Variabili aleatorie normali e binomiali

**Esercizio 1.** $X \sim N(4.3, 0.2^2)$
(a) $Pr\{X > 4.5\} = Pr\{Z > \frac{4.5-4.3}{0.2}\} = 1 - F_Z(1) = 0.16$
(b) $Pr\{X < 4\} = \cdots = F_Z(-1.5) = 1 - F_Z(1.5) = 0.067$
**Esercizio 2.** $S = \sum_{i=1}^{10} X_i \sim \text{Binom}(10, \theta)$
a) $\mathbb{P}(S \geq 6) = \sum_{i=6}^{10} \binom{10}{i} \theta^i (1-\theta)^{10-i}$.
b) $\mathbb{P}(S = 10) = \theta^{10}$.
c) $\mathbb{P}(S = 0) = (1-\theta)^{10}$. Inoltre: $\theta^{10} > (1-\theta)^{10} \Leftrightarrow \theta > 1/2$.
**Esercizio 3.** $p_A = 0.8$; $p_B = 0.7$.
a) $S = \sum X_i \sim \text{Binom}(n = 16, \theta = 0.8)$; $\mathbb{P}(S = 10) = 0.8^{16} = 0.028$.
b) $\mathbb{P}(S \geq 14) = \sum_{i=14}^{16} \binom{16}{i} (0.8)^i (0.2)^{16-i}$.
c) $E(S) = np_A = 12.8$.

**Esercizio 4.** Se $X \sim$ Bernoulli(0.7); il numero di pazienti che guariscono è $S = \sum_{i=1}^{10} X_i \sim \text{Binom}(n = 10, 0.7)$.
(a) $Pr\{S > 8\} = \mathbb{P}\{S = 9\} + \mathbb{P}\{S = 10\} \simeq 0.121 + 0.028 = 0.149$.
(b) $E(S) = 10 \cdot 0.7 = 7$.

**Esercizio 5.** Numero di laureati assunti $S \sim \text{Binom}(n = 9, 0.65)$
(a) $E(S) = n\theta = 9 \cdot 0.65 = 5.85$
(b) $Pr\{S \geq 5\} = \sum_{r=5}^{9} \binom{9}{r} (0.65)^r (0.35)^{9-r} \simeq 0.83$.
(c) $Pr\{S \leq 2\} = \cdots \simeq 0.011$.

**Esercizio 6.**
$Pr\{X < 245\} = Pr\{\frac{X-\mu}{\sigma} < \frac{245-\mu}{\sigma}\} = F_Z(\frac{245-\mu}{\sigma}) = 0.33 = F_Z(-0.44)$ quindi $\frac{245-\mu}{\sigma} = -0.44$
$Pr\{X > 260\} = Pr\{\frac{X-\mu}{\sigma} > \frac{260-\mu}{\sigma}\} = 1 - F_Z(\frac{260-\mu}{\sigma}) = 0.48$
quindi $F_Z(\frac{260-\mu}{\sigma}) = 0.52 = F_Z(0.06)$ e $\frac{260-\mu}{\sigma} = 0.06$.
Risolvendo otteniamo $\mu = 258.5$ e $\sigma = 30.5$.

**Esercizio 7.** Il numero di mancini nel campione ha distribuzione binomiale di parametri $n = 120, P = 0.0235$. Le probabilità richieste sono 0.23, 0.54, 0.69.

**Esercizio 8.** Il numero di cittadini nel campione che ha assicurazione sanitaria ha distribuzione binomiale di parametri $n = 5, P = 0.7$. La probabilità richiesta è 0.163.

**Esercizio 9.** È esattamente la stessa dell'esercizio precedente.

**Esercizio 10.** $X_i =$ etnia cittadino $i$-esimo campionario (1=ispanico, 0=non ispanico);

$Y_n = \sum_{i=1}^{n} X_i =$ numero di cittadini ispanici nel campione;

$\theta = 0.0996$

$X_i|\theta \sim \text{Ber}(\theta), \qquad Y_n|\theta \sim \text{Binom}(n, \theta)$.

a) $E_\theta[Y_n] = n\theta = (1200) * (0.096) = 119.52$.

b) Utilizzando l'approssimazione normale della distribuzione binomiale e ricordando che $V_\theta(Y_n) = n\theta(1-\theta)$ si ottiene

$$\mathbb{P}(Y_n < 100) \approx \mathbb{P}(Z < (100 - 119.52)/\sqrt{107.64}) = \Phi(-1.88) = 0.03,$$

dove $\Phi(\cdot)$ indica la funzione di ripartizione della v.a. normale standardizzata $(Z)$.

**Esercizio 1.**

$$f_{X_1,\cdots,X_5}(x_1,\cdots,x_5)\theta) = \frac{\theta^{\sum_{i=1}^5 x_i} e^{-5\theta}}{\prod_{i=1}^5 x_i!} \quad \text{quindi} \quad f_{X_1,\cdots,X_5}(2,3,1,5,5;\theta) = \frac{\theta^{16} e^{-5\theta}}{172800}$$

per $\theta = 2$ la probabilità è $\frac{e^{-10}2^{16}}{172800} \simeq 0.000017$ mentre per $\theta = 3$ si ha $\frac{e^{-15}3^{16}}{172800} \simeq 0.000076$.

**Esercizio 2.**
$f_{X_1,\cdots,X_5}(x_1,\cdots,x_5;\theta) = \theta^5 e^{-\theta\sum_{i=1}^5 x_i}$ quindi $f_{X_1,\cdots,X_5}(1.12, 0.88, 0.13, 0.42, 0.36;\theta) = \theta^5 e^{-2.91\theta}$.
Per $\theta = 0.5$ vale 0.0073 mentre per $\theta = 1$ vale 0.054.

**Esercizio 3.** Se indichiamo con $W = \frac{(n-1)S^2}{\sigma^2}$, sappiamo che $W \sim \chi^2_{n-1}$. Poiché $n = 5$, $W = \frac{4S^2}{\sigma^2} \sim \chi^2_4$.
$Pr\{\frac{S^2}{\sigma^2} \leq 1.8\} = Pr\{\frac{4S^2}{\sigma^2} \leq 7.2\} = F_{\chi^2_4}(7.2) \simeq 0.87$.

**Esercizio 4.** Usiamo l'approssimazione normale alla distribuzione della media campionaria $\bar{X} \sim N(100, \frac{20^2}{16})$.
$Pr\{\bar{X} < 104\} = Pr\{Z < \frac{104-100}{20/4}\} = F_Z(0.8) \simeq 0.79$.

**Esercizio 5.** La popolazione ha distribuzione incognita, ma con valore atteso $\mu = 5.75$ e varianza $\sigma^2 = 0.48^2$. Per $n$ sufficientemente grande, $\bar{X} \sim N(\mu, \sigma^2/n)$, quindi la somma $\sum X_i = n\bar{X} \simeq N(n\mu, n\sigma^2)$. Nel nostro caso $\sum X_i \simeq N(207, 8.2944)$.
$Pr\{202 \leq \sum X_i \leq 210\} = Pr\{\frac{202-207}{2.88} \leq Z \leq \frac{210-207}{2.88}\} = F_Z(1.04) - F_Z(-1.74) \simeq 0.809$.

**Esercizio 6.** $E(\bar{X}) = E(X) = 1.3$ e $V(\bar{X}) = \frac{V(X)}{n} = \frac{0.61}{2} = 0.305$.
La tabella di sinistra contiene tutti i campioni possibili, con la probabilità e la media campionaria corrispondente, la tabella di destra riassume la distribuzione della media campionaria.

| $(x_1, x_2)$ | $Pr$ | $\bar{x}$ |
|---|---|---|
| (0,0) | 0.04 | 0 |
| (0,1) | 0.06 | 0.5 |
| (0,2) | 0.10 | 1 |
| (1,0) | 0.06 | 0.5 |
| (1,1) | 0.09 | 1 |
| (1,2) | 0.15 | 1.5 |
| (2,0) | 0.10 | 1 |
| (2,1) | 0.15 | 1.5 |
| (2,2) | 0.25 | 2 |

| $\bar{x}$ | $Pr$ |
|---|---|
| 0 | 0.04 |
| 0.5 | 0.12 |
| 1 | 0.29 |
| 0.5 | 0.30 |
| 2 | 0.25 |

Dalla seconda tabella si desumono i valori di $E(\bar{X})$ e $V(\bar{X})$.

**Esercizio 7.**
Usando l'approssimazione normale si può supporre $\sum X_i \sim N(1200, 1200)$.
$Pr\{1150 \leq \sum X_i \leq 1280\} = Pr\{\frac{1150-1200}{\sqrt{1200}} \leq Z \leq \frac{1280-1200}{\sqrt{1200}}\} = Pr\{-1.44 \leq Z \leq 2.31\} \simeq 0.91$.

**Esercizio 8.**
Anche qui usando l'approssimazione normale si può supporre $\sum X_i \sim N(1200, 1200)$.
$Pr\{1180 \leq \sum X_i \leq 1270\} = Pr\{\frac{1180-1200}{\sqrt{1200}} \leq Z \leq \frac{1270-1200}{\sqrt{1200}}\} = Pr\{-0.58 \leq Z \leq 2.02\} \simeq 0.70$.

**Esercizio 9.**
$E(T_1) = E(X) = 8 \qquad V(T_1) = \frac{V(X)}{n} = \frac{8}{4} = 2$.
$E(T_2) = V(X) = 8 \qquad V(T_2) = V(\frac{\sigma^2}{n-1} \cdot \frac{\sum(X_i-\bar{X})^2}{\sigma^2}) = \frac{\sigma^4}{(n-1)^2} \cdot 2(n-1) = 2\frac{\sigma^4}{n-1} = \frac{128}{3}$.
$E(T_3) = E(\frac{1}{4}T_1 + \frac{3}{4}T_2) = \frac{1}{4}E(T_1) + \frac{3}{4}E(T_2) = 8$
Per l'indipendenza tra $\bar{X}$ e $S^2$ si ha $V(T_3) = V(\frac{1}{4}T_1 + \frac{3}{4}T_2) = \frac{1}{16}V(T_1) + \frac{9}{16}V(T_2) = \frac{193}{8}$.

**Esercizio 10.** $E(\frac{1}{2}\bar{X}) = \frac{\theta}{2} \quad V(\frac{1}{2}\bar{X}) = \frac{\theta^2}{8n}$
$E(2X_1 + 2X_n - 3) = 4\theta - 3 \quad V(2X_1 + 2X_n - 3) = 4\theta^2$.
$\bar{X} \sim N(\theta, \frac{\theta^2}{2n})$ quindi, nel caso $\theta = 2$, $n = 25$, $\bar{X} \sim N(2, \frac{4}{50})$.
$Pr\{2 < \bar{X} < 2.4\} = Pr\{0 < Z < 1.41\} \simeq 0.42$.

**Esercizio 11.** $E(\bar{X}) = \frac{\theta}{\theta+1} \quad V(\bar{X}) = \frac{\theta}{n\theta^2(\theta+2)}$
$E(2\sum X_i - 3) = 2n\frac{\theta}{\theta+1} - 3 \quad V(2\sum X_i - 3) = 4n\frac{\theta}{\theta^2(\theta+2)}$.
$\bar{X} \sim N(\frac{\theta}{\theta+1}, \frac{\theta}{n\theta^2(\theta+2)})$ quindi, nel caso $\theta = 0.2$, $n = 25$, $\bar{X} \sim N(0.16, 0.09)$.
$Pr\{\frac{1}{2} < \bar{X} < 1\} = \cdots \simeq 0.13$.

**Esercizio 12.**

a) La distribuzione esatta di $Y_n = \sum_{i=1}^{n} X_i$ è la distribuzione di Poisson di parametro $n\theta$. Quindi $Y_n$ può assumere i valori $y = 0, 1, \ldots, y \ldots$ con probabilità

$$f_{Y_n}(y;\theta) = \mathbb{P}(Y_n = y;\theta) = \frac{e^{-n\theta}(n\theta)^y}{y!}.$$

La variabile aleatoria $U_n = 2Y_n - 3$ assume di conseguenza i valori $u = 0, -3, -1, 1, 3, 5, \ldots 2y - 3, \ldots$ con probabilità

$$f_{U_n}(u;\theta) = \mathbb{P}(U_n = u;\theta) = \mathbb{P}(2Y_n - 3 = u;\theta) =$$

$$\mathbb{P}(Y_n = \frac{u+3}{2};\theta) = f_{Y_n}((u+3)/2;\theta) = \frac{e^{-n\theta}(n\theta)^{(u+3)/2}}{[(u+3)/2]!}.$$

b) Sono verificate le ipotesi del teorema del limite centrale: le v.a. $X_1, \ldots, X_n$ sono i.i.d., e valore atteso e varianza di $X_i$ sono finiti. Pertanto,

$$Y_n = \sum_{i=1}^{n} X_i \mathrel{\dot\sim} N(n\theta, n\theta) = N(2, 2).$$

Per le proprietà delle v.a. normali, una combinazione lineare di una v.a. normale ha ancora distribuzione normale. Poichè

$$E\theta[U_n] = 2E\theta[Y_n] - 3 = 2n\theta - 3, \qquad \mathbb{V}_\theta[U_n] = 4\mathbb{V}_\theta[Y_n] = 4n\theta,$$

si ha che

$$U_n = 2Y_n - 3 \mathrel{\dot\sim} N(2n\theta - 3, 4n\theta) = N(1, 8),$$

e quindi che

$$\mathbb{P}(U_n > 0) \approx \mathbb{P}\left(\frac{U_n - 1}{\sqrt{8}} > \frac{-1}{\sqrt{8}}\right) = \mathbb{P}\left(Z > -0.36\right) = 1 - \Phi(-0.36) = \Phi(0.36) = 0.64.$$

NOTA BENE: NON CONFONDERE L' APPROSSIMAZIONE NORMALE DI UNA DISTRIBUZIONE CAMPIONARIA CON L'APPROSSIMAZIONE NORMALE DELLA F.NE DI VEROSIMIGLIANZA.

## Verosimiglianza e sufficienza

**Esercizio 1.** Per il modello normale considerato, la distribuzione campionaria è

$$f_{X_1,\cdots,X_n}(x_1, \cdots, x_n;\theta) = \left(\frac{1}{\sqrt{2\pi\theta}}\right)^n e^{-\frac{1}{2\theta}\sum_i x_i^2}$$

quindi per il campione osservato $\mathbf{x}_n = (x_1, \cdots, x_n) = (2.52, \cdots, -0.27)$ abbiamo

$$f_n(\mathbf{x}_n; \theta) = L(\theta; \mathbf{x}_n) = \theta^{-5} e^{-\frac{34.83}{2\theta}}$$

La log-verosimiglianza è $\log L(\theta; \mathbf{x}_n) = -5 \log \theta - \frac{34.83}{2\theta}$. Se deriviamo la log-verosimiglianza e uguagliamo a zero otteniamo come punto di massimo $\hat{\theta} = \frac{34.83}{10} = 3.483$.

Ponendo $\theta = 3$ si ha $L(3, \mathbf{x}_n) = 0.000012$ mentre per $\theta = 5$ si ha $L(5, \mathbf{x}_n) = 0.0000098$. Per confrontarli vediamo che

$$\frac{L(3, \mathbf{x}_n)}{L(5, \mathbf{x}_n)} = 1.26$$

quindi, alla luce del nostro campione $\mathbf{x}_n$, il valore $\theta = 3$ è più verosimile di $\theta = 5$.

**Esercizio 2.** Si ha che:

$$\begin{aligned}
L(\theta; \mathbf{x}_n) &= \frac{\theta^n e^{n\theta}}{(\prod_i x_i)^{\theta+1}} \\
\log L(\theta; \mathbf{x}_n) &= n \log \theta + n\theta - (\theta + 1) \log(\prod_i x_i) \\
\frac{\partial}{\partial \theta} \log L(\theta; \mathbf{x}_n) &= \frac{n}{\theta} + n - \log(\prod_i x_i) \\
\hat{\theta} &= \frac{n}{\log(\prod_i x_i) - n} = \frac{n}{\sum_i \log x_i - n}
\end{aligned}$$

Sostituendo i valori campionari si ottiene la stima di massima verosimilianza.

**Esercizio 3.** Qui il supporto della v.a. $X$ da $\theta$, quindi è bene utilizzare le funzioni indicatrici.

$$\begin{aligned}
f_n(\mathbf{x}_n; \theta) &= e^{-(\sum_i x_i - n\theta)} \prod_i I_{(\theta, +\infty)}(x_i) \\
L(\theta; \mathbf{x}_n) &= e^{n\theta} I_{(0, x_{(1)})}(\theta) \quad \text{è positiva solo per } \theta \in (0, x_{(1)})
\end{aligned}$$

che risulta crescente in $(0, x_{(1)})$. Il massimo si raggiunge quindi nel valore più elevato che $\theta$ può assumere, ossia $x_{(1)}$.

**Esercizio 4.**

$$\begin{aligned}
f_n(\mathbf{x}_n; \theta) &= \frac{\theta^{\sum_i x_i} e^{-n\theta}}{\prod_i x_i!} \\
L(\theta; \mathbf{x}_n) &= \theta^{\sum_i x_i} e^{-n\theta} \\
L(\theta; \mathbf{x}_n) &= \theta^{28} e^{-10\theta} \quad \text{per questo campione}
\end{aligned}$$

quindi $L(2, \mathbf{x}_n) = 0.55$, $L(3, \mathbf{x}_n) = 2.14$, $L(4, \mathbf{x}_n) = 0.30$.

**Esercizio 5.**

$$\begin{aligned}
f_n(\mathbf{x}_n; \theta) &= 2^n \theta^n (\prod_i x_i) e^{-\theta \sum_i x_i^2} \\
L(\theta; \mathbf{x}_n) &= \theta^n e^{-\theta \sum_i x_i^2} \\
\log L(\theta; \mathbf{x}_n) &= n \log \theta - \theta \sum_i x_i^2 \\
\frac{\partial}{\partial \theta} \log L(\theta; x) &= \frac{n}{\theta} - \sum x_i^2 \\
\hat{\theta} &= \frac{n}{\sum x_i^2}
\end{aligned}$$

**Esercizio 6.**

$$\begin{aligned}
f_n(\mathbf{x}_n;\theta) &= 2^n\theta^{-n}(\prod_i x_i)e^{-(\sum_i x_i^2)/\theta} \\
L(\theta;\mathbf{x}_n) &= \theta^{-n}e^{-(\sum_i x_i^2)/\theta} \\
\log L(\theta;\mathbf{x}_n) &= -n\log\theta - \frac{\sum_i x_i^2}{\theta} \\
\frac{\partial}{\partial\theta}\log L(\theta;\mathbf{x}_n) &= -\frac{n}{\theta} + \frac{\sum_i x_i^2}{\theta^2} \\
\hat\theta &= \frac{\sum_i x_i^2}{n}
\end{aligned}$$

**Esercizio 7.**

$$\begin{aligned}
f_n(\mathbf{x}_n;\theta) &= \left(\frac{4}{\sqrt{\pi}}\right)^n\theta^{-3n}(\prod_i x_i^2)e^{-(\sum_i x_i^2)/\theta^2} \\
L(\theta;\mathbf{x}_n) &= \theta^{-3n}e^{-(\sum_i x_i^2)/\theta^2} \qquad \sum_i x_i^2 \text{ stat. suff.} \\
\log L(\theta;\mathbf{x}_n) &= -3n\log\theta - \frac{\sum x_i^2}{\theta^2} \\
\frac{\partial}{\partial\theta}\log L(\theta;\mathbf{x}_n) &= -\frac{3n}{\theta} + \frac{2\sum_i x_i^2}{\theta^3} \\
\hat\theta &= \sqrt{\frac{2}{3}\frac{\sum_i x_i^2}{n}}
\end{aligned}$$

**Esercizio 8.**

$$\begin{aligned}
f_n(\mathbf{x}_n;\theta) &= \frac{\theta^{\sum_i x_i}e^{-n\theta}}{\prod_i x_i!} \\
L(\theta;\mathbf{x}_n) &= \theta^{\sum_i x_i}e^{-n\theta} \\
\log L(\theta;\mathbf{x}_n) &= \sum_i x_i\log\theta - n\theta \\
\frac{\partial}{\partial\theta}\log L(\theta;\mathbf{x}_n) &= \frac{\sum_i x_i}{\theta} - n \\
\hat\theta &= \frac{\sum_i x_i}{n}
\end{aligned}$$

Poiché per il nostro campione $\sum x_i = 34$, $\hat\theta = 3.4$.

Chiamiamo il nuovo parametro $\alpha = P(X=0) = e^{-\theta}$. Per l'invarianza delle stime di MV abbiamo che $\hat\alpha = e^{-3.4} = 0.033$. In alternativa, arriviamo allo stesso risultato osservando che $\theta = -\log\alpha$ e quindi

$$\begin{aligned}
f_n(\mathbf{x}_n;\alpha) &= \frac{(-\log\alpha)^{\sum_i x_i}e^{n\log\alpha}}{\prod_i x_i!} \\
L(\alpha;\mathbf{x}_n) &= (-\log\alpha)^{\sum_i x_i}e^{-n\log\alpha} = (-\log\alpha)^{\sum_i x_i}\alpha^n \\
\log L(\alpha;\mathbf{x}_n) &= \sum_i x_i\log(-\log\alpha) + n\log\alpha \\
\frac{\partial}{\partial\alpha}\log L(\alpha;\mathbf{x}_n) &= \sum_i x_i\frac{1}{-\log\alpha}(-\frac{1}{\alpha}) + \frac{n}{\alpha} \\
\hat\alpha &= e^{-\frac{\sum_i x_i}{n}} = e^{-\hat\theta}
\end{aligned}$$

con il nostro campione la stima è $\hat{\alpha} = e^{-3.4} = 0.033$

**Esercizio 9.** Calcoliamo prima

$$E(X) = \int_0^\theta x \frac{2x}{\theta^2}\, dx = \cdots = \frac{2}{3}\theta.$$

Ovviamente $E(\bar{X}) = E(X) = \frac{2}{3}\theta$.

Qui il supporto di $X$ dipende da $\theta$. Conviene quindi introdurre le funzioni indicatrici.

$$\begin{aligned} f_n(\mathbf{x}_n;\theta) &= 2^n (\prod_i x_i!)\theta^{-2n} \prod_i I_{(0,\theta)}(x_i) \\ L(\theta;\mathbf{x}_n) &= \theta^{-2n} I_{(x_{(n)},+\infty)}(\theta) \qquad \text{è positiva solo per } \theta > x_{(n)}. \end{aligned}$$

La f.v. decresce in $(x_{(n)}, +\infty)$ e quindi il massimo si raggiunge nel valore minimo che $\theta$ può assumere, ossia $x_{(n)}$.

**Esercizio 10.** Qui l'insieme dei campioni dipende da $\theta$, quindi è bene lavorare con le funzioni indicatrici.

$$\begin{aligned} f_n(\mathbf{x}_n;\theta) &= 2^n \theta^{2n} (\prod_i x_i) \prod_i I_{(\theta,+\infty)}(x_i) \\ L(\theta;\mathbf{x}_n) &= \theta^{2n} I_{(0,x_{(1)})}(\theta) \quad \text{è positiva solo per } \theta \in (0, x_{(1)}) \\ \log L(\theta;\mathbf{x}_n) &= 2n \log\theta \quad \text{per } \theta \in (0, x_{(1)}) \\ \frac{\partial}{\partial\theta} \log L(\theta;\mathbf{x}_n) &= \frac{2n}{\theta} \quad \text{per } \theta \in (0, x_{(1)}) \\ \hat{\theta} &= x_{(1)} \end{aligned}$$

perché la derivata è positiva e quindi il massimo si raggiunge nel valore più alto che $\theta$ può assumere, ossia $x_{(1)}$.

**Esercizio 11.**

$$\begin{aligned} f_n(\mathbf{x}_n;\theta) &= \left( \prod_i \frac{\Gamma(\alpha + x_i)}{\Gamma(x_i+1)\Gamma(\alpha)} \right) \theta^{n\alpha} (1-\theta)^{\sum_i x_i} \\ L(\theta;\mathbf{x}_n) &= \theta^{n\alpha}(1-\theta)^{\sum x_i} \qquad \Rightarrow \qquad \sum_i x_i \text{ stat. suff.} \end{aligned}$$

**Esercizio 12.** Qui $n = 5$ e $\sum x_i = 20$, quindi

$$\begin{aligned} L(\theta;\mathbf{x}_n) &= \theta^5 (1-\theta)^{20} \\ \log L(\theta;\mathbf{x}_n) &= 5\log\theta + 20\log(1-\theta) \\ \frac{\partial}{\partial\theta} \log L(\theta;\mathbf{x}_n) &= \frac{5}{\theta} - \frac{20}{1-\theta} \\ \hat{\theta} &= \frac{1}{5} \end{aligned}$$

**Esercizio 13.** Calcoliamo prima

$$E(X) = \int_0^\theta x \frac{2}{\theta}(1 - \frac{x}{\theta})\, dx = \cdots = \frac{\theta}{3}.$$

Qui l'insieme dei campioni dipende da $\theta$, quindi è bene lavorare con le funzioni indicatrici.

$$\begin{aligned} f_n(\mathbf{x}_n;\theta) &= 2^n\theta^{-n}(\prod_i(1-\frac{x_i}{\theta})\prod_i I_{(0,\theta)}(x_i) \\ L(\theta;\mathbf{x}_n) &= \theta^{-n}(\prod_i(1-\frac{x_i}{\theta})I_{(x_{(n)},+\infty)}(\theta) \quad \text{è positiva solo per } \theta > x_{(n)}. \end{aligned}$$

**Esercizio 14.** È un esperimento bernoulliano, quindi

$$\begin{aligned} f_n(\mathbf{x}_n;\theta) &= \theta^{\sum_i x_i}(1-\theta)^{n-\sum_i x_i} \\ L(\theta;\mathbf{x}_n) &= \theta^{\sum_i x_i}(1-\theta)^{n-\sum_i x_i} \\ \log L(\theta;\mathbf{x}_n) &= \sum_i x_i\log\theta + (n-\sum_i x_i)\log(1-\theta) \\ \frac{\partial}{\partial\theta}\log L(\theta;\mathbf{x}_n) &= \frac{\sum_i x_i}{\theta} - \frac{n-\sum_i x_i}{1-\theta} \\ \hat{\theta} &= \frac{\sum_i x_i}{n}. \end{aligned}$$

Nel nostro campione $n = 371$ e $\sum x_i' = 18$, quindi $\hat{\theta} = 0.048$.

**Esercizio 15.** a) Svolgimento standard; b) $T(\mathbf{x}_n) = \sum_i[\ln x_i - 3]^2$; c) $\hat{\theta}(\mathbf{X}_n) = \sqrt{T(\mathbf{x}_n)/n}$; d) $\hat{\theta}(\mathbf{x}_n) = \sqrt{2}$.

**Esercizio 16.** a) Svolgimento standard; $L(\theta;\mathbf{x}_n) = \theta^n[\prod_i x_i]^{-(\theta+1)}$; b) segue dal criterio di fattorizzazione; c) $\hat{\theta}(\mathbf{X}_n) = n/(\sum_i \ln X_i)$; d) banale; e) $\hat{h} = \sqrt{\hat{\theta}} + 1$.

**Esercizio 17.** a) Svolgimento standard; b) $T(\mathbf{x}_n) = \sum_i |X_i|$; c) $\hat{\theta}(\mathbf{X}_n) = T(\mathbf{X}_n)/n$; d) $\hat{\theta}(\mathbf{x}_n) = 2/3$; 0.6 è più verosimile; e) $\hat{h} = \sqrt{1/6}$.

**Esercizio 18.** a) $\bar{L}(\theta,\mathbf{x}_n) = (\theta/\bar{x}_n)^{\sum_i x_i}e^{-n(\theta-\bar{x}_n)}$; b) $I_n^{oss} = n/\bar{x}_n$; c) $\bar{L}$ è approssimata da una densità normale di parametri $(\bar{x}_n, \bar{x}_n/n)$; d) $\bar{x}_n \pm k \times \sqrt{\bar{x}_n/n}$, con $k = \sqrt{-2\ln q}$.

**Esercizio 19.**

a) Sappiamo che le stime di MV sono, in questo modello,

$$\hat{\mu} = \bar{x}_n = \frac{49.6}{16} = 3.10, \qquad \hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^n (x_i - \bar{x}_n)^2 = \frac{2.56}{16} \quad \Rightarrow \quad \hat{\sigma} = 0.40.$$

b) Indicando con $Z$ la v.a. $\mathrm{N}(0,1)$ e con $\Phi(\cdot)$ la sua f. di ripartizione, si ha:

$$\mathbb{P}(X > 3) = \mathbb{P}\left(Z > \frac{3-3.1}{0.4}\right) = \mathbb{P}(Z > -0.25) = 1 - \mathbb{P}(Z < -0.25) = 1 - \Phi(-0.25) \simeq 0.6$$

c) $X_1,\ldots,X_n$ i.i.d. $\Rightarrow Y_n = \sum_{i=1}^n X_i \sim \mathrm{N}(n\mu; n\sigma^2)$ = (in questo caso) = $\mathrm{N}(31, 10(0.4)^2)$. Quindi:

$$\mathbb{P}(Y_n < 35) = \mathbb{P}\left(Z < \frac{35-31}{0.4\sqrt{10}}\right) = \mathbb{P}(Z < 3.17) = \Phi(3.17) \simeq 0.99.$$

**Esercizio 20.**

a) Si osservi che $X_i|\theta \sim \text{Bin}(2, \theta) \quad \Rightarrow \quad E_\theta[X] = 2\theta, \qquad V_\theta[X] = 2\theta(1-\theta)$.
Per il calcolo esplicito si ha:

$$E_\theta[X] = \sum_{x=0}^{2} x f_X(x;\theta) = 0 \times \binom{2}{0}\theta^0(1-\theta)^2 + 1 \times \binom{2}{1}\theta^1(1-\theta)^1 + 2 \times \binom{2}{2}\theta^2(1-\theta)^0 = ... = 2\theta,$$

e che

$$E_\theta[X^2] = \sum_{x=0}^{2} x^2 f_X(x;\theta) = ... = 2\theta^2 + 2\theta,$$

da cui

$$V_\theta[X] = E_\theta[X^2] - (E_\theta[X])^2 = 2\theta^2 + 2\theta - (2\theta)^2 = 2\theta(1-\theta).$$

b) Per le note proprietà di media e varianza campionarie per campioni casuali, si ha che

$$E_\theta[\bar{X}_n] = E_\theta[X] = 2\theta$$

e che

$$V_\theta[\bar{X}_n] = \frac{1}{n} V_\theta[X] = \frac{2}{n}\theta(1-\theta).$$

c) Si ha $L(\theta; \mathbf{x}_n) = \left[\prod_{i=1}^n \binom{2}{x_i}\right] \theta^{\sum_{i=1}^n x_i}(1-\theta)^{2n - \sum_{i=1}^n x_i}$. Pertanto:

$$L(\theta; \mathbf{x}_n) = h(\mathbf{x}_n) g(T(\mathbf{x}_n); \theta), \quad \text{dove} \quad h(\mathbf{x}_n) = \prod_{i=1}^{n} \binom{2}{x_i}, \quad g(T(\mathbf{x}_n); \theta) = \theta^{T(\mathbf{x}_n)} \theta^{2n - T(\mathbf{x}_n)}.$$

In base al criterio di fattorizzazione si ha quindi che $T(\mathbf{x}_n) = \sum_{i=1}^n x_i$ è statistica sufficiente unidimensionale.

**Esercizio 21.**

a)

$$\frac{d}{d\theta} \ln L(\theta; \mathbf{x}_n) = \frac{d}{d\theta}[\sum_{i=1}^{n} x_i \ln\theta + (2n - \sum_{i=1}^{n} x_i)\ln(1-\theta)] = \frac{\sum_{i=1}^n x_i}{\theta} - \frac{2n - \sum_{i=1}^n x_i}{1-\theta}.$$

Uguagliando a zero e risolvendo rispetto a $\theta$ si trova

$$\hat{\theta}_{smv}(\mathbf{x}_n) = \frac{\sum_{i=1}^n x_i}{2n} = 10/40 = 1/4,$$

che è punto di massimo in quanto la derivata seconda di $\ln L(\theta; \mathbf{x}_n)$ è negativa $\forall\theta$.

b) Poichè

$$= -\frac{d^2}{d\theta^2}[\frac{\sum_{i=1}^n x_i}{\theta} - \frac{2n - \sum_{i=1}^n x_i}{1-\theta}] = \frac{\sum_{i=1}^n x_i}{\theta^2} + \frac{2n - \sum_{i=1}^n x_i}{(1-\theta)^2},$$

Si ha

$$I_n^{oss} = -\frac{d^2}{d\theta^2} \ln L(\theta; \mathbf{x}_n) \mid_{\theta = \hat{\theta}_{smv}} = \frac{\sum_{i=1}^n x_i}{\hat{\theta}_{smv}^2} + \frac{2n - \sum_{i=1}^n x_i}{(1 - \hat{\theta}_{smv})^2} = \frac{10}{1/16} + \frac{40 - 10}{9/16} = \frac{640}{3}.$$

Quindi

$$\bar{L}_N(\theta; \mathbf{x}_n) = \exp\{-\frac{1}{2}(\theta - 1/4)^2(640/3)\}.$$

c) Poichè $\sqrt{-2\ln(0.85)} = 0.57$, si ha

$$\tilde{L}_q(\mathbf{x}_n) = [1/4 - 0.57\sqrt{3/640}, 1/4 + 0.57\sqrt{3/640}] = [0.211, 0.289].$$

d) $\widehat{E_\theta[X]} = \widehat{[2\theta]} = 2\hat{\theta}_{smv}(\mathbf{x}_n) = 1/2.$

**Esercizio 22.** a)

$$f_n(\mathbf{x}_n;\theta) = 2^n\theta^n\left(\prod_{i=1}^n x_i\right)e^{-\theta\sum_{i=1}^n x_i^2}; \qquad \ell(\theta;\mathbf{x}_n) = \ln L(\theta;\mathbf{x}_n) \propto n\ln\theta - \theta\sum_{i=1}^n x_i^2.$$

Pertanto

$$\frac{d}{d\theta}\ell(\theta;\mathbf{x}_n) = 0 \quad\Leftrightarrow\quad \frac{n}{\theta} - \sum_{i=1}^n x_i^2 = 0 \quad\Leftrightarrow\quad \theta = \frac{n}{\sum_{i=1}^n x_i^2}.$$

Poichè la derivata seconda di $\ell(\theta;\mathbf{x}_n)$ è $-n/\theta^2$, ovvero sempre negativa, la radice trovata è punto di massimo per $\ell$ (e per $L$) e quindi stima di massima verosimiglianza, $\hat{\theta}_{mv}(\mathbf{x}_n)$. Per il campione osservato si ha: $\hat{\theta}_{mv}(\mathbf{x}_n) = 100/25.5 = 3.92$.

b) Intervallo di verosimiglianza approssimato di livello $q$ è:

$$\hat{\theta}_{mv}(\mathbf{x}_n) \pm K \times \sqrt{1/I_n^{oss}},$$

dove

$$k = \sqrt{-2\ln q} = 1.96, \qquad I_n^{oss} = -\frac{d}{d\theta}\ln L(\theta;\mathbf{x}_n)|_{\theta=\hat{\theta}_{mv}(\mathbf{x}_n)} = \frac{n}{[\hat{\theta}_{mv}(\mathbf{x}_n)]^2} = \frac{(\sum_{i=1}^n x_i^2)^2}{n} = 6.05.$$

Si ha quindi che l'intervallo richiesto è: $(3.16, 4.68)$.

**Esercizio 23.**
a)

$$E_\theta[X] = \int_\theta^{2\theta}\frac{x}{\theta}d\theta = \ldots = \frac{3}{2}\theta.$$
$$E_\theta[X^2] = \int_\theta^{2\theta}\frac{x^2}{\theta}d\theta = \ldots = \frac{7}{3}\theta^2.$$

$V_\theta[X] = E_\theta[X^2] - (E_\theta[X])^2 = \ldots = \frac{1}{12}\theta^2.$

b)
Segue dal teorema del limite centrale (le cui ipotesi sono soddisfatte) osservando che:

$$E_\theta[\bar{X}_n] = E_\theta[X] = \frac{3}{2}\theta, \qquad V_\theta[\bar{X}_n] = \frac{V_\theta[X]}{n} = \frac{1}{12n}\theta^2.$$

c)

$$\mathbb{P}\left(\bar{X}_n > \frac{19}{6}\right) \simeq \mathbb{P}\left(Z > \frac{19/6 - 3}{1/12}\right) = \mathbb{P}(Z > 2) = 1 - \Phi(2) = 0.023.$$

d)
Poichè

$$\theta < x_1 < \ldots < x_n < 2\theta \qquad \Leftrightarrow \theta < x_{(1)} < \ldots < x_{(n)} < 2\theta,$$

si ha che $\prod_{i=1}^n I_{(\theta,2\theta)}(x_i) = I_{(\theta,2\theta)}(x_{(1)})\ I_{(\theta,2\theta)}(x_{(n)})$. Pertanto

$$L(\theta;\mathbf{x}_n) = \frac{1}{\theta^n}\prod_{i=1}^n I_{(\theta,2\theta)}(x_i) = \frac{1}{\theta^n}\ I_{(\theta,2\theta)}(x_{(1)})\ I_{(\theta,2\theta)}(x_{(n)}),$$

e risultato discende dal criterio di fattorizzazione.

**Esercizio 24.** a)

La funzione di verosimiglianza è:

$$L(\theta;\mathbf{x}_n) = \theta^n(1-\theta)^{\sum_{i=1}^n x_i - n}.$$

Per il criterio di fattorizzazione segue che $\sum_{i=1}^n x_i$ è una stat. sufficiente per il modello.

b)

Risolvendo l'equazione di log-verosimiglianza si trova che la soluzione

$$\hat{\theta}_{MV} = \frac{n}{\sum_{i=1}^n x_i},$$

che risulta essere effettivamente un punto di massimo per la funzione.

c)

Si ha che:

$$f_X(x;\theta) = \exp(\log[\theta(1-\theta)^{x-1}]) = (\text{propr. dei logaritmi}) = \exp(x\log(1-\theta) + \log\frac{\theta}{1-\theta}).$$

Il modello considerato è quindi famiglia esponenziale con

$$h(x) = 1, \qquad T(x) = x, \qquad \eta(\theta) = \log(1-\theta), \qquad c(\theta) = \log\frac{\theta}{1-\theta}.$$

d)

Per un $\theta$ generico si ha:

$$\mathbb{P}(X > 3;\theta) = 1 - \mathbb{P}(X \le 3;\theta) = 1 - \sum_{x=1}^3 f_X(x;\theta) = 1 - [\theta + \theta(1-\theta) + \theta(1-\theta)^2]|.$$

Il risultato si ottiene sostituendo nella precedente espressione a $\theta$ il valore della stima di massima verosimiglianza che, con $n=5$ e $\sum_{i=1}^n x_i = 15$, risulta essere in quest'esempio $\hat{\theta}_{MV} = 5/15 = 1/3$.

**Esercizio 25.**

a) Modello statistico:

$$\left(\mathcal{X}^n = (\mathbb{R}^+)^n,\ \ f_n(\mathbf{x};\theta) = \theta^{-2n}\ [\prod_{i=1}^n x_i I_{\mathbb{R}^+}(x_i)]\ \exp\left\{-\frac{\sum_{i=1}^n x_i}{\theta}\right\},\ \ \Theta = \mathbb{R}^+\right).$$

La distribuzione di $X$ appartiene alla famiglia esponenziale in quanto può essere scritta come:

$$f_X(x;\theta) = x\ \exp\left\{-2\log\theta - \frac{x}{\theta}\right\} I_{\mathbb{R}^+}(x) = h(x)\exp\{\eta(\theta)T(x) - B(\theta)\},$$

con:

$$h(x) = xI_{\mathbb{R}^+}(x) \qquad \eta(\theta) = -\frac{1}{\theta} \qquad T(x) = x \qquad B(\theta) = 2\log\theta.$$

b) La funzione di verosimiglianza è:

$$L(\theta; \mathbf{x}_n) = \theta^{-2n} \, e^{-\frac{\sum_{i=1}^n x_i}{\theta}}$$

c) Una statistica sufficiente è data da $T(\mathbf{x}) = \sum_{i=1}^n x_i$.

d) Per determinare lo stimatore di massima verosimiglianza si considera:

$$\ell(\theta) = \log L(\theta; \mathbf{x}_n) = -2n \log \theta - \frac{\sum_{i=1}^n x_i}{\theta}$$

che ha derivata prima data da:

$$\ell'(\theta) = -\frac{2n}{\theta} + \frac{\sum_{i=1}^n x_i}{\theta^2} \qquad \text{e} \qquad \text{quindi} \qquad \hat{\theta}_{mv} = \frac{\sum_{i=1}^n x_i}{2n}.$$

Inoltre $\ell''(\theta) = 2n/\theta^2 - 2\sum_{i=1}^n x_i/\theta^3$ che risulta essere $< 0$ in $\theta = \hat{\theta}_{mv}$.

e) L'informazione osservata di Fisher è data da:

$$I(\hat{\theta}) = -\ell''(\theta)\big|_{\theta=\hat{\theta}_{mv}} = \frac{2n}{\hat{\theta}^2_{mv}} = \frac{8n^3}{(\sum_{i=1}^n x_i)^2}.$$

**Esercizio 26.**

a) Il valore della stima di massima verosimiglianza è:

$$\hat{\theta}_{mv} = \frac{\sum_{i=1}^n x_i}{2n} = \frac{17.91}{60} = 0.2985.$$

b) Il valore dell'informazione osservata di Fisher:

$$I_n^{oss}(\mathbf{x}_n) = \frac{8n^3}{(\sum_{i=1}^n x_i)^2} = \frac{216.000}{320.77} = 673.38.$$

c) L'intervallo di verosimiglianza approssimato si ottiene sfruttando l'approssimazione normale della funzione di verosimiglianza (DA NON CONFONDERE CON L'APPROSSIMAZIONE NORMALE DI UNA DISTRIBUZIONE CAMPIONARIA) nell'intorno del punto $\theta = \hat{\theta}_{mv}$, da cui si ottiene che, ponendo $k = \sqrt{-2\log q} = 1.96$:

$$L_q = \hat{\theta}_{mv} \pm k \times [I_n^{oss}(\mathbf{x}_n)]^{-1/2} = 0.2985 \pm \frac{1.96}{\sqrt{673.38}} = 0.2985 \pm 0.0755.$$

e l'intervallo è $(0.22, 0.37)$.

d) Le ipotesi del TLC sono verificate: variabili aleatorie i.i.d. con valore atteso e varianza finiti. Dato che $\sum_{i=1}^n X_i \mathrel{\dot\sim} N(2n\theta, 2n\theta^2) =$ (per $\theta = \hat{\theta}_{mv}$)$= N(17.91, 5.34)$, si ottiene:

$$\mathbb{P}\left(\sum_{i=1}^n X_i < 10\right) \approx \mathbb{P}\left(Z < \frac{10 - 17.91}{2.32}\right) = \mathbb{P}\left(Z < -3.41)\right) = \Phi(-3.41) = \texttt{pnorm}(-3.41) = 0.00032.$$

NOTA BENE: NON CONFONDERE L' APPROSSIMAZIONE NORMALE DI UNA DISTRIBUZIONE CAMPIONARIA CON L'APPROSSIMAZIONE NORMALE DELLA F.NE DI VEROSIMIGLIANZA.

**Esercizio 27.**

**1.** Il modello statistico probabilistico per il campione casuale $\mathbf{X} = (X_1, \dots X_n)$ è costituito dalla famiglia di distribuzioni congiunte dell'intero vettore campionario $(X_1, ..., X_n)$ (scandita dallo spazio parametrico ) e dallo spazio campionario corrispondente. In simboli

$$\left\{ f(x_1, ..., x_n; \theta) = \prod_{i=1}^{n} f_X(x_i; \theta) = \frac{3^n \; \prod_{i=1}^{n} x_i^2}{\theta^{3n}} \; \exp\left\{ - \sum_{i=1}^{n} \frac{x_i^3}{\theta^3} \right\} \; ; \; \theta \in \Theta = (0, \infty) \right\}$$
$$\mathcal{X}^n = (0, \infty)^n = (0, \infty) \times ... \times (0, \infty)$$

Per verificare che $f_X(x; \theta)$ appartiene alla famiglia esponenziale è sufficiente esibire le funzioni $h(x)$, $\eta(\theta)$, $T(x)$, e $B(\theta)$ tali che

$$f_X(x; \theta) = h(x) \exp\left\{ \eta(\theta) T(x) - B(\theta) \right\}$$

Nel caso in questione si potevano prendere

$$\begin{aligned} h(x) &= 3x^2 \\ \eta(\theta) &= \frac{1}{\theta^3} \\ T(x) &= -x^3 \\ B(\theta) &= \log \theta^{-3} = 3 \log \theta \end{aligned}$$

**2.** La funzione di verosimiglianza è

$$L_{\mathbf{x}_{oss}}(\theta) = \frac{3^n \; \prod_{i=1}^{n} x_i^2}{\theta^{3n}} \; \exp\left\{ - \sum_{i=1}^{n} \frac{x_i^3}{\theta^3} \right\} \; \propto \; \frac{1}{\theta^{3n}} \; \exp\left\{ - \sum_{i=1}^{n} \frac{x_i^3}{\theta^3} \right\} \quad \theta \in (0, \infty)$$

Notare che il simbolo $\propto$ (leggasi 'è proporzionale a') indica che è stato rimosso dalla verosimiglianza un termine moltiplicativo, costante rispetto al parametro $\theta$, che non influisce nella determinazione del comportamento della verosimiglianza come funzione se non per un fattore di scala e che comunque non interviene nella determinazione dell'argomento di massimo o nelle eventuali derivate prima e seconda della logverosimiglianza. Tutto ciò che segue il simbolo $\propto$ viene denominato *nucleo funzionale* della verosimiglianza.

La statistica sufficiente si può ottenere con una delle due argomentazioni seguenti, entrambe valide.

La prima è la via più rapida ed è la seguente: avendo riconosciuto che il modello della singola osservazione appartiene alla famiglia esponenziale allora ogni funzione biunivoca di

$$\sum_{i=1}^{n} T(x_i) = - \sum_{i=1}^{n} x_i^3$$

è una statistica sufficiente (per una nota proprietà delle famiglie esponenziali) e con analoga argomentazione che sfrutta le proprietà delle famiglie esponenziali è possibile affermare che la stessa statistica è anche minimale oltre che sufficiente per il parametro $\theta$.

La seconda argomentazione discende dalla definizione di sufficienza attraverso il criterio di fattorizzazione di Neyman. In tal caso è sufficiente verificare che si ha

$$f(x_1, ..., x_n; \theta) = \prod_{i=1}^{n} f_X(x_i; \theta) = \frac{3^n \; \prod_{i=1}^{n} x_i^2}{\theta^{3n}} \; \exp\left\{ - \sum_{i=1}^{n} \frac{x_i^3}{\theta^3} \right\} = k(x_1, ..., x_n) g(S(x_1, ..., x_n), \theta)$$

per opportune funzioni $S(x_1,...,x_n)$, $k(x_1,...,x_n)$ (funzioni della sola $n$-upla campionaria) e $g(s,\theta)$. Nel caso in questione si poteva considerare

$$\begin{aligned} S(x_1,...,x_n) &= \sum_{i=1}^{n} x_i^3 \\ k(x_1,...,x_n) &= 3^n \prod_{i=1}^{n} x_i^2 \\ g(s,\theta) &= \frac{1}{\theta^{3n}} \exp\left\{-\frac{s}{\theta^3}\right\} \end{aligned}$$

Verificata la sufficienza in questo modo si doveva verificare la proprietà di minimalità della statistica sufficiente attraverso il criterio di Lehmann e Scheffé facendo vedere che il rapporto

$$\frac{L_{\mathbf{x}_{oss}}(\theta)}{L_{y_{oss}}(\theta)} = \frac{\frac{3^n \prod_{i=1}^n x_i^2}{\theta^{3n}} \exp\left\{-\sum_{i=1}^n \frac{x_i^3}{\theta^3}\right\}}{\frac{3^n \prod_{i=1}^n y_i^2}{\theta^{3n}} \exp\left\{-\sum_{i=1}^n \frac{y_i^3}{\theta^3}\right\}} = c((x_1,...,x_n),(y_1,...,y_n) \Longleftrightarrow S(x_1,...,x_n) = S(y_1,...,y_n)$$

In tal caso il rapporto si semplificava e si poteva dedurre che

$$\begin{aligned} \frac{L_{\mathbf{x}_{oss}}(\theta)}{L_{y_{oss}}(\theta)} &= \frac{\prod_{i=1}^n x_i^2}{\prod_{i=1}^n y_i^2} \exp\left\{-\sum_{i=1}^{n} \frac{\left(x_i^3 - y_i^3\right)}{\theta^3}\right\} = c((x_1,...,x_n),(y_1,...,y_n) \Longleftrightarrow \\ & \exp\left\{-\sum_{i=1}^{n} \frac{\left(x_i^3 - y_i^3\right)}{\theta^3}\right\} = c((x_1,...,x_n),(y_1,...,y_n) \Longleftrightarrow \\ & \sum_{i=1}^{n} \left(x_i^3 - y_i^3\right) = 0 \Longleftrightarrow \\ & S(x_1,...,x_n) = S(y_1,...,y_n) \end{aligned}$$

**3.** - Essendo il modello in questione un modello regolare la stima di massima verosimiglianza si ottiene dal solito procedimento analitico, considerando la logverosimiglianza e risolvendo l'equazione di verosimiglianza ovvero quella ottenuta azzerando la derivata prima della logverosimiglianza. Da ciò deriva

$$\hat{\theta}(x_1,...,x_n) = \sqrt[3]{\frac{1}{n}\sum_{i=1}^{n} x_i^3}$$

L'informazione osservata di Fisher è

$$\mathcal{I}(\hat{\theta}) = -\frac{3n}{\hat{\theta}^2} + \frac{12\sum_{i=1}^n x_i^3}{\hat{\theta}^5}$$

a cui corrispondono i valori numerici

$$\hat{\theta} = 5.921825$$

e

$$\mathcal{I}(\hat{\theta}) = 0.7699324$$

**Esercizio 28.**

1. Il modello statistico probabilistico è dato dalla terna costituita da: *i)* spazio campionario, *ii)* dalla singola distribuzione (in questo caso densità di probabilità) dipendente da un parametro e *iii)* dallo spazio parametrico. Complessivamente fornisce una specificazione esauriente della

famiglia di distribuzioni che si suppone governi l'aleatorietà della singola osservazione campionaria. Nel caso in questione il modello per la singola osservazione $X$ si scrive formalmente come segue:

$$\left\{ \mathcal{X} = (0, \infty)\,;\; f_X(x;\theta) = 2\,\theta\, e^{-2\,\theta x}\,,\; \theta \in \Theta = (0, \infty) \right\}$$

Il modello appartiene alla famiglia esponenziale dal momento che è possibile scrivere la funzione di densità (dipendente dal parametro $\theta$) nel seguente modo

$$f_X(x;\theta) = h(x) \exp\left\{ \eta(\theta) T(x) - \beta(\theta) \right\}$$

per un'opportuna scelta delle funzioni

$$\begin{aligned} h(x) &= 2 \\ \eta(\theta) &= -2\theta \\ T(x) &= x \\ \beta(\theta) &= -\log\theta \end{aligned}$$

2. Dall'espressione della funzione di verosimiglianza, a meno di una costante moltiplicativa (ininfluente ai fini della determianzione degli stimatori)

$$L_{\boldsymbol{X}}(\theta) = \prod_{i=1}^{n} f_X(X_i;\theta) = 2^n\, \theta^n\, e^{-2\,\theta \sum_{i=1}^{n} x_i} \propto \theta^n\, e^{-2\,\theta \sum_{i=1}^{n} X_i}$$

e dalla sua trasformazione logaritmica

$$\ell(\theta) = n \log\theta - 2\,\theta \sum_{i=1}^{n} X_i$$

si ottiene per via analitica l'argomento di massimo risolvendo l'equazione

$$\frac{d}{d\theta}\ell(\theta) = \frac{n}{\theta} - 2\sum_{i=1}^{n} X_i = 0$$

L'unica soluzione corrisponde a $\theta$ uguale a

$$\hat{\theta}(X_1, ..., X_n) = \frac{n}{2\sum_{i=1}^{n} X_i} = \frac{1}{2\bar{X}}$$

che corrisponde allo stimatore di massima verosimiglianza dal momento che il punto stesso è l'argomento di massimo <u>assoluto</u> della funzione. Infatti ciò discende dalla verifica del cambio di segno della derivata che cambia nel punto di stazionario da positivo a negativo.

Per il parametro $\gamma = \frac{2}{\theta}$ si può rapidamente concludere che, sfruttando la proprietà di invarianza degli stimatori di massima verosimiglianza, lo stimatore di massima verosimiglianza per $\gamma = g(\theta) = \frac{2}{\theta}$ sarà

$$\hat{\gamma}(X_1, ..., X_n) = g(\hat{\theta}(X_1, ..., X_n)) = \frac{2}{\frac{1}{2\bar{X}}} = 4\bar{X}.$$

In corrispondenza dei valori campionari osservati, per i quali risulta $\bar{x} = 1500/150 = 10$, si ottengono le seguenti stime di massima verosimiglianza per $\theta$ e per $\gamma$:

$$\begin{aligned} \hat{\theta} &= \frac{1}{2 \cdot 10} = \frac{1}{20} = 0.05 \\ \hat{\gamma} &= 4 \cdot 10 = 40 \end{aligned}$$

3. Per ottenere l'espressione dell'informazione di Fisher osservata si deve determinare la derivata seconda della funzione di logverosimiglianza e cambiarla di segno ovvero

$$-\frac{d^2}{d\theta^2}\ell(\theta) \quad = \quad \frac{n}{\theta^2}$$

e calcolarla nel punto $\theta = \hat{\theta}(X_1, ..., X_n)$ ottenendo

$$\begin{aligned} I(\boldsymbol{X}_{oss}) &= \frac{n}{\hat{\theta}(X_1, ..., X_n)^2} \\ &= n \cdot (2\bar{X})^2 \end{aligned}$$

In corrispondenza del campione osservato l'informazione osservata vale

$$I(\boldsymbol{X}_{oss}) = 150 \cdot (2 \cdot 10)^2 = 150 \cdot 400 = 60000$$

**Esercizio 29.**

1. La funzione di verosimiglianza è:
$L_{\mathbf{x}_{oss}}(\theta) = \prod_{i=1}^{n} \big[\, (x_i + 1)\,(1-\theta)^2\,\theta^{x_i} \big] = \big[\, \prod_{i=1}^{n} (x_i+1) \,\big] (1-\theta)^{2n}\, \theta^{\sum x_i}$.
La funzione di log-verosimiglianza è:
$\ell(\theta) = \sum_{i=1}^{n} \log(x_i + 1) + 2n\log(1-\theta) + \sum_{i=1}^{n} x_i \log\theta$, con derivata prima
$\ell'(\theta) = -\frac{2n}{1-\theta} + \frac{\sum_{i=1}^{n} x_i}{\theta}$, che si azzera per $-2\,n\,\theta + (1-\theta)\sum_{i=1}^{n} x_i = 0$ e cioè per $\hat{\theta} = \frac{\sum_{i=1}^{n} x_i}{2n + \sum_{i=1}^{n} x_i}$. La derivata seconda della funzione di log-verosimiglianza è data da:

$$\ell''(\theta) = -\frac{2n}{(1-\theta)^2} - \frac{\sum_{i=1}^{n} x_i}{\theta^2} < 0$$

per ogni valore di $\theta$, quindi $\hat{\theta}$ è stima di massima verosimiglilanza.

2. Dato che $n = 8$, $\sum_{i=1}^{n} = 48$ si ha $\hat{\theta}_{MLE} = \frac{48}{16+48} = \frac{48}{64} = \frac{3}{4}$

3. 
$$\frac{L_{\mathbf{x}_{oss}}(\theta_1)}{L_{\mathbf{x}_{oss}}(\theta_2)} = \frac{\left(\frac{3}{10}\right)^{16}\left(\frac{7}{10}\right)^{48}}{\left(\frac{2}{10}\right)^{16}\left(\frac{8}{10}\right)^{48}} = 1.08 > 1.$$

Il valore $\theta_1 = \frac{7}{10}$ è il più verosimile.

4. 
$$\mathbb{P}(X = 0; \theta = \hat{\theta}) = (1-\hat{\theta})^2 = \left(\frac{1}{4}\right)^2 \qquad \mathbb{P}(X = 1; \theta = \hat{\theta}) = 2(1-\hat{\theta})^2\,\hat{\theta} = 2\left(\frac{1}{4}\right)^2 \frac{3}{4}.$$

e

$$\mathbb{P}\big(\,\mathbf{X}_3 = (0,1,0)\,\big) = 2\left(\frac{1}{4}\right)^6 \frac{3}{4}.$$

**Esercizio 30.**

1. La funzione di verosimiglianza è:

$$L_{\mathbf{x}_{oss}}(\theta) = \prod_{i=1}^{n}\left[\frac{1}{2}\,(x_i+1)\,(x_i+2)\,\theta^3\,(1-\theta)^{x_i}\right] = \left[\prod_{i=1}^{n}(x_i+1)\,(x_i+2)\right]2^{-n}\,\theta^{3n}\,(1-\theta)^{\sum x_i}.$$

   La funzione di log-verosimiglianza è:

$$\ell(\theta) = \sum_{i=1}^{n}\log(x_i+1) + \sum_{i=1}^{n}\log(x_i+2) - n\log 2 + 3n\log\theta + \sum_{i=1}^{n}x_i\log(1-\theta),$$

   con derivata prima $\ell'(\theta) = \dfrac{3n}{\theta} - \dfrac{\sum_{i=1}^{n}x_i}{(1-\theta)}$, che si azzera per

   $3\,n\,(1-\theta) - \theta\sum_{i=1}^{n}x_i = 0$ e cioè per $\hat{\theta} = \dfrac{3n}{3n+\sum_{i=1}^{n}x_i}$. La derivata seconda della funzione di log-verosimiglianza è data da:

$$\ell''(\theta) = -\frac{3n}{\theta^2} - \frac{\sum_{i=1}^{n}x_i}{(1-\theta)^2} < 0$$

   per ogni valore di $\theta$, quindi $\hat{\theta}$ è stima di massima verosimiglilanza.

2. Dato che $n=6$, $\sum_{i=1}^{n} = 9$ si ha $\hat{\theta}_{MLE} = \dfrac{18}{18+9} = \dfrac{18}{27} = \dfrac{2}{3}$

3. 
$$\frac{L_{\mathbf{x}_{oss}}(\theta_1)}{L_{\mathbf{x}_{oss}}(\theta_2)} = \frac{\left(\frac{23}{30}\right)^{18}\left(\frac{7}{30}\right)^{9}}{\left(\frac{17}{30}\right)^{18}\left(\frac{13}{30}\right)^{9}} = 0.877 < 1.$$

   Il valore $\theta_2 = \dfrac{17}{30}$ è il più verosimile.

4. 
$$\mathbb{P}(X=0;\theta=\hat{\theta}) = \hat{\theta}^{\,3} = \left(\frac{1}{3}\right)^3 \qquad \mathbb{P}(X=1;\theta=\hat{\theta}) = 3\,\hat{\theta}^{\,3}\,(1-\hat{\theta}) = \left(\frac{1}{3}\right)^2\frac{2}{3}.$$

   e

$$\mathbb{P}\big(\,\mathbf{X}_3 = (0,1,0)\,\big) = \left(\frac{1}{3}\right)^8\frac{2}{3}.$$

**Esercizio 31.**

1. $\left(\mathcal{X}^n = (\mathbb{R}_+)^n, f_n(\mathbf{x}:\theta) = \prod_{i=1}^{n}\left[\,\theta(1+x_i)^{-(1+\theta)}\,\right] = \theta^n\prod_{i=1}^{n}(1+x_i)^{-(1+\theta)}, \Theta = \mathbb{R}_+\right)$

2. Dato che
$$f(x;\theta) = \theta(1+x)^{-(1+\theta)} = \frac{1}{(1+x)}\exp\left\{\log\theta - \theta\log(1+x)\right\}$$

   Ponendo $h(x) = (1+x)^{-1}$, $\eta(\theta) = -\theta$, $T(x) = \log(x+1)$ e $B(\theta) = \log\theta$, si verifica che la legge di probabilità della variabile di base appartiene alla famiglia esponenziale. Una statistica sufficiente è data da $\sum_{i=1}^{n}T(x_i) = \sum\log(1+x_i)$

3. La funzione di verosimiglianza è data da:

$$L_{\mathbf{x}_{oss}}(\theta) = \frac{\theta^n}{\prod_{i=1}^n (1+x_i)} \prod_{i=1}^n (1+x_i)^{-\theta}.$$

Il suo nucleo è la funzione $g(T(\mathbf{x}_{oss}), \theta) = \theta^n \prod_{i=1}^n (1+x_i)^{-\theta}$. La funzione di log verosimiglianza e la sua derivata prima sono: $\ell(\theta) = n \log \theta - \theta \sum_{i=1}^n \log(1+x_i)$ e $\ell'(\theta) = \frac{n}{\theta} - \sum_{i=1}^n \log(1+x_i)$. La derivata della funzione di log verosimiglianza si annulla per $\frac{n}{\theta} - \sum_{i=1}^n \log(1+x_i) = 0$ da questa equazione si ottiene $\hat{\theta} = \frac{n}{\sum_{i=1}^n \log(1+x_i)}$, che è la stima di massima verosimiglianza, dato che la derivata seconda delle funzione di log verosimiglianza $\ell''(\theta) = -n/\theta^2 < 0$ per tutti i valori di $\theta$.

4. L'informazione osservata di Fisher è data da:

$$\mathcal{I}_{\mathbf{x}_{oss}}(\hat{\theta}) = -\ell''(\theta)\big|_{\theta=\hat{\theta}} = \frac{\left[\sum_{i=1}^n \log(1+x_i)\right]^2}{n}$$

5. In corrispondenza al campione osservato si ottiene $\hat{\theta}_{MLE} = \frac{50}{106.6} = 0.47$ e $\mathcal{I}_{\mathbf{x}_{oss}}(\hat{\theta}) = \frac{(106.6)^2}{50} = 227.27$. Quindi

$$L_{\mathbf{x}_{oss}}(\theta) \approx \frac{\sqrt{227.27}}{\sqrt{2\,\pi}} \exp\left\{-\frac{227.27}{2}(\theta - 0.47)^2\right\}.$$

**Esercizio 32.**

1. 

$$\left(\mathcal{X}^n = (I\!R_+)^n, f_n(\mathbf{x} : \theta) = \prod_{i=1}^n \left[\frac{2x_i}{\theta^2} \exp\left\{-\frac{x_i^2}{\theta^2}\right\}\right] = \frac{2^n}{\theta^{2n}} \left[\prod_{i=1}^n x_i\right] e^{-\frac{\sum_{i=1}^n x_i^2}{\theta^2}}, \Theta = I\!R_+\right)$$

2. Dato che

$$f(x;\theta) = \frac{2\,x}{\theta^2}\; e^{-\frac{x^2}{\theta^2}} = 2\,x \exp\left\{-2\log\theta - \frac{x^2}{\theta^2}\right\}$$

Ponendo $h(x) = 2x$, $\eta(\theta) = -\theta^{-2}$, $T(x) = x^2$ e $B(\theta) = -2\log\theta$, si verifica che la legge di probabilità della variabile di base appartiene alla famiglia esponenziale. Una statistica sufficiente è data da $\sum_{i=1}^n T(x_i) = \sum x_i^2$

3. La funzione di verosimiglianza è data da:

$$L_{\mathbf{x}_{oss}}(\theta) = \left[2^n \prod_{i=1}^n x_i\right] \theta^{-2n}\, e^{-\frac{\sum_{i=1}^n x_i^2}{\theta^2}}.$$

Il suo nucleo è la funzione

$$g(T(\mathbf{x}_{oss}), \theta) = \theta^{-2n}\, e^{-\frac{\sum_{i=1}^n x_i^2}{\theta^2}}.$$

La funzione di log verosimiglianza e la sua derivata prima sono:

$$\ell(\theta) = -2n\log\theta - \frac{\sum_{i=1}^n x_i^2}{\theta^2} \qquad \text{e} \qquad \ell'(\theta) = -\frac{2n}{\theta} + \frac{2\sum_{i=1}^n x_i^2}{\theta^3}.$$

La derivata della funzione di log verosimiglianza si annulla per $-n + \frac{\sum_{i=1}^n x_i^2}{\theta^2} = 0$. Da questa equazione si ottiene $\hat{\theta} = \sqrt{\frac{\sum_{i=1}^n x_i^2}{n}}$, che è la stima di massima verosimiglianza, dato che la derivata seconda delle funzione di log verosimiglianza calcolata in $\theta = \hat{\theta}$ è $\ell''(\theta)\big|_{\theta=\hat{\theta}} = \frac{2n}{\theta^2} - \frac{6}{\theta^4}\sum_{i=1}^n x_i^2\big|_{\theta=\hat{\theta}} = \frac{2n\frac{\sum x_i^2}{n} - 6\sum x_i^2}{(\sum x_i^2/n)^4} = \frac{-4\sum x_i^2}{(\sum x_i^2/n)^4} < 0$ .

4. L'informazione osservata di Fisher è data da:

$$\mathcal{I}_{\mathbf{x}_{oss}}(\hat{\theta}) = -\ell''(\theta)\big|_{\theta=\hat{\theta}} = \frac{4n^4}{(\sum x_i^2)^3}$$

5. In corrispondenza al campione osservato si ottiene $\hat{\theta}_{MLE} = \frac{29.01}{10} = 2.9$ e $\mathcal{I}_{\mathbf{x}_{oss}}(\hat{\theta}) = \frac{4 \cdot 10^4}{(29.01^2)^3} = 0.67$. Quindi

$$L_{\mathbf{x}_{oss}}(\theta) \approx \frac{\sqrt{0.67}}{\sqrt{2\,\pi}} \exp\left\{-\frac{0.67}{2}(\theta - 2.9)^2\right\}.$$

## II PARTE: GLI STIMATORI E LE LORO PROPRIETA'

**Esercizio 1*.** Sia $\mathbf{X}_n = (X_1, \ldots, X_n)$ un campione casuale dalla popolazione con distribuzione di probabilità

$$f_X(x;\theta,\alpha) = \frac{\Gamma(\alpha+x)}{\Gamma(x+1)\Gamma(\alpha)}\,\theta^\alpha(1-\theta)^x, \qquad x=0,1,2,\ldots, \quad \theta\in(0,1), \quad \alpha>0.$$

Assumendo che $\alpha$ sia una quantità nota,

a) determinare, se esiste, una statistica sufficiente unidimensionale;

b) determinare il limite inferiore di Cramer-Rao (**Sugg.**: $E_\theta[X] = \alpha(1-\theta)/\theta$).

**Esercizio 2*.** Sia $\mathbf{X}_n = (X_1, \ldots, X_n)$ un campione casuale dalla popolazione con funzione di densità

$$f_X(x;\theta) = \frac{2}{\theta}\left(1-\frac{x}{\theta}\right) I_{[0,\theta]}(x), \qquad \theta>0.$$

a) Determinare la funzione di verosimiglianza e stabilire se esiste una statistica sufficiente unidimensionale.

b) Verificare che $E_\theta(X) = \frac{\theta}{3}$.

c) Determinare uno stimatore non distorto di $\theta$, che sia funzione della media campionaria, $\bar{X}$.

d) Calcolare la varianza dello stimatore determinato al punto precedente e studiarne la consistenza.

**Esercizio 3.** Sia $\mathbf{X}_n = (X_1, \ldots, X_n)$ un campione casuale dalla popolazione con funzione di densità

$$f_X(x;\theta) = \frac{2x}{\theta^2} \qquad 0<x<\theta \qquad \theta>0.$$

a) Calcolare $E_\theta(\bar{X})$, dove $\bar{X} = \sum_{i=1}^n X_i/n$ indica la media campionaria.

b) Determinare lo stimatore dei momenti di $\theta$ e stabilire se è non distorto.

c) Determinare la funzione di verosimiglianza e lo stimatore di massima verosimiglianza del parametro $\theta$.

**Esercizio 4*.** Sia $X_1, \ldots, X_n$ un campione casuale da una popolazione con distribuzione uniforme nell'intervallo $(0,\theta)$, dove $\theta$ è un parametro incognito.

a) Verificare che

$$\mathrm{E}_\theta(X_i) = \frac{\theta}{2} \qquad \mathrm{V}_\theta(X_i) = \frac{\theta^2}{12}.$$

b) Calcolare $\mathrm{E}(\bar{X})$ e $\mathrm{V}(\bar{X})$, dove $\bar{X}$ rappresenta la variabile aleatoria media campionaria.

c) Determinare uno stimatore non distorto di $\theta$, che sia funzione di $\bar{X}$.

d) Verificare la non distorsione dello stimatore definito ponendo

$$T(\mathbf{X}_n) = 2\frac{(n-1)X_1 + X_2}{n},$$

e confrontare la varianza di $T(\mathbf{X}_n)$ con quella dello stimatore ottenuto al punto *c*).

**Esercizio 5*.** Sia $\mathbf{X}_n = (X_1, \ldots, X_n)$ un campione casuale dalla popolazione $N(0, \theta)$. Verificare che la statistica $S_0^2 = \sum_{i=1}^n X_i^2/n$ è uno stimatore non distorto di $\theta$. Lo stimatore considerato è UMVUE?

**Esercizio 6*.** Sia $\mathbf{x}_n = (3, 4, 2, 7, 4, 5, 8, 1, 0, 0)$ un campione casuale dalla popolazione con funzione di massa di probabilità

$$f_X(x; \theta) = e^{-\theta} \, \frac{\theta^x}{x!} \qquad x = 0, 1, 2, \ldots \qquad \theta > 0.$$

a) Calcolare la stima di massima verosimiglianza del parametro $\theta$;

b) Calcolare la probabilità dell'evento $\{X_1 = 1\}$ e, osservando che si tratta di una funzione del parametro incognito $\theta$, calcolare la stima di massima verosimiglianza di tale quantità.

**Esercizio 7*.** Sia $X_1, \ldots, X_n$ un campione casuale in cui $E[X_i] = \mu$ incognita e $\mathrm{V}[X_i] = \sigma^2$ nota e finita. Si consideri la classe di stimatori di $\mu$ ottenuta considerando una generica combinazione lineare delle v.a. $X_i$:

$$T(\mathbf{X}_n) = \sum_{i=1}^n a_i X_i, \qquad a_i \in I\!R.$$

a) Determinare il valore atteso di $T(\mathbf{X}_n)$ e stabilire la condizione sui coefficienti $a_i$ affinchè lo stimatore sia non distorto.

b) Determinare l'espressione di $\mathrm{MSE}_\mu[T(\mathbf{X}_n)]$, per uno stimatore non distorto del tipo considerato.

c) Si scriva l'espressione dei due stimatori $T_1(\mathbf{X}_n)$ e $T_2(\mathbf{X}_n)$, ottenuti da $T(\mathbf{X}_n)$ ponendo
   - $a_i = \frac{1}{n}, \quad i = 1, \ldots, n \qquad$ per $T_1(\mathbf{X}_n)$
   - $a_1 = \frac{n-2}{n}, \quad a_2 = a_3 = \frac{1}{n} \quad$ e $\quad a_4 = \ldots = a_n = 0 \qquad$ per $T_2(\mathbf{X}_n)$.

d) Determinare l'espressione di $\mathrm{MSE}_\mu[T_i(\mathbf{X}_n)], \; i = 1, 2$, e stabilire quale tra i due è più efficiente.

e) Stabilire se i due stimatori considerati sono consistenti.

**Esercizio 8*.** Si consideri una v.a. $X$ con funzione di densità

$$f_X(x; \theta) = \frac{2}{\theta} \, x \, e^{-\frac{x^2}{\theta}} \, I_{(0,+\infty)}(x), \qquad \theta > 0.$$

Si determinini il limite inferiore di Cramer-Rao per gli stimatori non distorti di $\theta$.
(Sugg.: è noto che $E(X^2) = \theta$.)

**Esercizio 9*.** Siano $\bar{X}_1$ e $\bar{X}_2$ le medie campionarie di due campioni casuali indipendenti di dimensioni rispettivamente uguali a $n_1$ e $n_2$, entrambi provenienti dalla stessa popolazione normale di parametri $(\mu; \sigma^2)$. Si consideri la classe di stimatori di $\mu$ ottenuta considerando una generica combinazione lineare delle v.a. $\bar{X}_i$:

$$T_q(\mathbf{X}_n) = q\bar{X}_1 + (1-q)\bar{X}_2 \qquad q \in I\!R.$$

a) Verificare che, per ogni valore di $q \in I\!R$, lo stimatore $T_q(\mathbf{X}_n)$ è non distorto.

b) Determinare l'espressione di $\mathrm{MSE}_\mu[T_q(\mathbf{X}_n)]$.

c) Si scriva l'espressione dei due stimatori $T_1(\mathbf{X}_n)$ e $T_2(\mathbf{X}_n)$, il primo ottenuto ponendo (nell'espressione generica di $T_q(\mathbf{X}_n)$) $q = \frac{n_1}{n_1+n_2}$ e il secondo ponendo $q = \frac{1}{2}$.

d) Determinare l'espressione di $\mathrm{MSE}_\mu[T_i(\mathbf{X}_n)],\ i = 1, 2$.

e) Stabilire se i due stimatori considerati sono consistenti (al crescere di $n_1$ e di $n_2$).

**Esercizio 10*.** Si consideri una v.a. $X$ con funzione di densità

$$f_X(x;\theta) = 2\theta\ x\ e^{-\theta x^2}\ I_{(0,+\infty)}(x), \qquad \theta > 0.$$

Si determinini il limite inferiore di Cramer-Rao per gli stimatori non distorti di $\theta$.

**Esercizio 11.** Sia $X_1, \ldots, X_n$ un campione casuale da una popolazione $N(\theta, 1)$. Verificare che lo stimatore $\sum_{i=1}^n X_i$ è uno stimatore UMVUE di $n\theta$.

**Esercizio 12*.** Siano $\bar{X}_1$ e $\bar{X}_2$ le medie campionarie di due campioni casuali indipendenti di dimensioni rispettivamente uguali a $n_1$ e $n_2$ entrambi provenienti dalla stessa popolazione con media e varianza rispettivamente uguali a $\mu$ e a $\sigma^2$. Si consideri lo stimatore di $\mu$

$$T(\bar{X}_1, \bar{X}_2) = \frac{1}{3}\bar{X}_1 + \frac{2}{3}\bar{X}_2.$$

a) Verificare se lo stimatore è non distorto.

b) Determinarne la varianza dello stimatore, il suo errore quadratico medio e studiarne la consistenza.

**Esercizio 13*.** Sia $X_1, \ldots X_n$ un campione casuale da una popolazione bernoulliana di parametro incognito $\theta$. Si considerino i seguenti due stimatori per $\theta$:

$$T_1(\mathbf{X}_n) = \bar{X}_n \qquad \text{e} \qquad T_2(\mathbf{X}_n) = \frac{1 + \sum_{i=1}^n X_i}{n+2}.$$

a) Determinare l'errore quadratico medio dei due stimatori ($MSE_\theta[T_i(\mathbf{X}_n)], \quad i = 1, 2$);

b) Studiare, al crescere della dimensione campionaria, $n$, il comportamento della distorsione dei due stimatori ($B_\theta[T_i(\mathbf{X}_n)], \quad i = 1, 2$)

c) Studiare la consistenza dei due stimatori.

**Esercizio 14.** Sia $X_1, \ldots X_n$ un campione casuale da una popolazione uniforme in $(0, \theta)$, $\theta > 0$. Si consideri la statistica campionaria

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

per la quale è noto che, $\forall \theta > 0$,

$$E_\theta[X_{(n)}] = \frac{n}{n+1}\,\theta \qquad \text{e} \qquad V_\theta[X_{(n)}] = \frac{n}{(n+1)^2(n+2)}\,\theta^2.$$

a) Determinare uno stimatore non distorto di $\theta$ basato sullo stimatore di massima verosimiglianza, $X_{(n)}$.

b) Determinare lo stimatore dei momenti.

c) Determinare gli errori quadratici medi dei tre stimatori considerati, confrontarli e stabilire se gli stimatori sono consistenti (in errore quadratico medio).

c) Supponendo di avere osservato il campione di dati $\mathbf{x}_n = (1, 9, 3, 4, 5, 3, 2, 0, 10, 5)$, determinare le stime puntuali di $\theta$ basate sui tre stimatori considerati.

**Esercizio 15*.** Sia $X_1, \ldots, X_n$ un campione casuale da una popolazione bernoulliana di parametro incognito $\theta$. Per ciascuno dei due seguenti stimatori di $\theta$:

$$T_1(\mathbf{X}_n) = \bar{X}_n \qquad \text{e} \qquad T_2(\mathbf{X}_n) = \frac{n\bar{X}_n + \sqrt{n/4}}{n + \sqrt{n}},$$

a) determinare il valore atteso e la distorsione;

b) determinare l'errore quadratico medio e studiare la consistenza;

c) verificare se esiste un valore di $\theta$ per il quale si abbia che $E_\theta[T_2] = \theta$.

**Esercizio 16*.** Sia $X_1, \ldots, X_n$ un campione casuale da una popolazione con valore atteso pari a $\theta$ e varianza pari a $\sigma^2$. Per il seguente stimatore del parametro $\theta$

$$T(\mathbf{X}_n) = \frac{X_1 + \ldots + X_{n-1}}{n-1} + \frac{X_n}{n}$$

a) determinare il valore atteso e studiare la distorsione;

b) determinare l'errore quadratico medio e studiare la consistenza.

**Esercizio 17*.** Siano $\mathbf{X}_{n_1}$ e $\mathbf{X}_{n_2}$ due campioni casuali indipendenti di dimensioni $n_1$ e $n_2$ ($n_1 < n_2$), provenienti da una popolazione normale di parametri $\theta$ e $\sigma^2$. Si considerino i seguenti quattro stimatori per $\theta$:

$$T_1 = \bar{X}_{n_1}, \qquad T_2 = \bar{X}_{n_2}, \qquad T_3 = \frac{\bar{X}_{n_1} + \bar{X}_{n_2}}{2}, \qquad T_4 = \frac{n_1\bar{X}_{n_1} + n_2\bar{X}_{n_2}}{n_1 + n_2}.$$

Di tali stimatori:

a) determinare la distorsione e la varianza;

b) calcolare l'errore quadratico medio e stabilire qual è il più efficiente;

c) studiare la consistenza.

**Esercizio 18*.** Sia $X_1, \ldots, X_n$ un campione casuale da una popolazione $\mathrm{N}(\theta, 1)$. Si consideri la seguente famiglia di stimatori del parametro incognito $\theta$:

$$T_a(\mathbf{X}_n) = \omega_n\bar{X}_n + (1 - \omega_n)a, \qquad a \in \mathbb{R},$$

definita come media ponderata dello stimatore UMVUE, $\bar{X}_n$, e della costante reale $a$, con pesi $\omega_n = n/(n+1)$ e $1 - \omega_n = 1/(n+1)$.

a) Per gli stimatori $T_a$, determinare distorsione, varianza ed errore quadratico medio e studiarne la consistenza.

b) Stabilire se esistono dei valori di $\theta$ per i quali lo stimatore $T_a$, ottenuto ponendo $a = 0$, risulta migliore di $\bar{X}_n$.

c) Per i due stimatori considerati al punto (b) (ovvero $T_0$ e $\bar{X}_n$), tracciare i grafici (approssimativi) delle due funzioni MSE, al variare di $\theta$ in $\mathbb{R}$.

**Esercizio 19*.** Si consideri un campione casuale di $n$ osservazioni da una popolazione $X$ con distribuzione di probabilità

$$f_X(x; \theta) = \frac{\theta}{(1+x)^{1+\theta}} \qquad x > 0, \qquad \theta > 0.$$

a) Scrivere il modello statistico-probabilistico associato al campione casuale.

b) Determinare l'espressione della funzione di verosimiglianza del parametro associata a un generico campione osservato, $\mathbf{x}_n$; individuare il nucleo della funzione di verosimiglianza e, se esiste, una statistica sufficiente unidimensionale.

c) Determinare la stima di massima verosimiglianza di $\theta$ e calcolarne il valore per un campione di dimensione $n = 10$ in cui si ha che

$$\sum_{i=1}^{10} \ln(1 + x_i) = 5.$$

d) Verificare se la famiglia di densità $\mathcal{F} = \{f_X(\cdot;\theta);\ \theta \in \Theta\}$ è una famiglia esponenziale.

**Esercizio 20*.** Sia $X_1, \dots X_n$ un campione casuale da una popolazione binomiale di parametri incogniti $(m, \theta)$, con funzione di massa di probabilità

$$f_X(x;\theta) = \binom{m}{x} \theta^x (1-\theta)^{m-x}, \qquad x = 0, 1, 2, \dots, m.$$

a) Determinare lo stimatore dei momenti di $\theta$, $\hat{\theta}_M(\mathbf{X}_n)$.

b) Determinare distorsione, varianza ed errore quadratico medio dello stimatore $\hat{\theta}_M(\mathbf{X}_n)$.

c) Studiare la consistenza di $\hat{\theta}_M(\mathbf{X}_n)$.

d) Determinare lo stimatore di massima verosimiglianza di $\theta$ e confrontarne le proprietà con quelle dello stimatore dei momenti.

e) Stabilire se esiste lo stimatore UMVUE e, in caso di risposta affermativa, determinarlo.

**Esercizio 21*.** Sia $X_1, \dots X_n$ un campione casuale da una popolazione con funzione di massa di probabilità

$$f_X(x;\theta) = (2\theta - 1)^x (2 - 2\theta)^{1-x} \qquad x = 0, 1, \qquad \theta \in \left(\frac{1}{2}, 1\right).$$

1. Verificare che
$$\mathbb{E}[X] = 2\theta - 1, \qquad \mathbb{V}[X] = (2\theta - 1)(2 - 2\theta).$$

2. Determinare lo stimatore dei momenti di $\theta$, $\hat{\theta}_M(\mathbf{X}_n)$, calcolarne distorsione, varianza ed errore quadratico medio e stabilire se si tratta di stimatore UMVUE.

3. Studiare la consistenza di $\hat{\theta}_M(\mathbf{X}_n)$ e determinare l'approssimazione normale della sua distribuzione campionaria.

4. Verificare che lo stimatore di massima verosimiglianza di $\theta$ coincide con lo stimatore dei momenti.

**Esercizio 22*.** Si consideri un campione casuale di $n$ osservazioni da una popolazione $X$ con funzione di densità:

$$f_X(x;\theta) = \frac{x^2}{2\,\theta^3} \exp\left\{-x/\theta\right\} \qquad \theta > 0, \qquad x > 0.$$

Per questa variabile aleatoria si ha che $\mathbb{E}[X] = 3\,\theta$ e $\mathbb{V}[X] = 3\,\theta^2$.

1. Scrivere il modello statistico probabilistico per il campione casuale $\mathbf{X}_n$, verificare che $f_X(x;\theta)$ appartiene alla famiglia esponenziale e ricavare una statistica sufficiente per $\theta$.

2. Scrivere la funzione di verosimiglianza. Ottenere lo stimatore $\hat{\theta}_{MLE}$ di massima verosimiglianza di $\theta$ e lo stimatore di massima verosimiglianza di $\psi = \theta + \theta^2$, funzione del parametro $\theta$.

3. Calcolare il momento primo e secondo della variabile aleatoria $\bar{X}_n$, media campionaria.

4. Verificare che $\hat{\psi}_{MLE}$ è uno stimatore distorto di $\psi$, calcolarne la distorsione e stabilire se è asintoticamente non distorto.

**Esercizio 23*.** Sia $\mathbf{X}_n = (X_1, X_2, \ldots, X_n)$ un campione casuale di $n$ osservazioni estratte da una popolazione $X$ con distribuzione di probabilità:

$$f_X(x;\theta) = \theta^{-1}\, x^{-\frac{1}{\theta}-1} \qquad x > 1 \qquad \theta \in (0,1)$$

Per questa variabile aleatoria si ha che $\mathbb{E}[\log X] = \theta$ e $\mathbb{V}[\log X] = \theta^2$

1. Determinare la funzione di verosimiglianza $L_{\mathbf{x}_{oss}}(\theta)$ ed ottenere lo stimatore di massima verosimiglianza $\hat{\theta}_{MLE}$

2. Verificare che l'espressione dell'informazione osservata di Fisher è $\frac{n^3}{(\sum_{i=1}^n \log x_i)^2}$. Scrivere l'approssimazione normale della funzione di verosimiglianza e spiegare perché utilizzando questa approssimazione si può ricavare un intervallo di verosimiglianza approssimato ad un livello fissato $q$.

   **Parte facoltativa.** Scrivere gli estremi dell'intervallo approssimato.

3. Studiare la correttezza e la consistenza dello stimatore $\hat{\theta}_{MLE}$.

4. Verificare se $\hat{\theta}_{MLE}$ è uno stimatore UMVUE.

**Esercizio 24.** Sia $X_1, \ldots, X_n$ un campione casuale proveniente dalla popolazione con funzione di densità di probabilità

$$f_X(x;\theta) = \frac{3}{\theta^3}x^2, \qquad 0 < x < \theta, \qquad \theta > 0.$$

1. Verificare che

   $$\mathbb{E}[X] = \frac{3}{4}\theta, \qquad \mathbb{V}[X] = \frac{3}{80}\theta^2$$

   e che lo stimatore dei momenti di $\theta$ è

   $$\hat{\theta}_M = \frac{4}{3}\,\bar{X}_n$$

2. Determinare l'errore quadratico medio dello stimatore dei momenti e studiarne la consistenza.

3. Determinare la distribuzione asintotica dello stimatore $\hat{\theta}_M$ .

4. Determinare la stima con metodo dei momenti supponendo di avere osservato un campione di dimensione $n = 36$ per il quale $\sum_{i=1}^n x_i = \frac{3}{2}$.

**Esercizio 25.** Sia $X_1, \ldots, X_n$ un campione casuale proveniente dalla popolazione con funzione di densità di probabilità

$$f_X(x;\theta) = \frac{1}{2\sqrt{\theta}}\frac{1}{\sqrt{x}}, \qquad 0 < x < \theta, \qquad \theta > 0.$$

1. Verificare che
$$\mathbb{E}[X] = \frac{1}{3}\theta, \qquad \mathbb{V}[X] = \frac{4}{45}\theta^2$$
e che lo stimatore dei momenti di $\theta$ è
$$\hat{\theta}_M = 3\bar{X}_n$$

2. Determinare l'errore quadratico medio dello stimatore dei momenti e studiarne la consistenza.

3. Determinare la distribuzione asintotica dello stimatore $\hat{\theta}_M$.

4. Determinare la stima dei momenti supponendo di avere osservato un campione di dimensione $n = 36$ per il quale $\sum_{i=1}^{n} x_i = \frac{1}{3}$.

**Esercizio 26.** Siano $X_1^1, \ldots, X_{n_1}^1$ e $X_1^2, \ldots, X_{n_2}^2$ due campioni casuali indipendenti, rispettivamente di ampiezza $n_1$ e $n_2$, provenienti da distribuzioni $N(\theta_1, 1)$ e $N(\theta_2, 1)$. Si consideri il parametro incognito
$$\theta = \frac{\theta_1 + \theta_2}{2}$$
e lo stimatore
$$T(\mathbf{X}_n) = \frac{\bar{X}_1 + \bar{X}_2}{2},$$
dove $\bar{X}_1$ e $\bar{X}_2$ sono le medie campionarie dei due campioni.

1. Verificare che lo stimatore $T(\mathbf{X}_n)$ è non distorto e consistente per il parametro $\theta$.

2. Determinare la distribuzione campionaria di $T(\mathbf{X}_n)$.

**Esercizio 27.** Sia $X_1, \ldots, X_n$ un campione casuale proveniente da una distribuzione uniforme nell'intervallo $[0, \theta]$.

1. Si confrontino e si discutano le proprietà inferenziali dei seguenti due stimatori per campioni di ampiezza $n$ fissata:
$$\begin{aligned} T_1(\mathbf{X}_n) &= \frac{n+1}{n} X_{(n)}, \\ T_2(\mathbf{X}_n) &= \frac{n+2}{n+1} X_{(n)}. \end{aligned}$$
*Suggerimento*: si ricordi che
$$\begin{aligned} \mathbb{E}\left[\frac{X_{(n)}}{\theta}\right] &= \frac{n}{n+1} \\ \mathbb{V}\left[\frac{X_{(n)}}{\theta}\right] &= \frac{n}{(n+1)^2(n+2)} \end{aligned}$$

2. Determinare $T_3(\mathbf{X}_n) = \hat{\theta}_M$ lo stimatore dei momenti di $\theta$ e la sua distribuzione asintotica.

3. Dato un campione osservato di dimensione $n = 20$ in cui
$$x_{(1)} = 0.2, \qquad \bar{x}_n = 0.4, \qquad S_n^2 = 0.1, \qquad x_{(n)} = 0.9$$
determinare le tre stime puntuali per il parametro incognito.

## SOLUZIONI

Proprietà degli stimatori

**Esercizio 1.**

$$\begin{aligned}
f_n(\mathbf{x}_n;\theta) &= \left(\prod_{i=1}^n \frac{\Gamma(\alpha+x_i)}{\Gamma(x_i+1)\Gamma(\alpha)}\right)\theta^{n\alpha}(1-\theta)^{\sum_{i=1}^n x_i}\\
L(\theta;\mathbf{x}_n) &= \theta^{n\alpha}(1-\theta)^{\sum_{i=1}^n x_i}\\
\log L(\theta;\mathbf{x}_n) &= n\alpha\log\theta+(\sum_{i=1}^n x_i)\log(1-\theta)\\
\frac{\partial^2}{\partial\theta^2}\log L(\theta;\mathbf{x}_n) &= -\frac{n\alpha}{\theta^2}-\frac{\sum_{i=1}^n x_i}{(1-\theta)^2}\\
I_n(\theta) &= -E_\theta\left(\frac{\partial^2}{\partial\theta^2}\log L(\theta;\mathbf{X}_n)\right)=\frac{n\alpha}{\theta^2(1-\theta)}.
\end{aligned}$$

La statistica sufficiente è $\sum_{i=1}^n X_i$; il limite inferiore di Cramer-Rao è $\frac{\theta^2(1-\theta)}{n\alpha}$.

**Esercizio 2.**

$$\begin{aligned}
f_n(\mathbf{x}_n;\theta) &= \frac{2^n}{\theta^n}\prod_{i=1}^n\left(1-\frac{x_i}{\theta}\right)\prod_{i=1}^n I_{[0,\theta]}(x_i)\\
L(\theta;\mathbf{x}_n) &= \frac{1}{\theta^n}\prod_{i=1}^n\left(1-\frac{x_i}{\theta}\right)I_{(x_{(n)},+\infty)}(\theta)
\end{aligned}$$

Non esiste una statistica sufficiente di dimensione 1.

$$E_\theta(X)=\int_0^\theta x\frac{2}{\theta}\left(1-\frac{x}{\theta}\right)\,dx=\cdots=\frac{2}{\theta}\left[\frac{x^2}{2}-\frac{x^3}{3\theta}\right]_0^\theta=\frac{\theta}{3}$$

quindi uno stimatore non distorto basato sulla media campionaria è $\hat\theta=3\bar X$.

Per la varianza di $\hat\theta$, si noti che

$$V_\theta(\hat\theta)=V_\theta(3\bar X)=9\cdot V_\theta(\bar X)=\frac{9}{n}V_\theta(X).$$

Troviamo la varianza della popolazione come $V_\theta(X)=E_\Theta(X^2)-(E_\theta(X))^2$

$$E_\theta(X^2)=\int_0^\theta x^2\frac{2}{\theta}\left(1-\frac{x}{\theta}\right)\,dx=\cdots=\frac{2}{\theta}\left[\frac{x^3}{3}-\frac{x^4}{4\theta}\right]_0^\theta=\frac{\theta^2}{6}$$

da cui $V_\theta(X)=\frac{\theta^2}{6}-(\frac{\theta}{3})^2=\frac{\theta^2}{18}$ e $V_\theta(\hat\theta)=\frac{\theta^2}{2n}$. Per $n\to+\infty$ si ha che $MSE_\theta(\hat\theta)=V_\theta(\hat\theta)\to 0$, quindi $\hat\theta$ è consistente.

**Esercizio 3.** a) Cominciamo col calcolare $E_\theta(X)$ e $V_\theta(X)$ (ci serviranno entrambi):

$$\begin{aligned}
E_\theta(X) &= \int_0^\theta x\frac{2x}{\theta^2}=\frac{2\theta}{3}\\
E_\theta(X^2) &= \int_0^\theta x^2\frac{2x}{\theta^2}=\frac{\theta^2}{2}
\end{aligned}$$

per cui $V_\theta(X) = \frac{\theta^2}{2} - \frac{4\theta^2}{9} = \frac{\theta^2}{18}$. $\qquad E_\theta(\bar{X}) = E_\theta(X) = \frac{2\theta}{3}$.

Lo stimatore con il metodo dei momenti si ottiene ponendo $E_\theta(X) = \bar{X}$. Nel nostro caso, indicando con $\hat{\theta}_M$ lo stimatore di $\theta$ ottenuto con il metodo dei momenti

$$\theta = \frac{3E_\theta(X)}{2} \qquad \text{quindi} \qquad \hat{\theta}_M = \frac{3\bar{X}}{2}.$$

b) Si vede facilmente che $\hat{\theta}_M$ è non distorto.

c)Si ha che:

$$\begin{aligned} f_n(\mathbf{x}_n;\theta) &= \frac{2^n \prod_{i=1}^n x_i}{\theta^{2n}} \prod_{i=1}^n I_{[0,\theta]}(x_i) \\ L(\theta;\mathbf{x}_n) &= \frac{1}{\theta^{2n}} I_{(x_{(n)},+\infty)}(\theta). \end{aligned}$$

l$L(\theta;\mathbf{x}_n)$ è decrescente rispetto a $\theta$. Il massimo si ottiene per $\theta = x_{(n)}$, da cui $\hat{\theta}_{MV} = x_{(n)}$.

**Esercizio 4.** $E_\theta(\bar{X}) = E_\theta(X_i) = \frac{\theta}{2}$; $V_\theta(\bar{X}) = \frac{V_\theta(X_i)}{n} = \frac{\theta^2}{12n}$.

$T_1 = 2\bar{X}$ è funzione di $\bar{X}$ ed è non distorto, infatti $E_\theta(T_1) = 2E_\theta(\bar{X}) = \theta$.

Si ha che $MSE_\theta(T_1) = V_\theta(T_1) = 4V_\theta(\bar{X}) = \frac{\theta^2}{3n}$.

Anche lo stimatore $T = 2\frac{(n-1)X_1 + X_2}{n}$ è non distorto ma è meno efficiente di $T_1$, infatti

$$\begin{aligned} E_\theta(T) &= E\left(2\frac{(n-1)X_1+X_2}{n}\right) = \frac{2}{n}\left((n-1)E_\theta(X_1) + E_\theta(X_n)\right) = \frac{2}{n}\left((n-1)\frac{\theta}{2} + \frac{\theta}{2}\right) = \theta \\ V_\theta(T) &= \frac{4}{n^2}\left((n-1)^2 V_\theta(X_1) + V_\theta((X_n)\right) = \frac{4}{n^2}\left((n-1)^2\frac{\theta^2}{12} + \frac{\theta^2}{12}\right) = \frac{\theta^2((n-1)^2+1)}{3n^2} \end{aligned}$$

e quindi l'efficienza relativa

$$eff(T_1/T) = \frac{MSE_\theta(T)}{MSE_\theta(T_1)} = \frac{V_\theta(T)}{V_\theta(T_1)} = \frac{(n-1)^2+1}{n} > 1.$$

**Esercizio 5.** Poiché $E_\theta(X) = 0$, si ha $E_\theta(X^2) = V_\theta(X) = \theta$ e quindi $E_\theta(X_i^2) = \theta$. Indicando con $T = \frac{1}{n}\sum_{i=1}^n X_i^2$, abbiamo che

$$E_\theta(T) = \frac{1}{n}\sum_{i=1}^n E_\theta(X_i^2) = \frac{1}{n}n\theta = \theta$$

per cui $T$ è corretto.

Calcoliamo il $MSE[T]$, che coincide con $V_\theta[T]$. Poichè $X_i|\theta \sim N(0,\theta)$, si ha che $\frac{X_i}{\sqrt{\theta}}|\theta \sim N(0,1)$, $\frac{X_i^2}{\theta}|\theta \sim \chi_1^2$ e, per l'indipendenza delle $X_i$ e la proprietà di additività del chi quadrato,

$$W = \frac{\sum_{i=1}^n X_i^2}{\theta} \sim \chi_n^2.$$

Pertanto, $V_\theta[W] = 2n$. Poichè $T = \frac{\theta W}{n}$, si ha che $V_\theta[T] = \frac{\theta^2}{n^2}V_\theta[W] = \frac{2\theta^2}{n}$. In alternativa:

$$MSE_\theta(T) \;=\; V_\theta(T) = \frac{V_\theta(X_i^2)}{n} = \frac{E_\theta(X_i^4) - (E_\theta(X_i^2))^2}{n} = \frac{3\theta^2 - \theta^2}{n} = \frac{2\theta^2}{n}.$$

Ora calcoliamo il limite inferiore di Cramer-Rao

$$\begin{aligned} L(\theta, \mathbf{x}_n) &= \theta^{-\frac{n}{2}} e^{-\frac{1}{2\theta}\sum_{i=1}^n X_i^2} = \theta^{-\frac{n}{2}} e^{-\frac{nT}{2\theta}} \\ \log L(\theta, \mathbf{x}_n) &= -\frac{n}{2}\log\theta - \frac{nT}{2\theta} \\ \frac{\partial^2}{\partial\theta^2}\log L(\theta, \mathbf{x}_n) &= \frac{n}{2\theta^2} - \frac{nT}{\theta^3} \\ -E\left(\frac{\partial^2}{\partial\theta^2}\log L(\theta; \mathbf{X}_n)\right) &= E\left(-\frac{n}{2\theta^2} + \frac{nT}{\theta^3}\right) = -\frac{n}{2\theta^2} + \frac{n\theta}{\theta^3} = \frac{n}{2\theta^2} \end{aligned}$$

e quindi l'estremo di Cramer-Rao è $2\theta^2/n$ e coincide con $MSE_\theta(T)$. $T$ è UMVUE.

**Esercizio 6.** Si ha:

$$\begin{aligned} L(\theta; \mathbf{x}_n) &= e^{-n\theta}\theta^{\sum_{i=1}^n x_i} \\ \log L(\theta; \mathbf{x}_n) &= -n\theta + \sum_{i=1}^n x_i \log\theta \\ \frac{\partial}{\partial\theta}\log L(\theta; \mathbf{x}_n) &= -n + \frac{\sum_{i=1}^n x_i}{\theta} \end{aligned}$$

quindi lo stimatore di massima verosimiglianza è $\hat{\theta} = \frac{\sum_{i=1}^n X_i}{n} = \bar{X}$. La stima di massima verosimiglianza relativa al campione osservato è 3.4.

Sappiamo che $P(X_1 = 1) = \theta e^{-\theta}$, quindi la stima di massima verosimiglianza sarà

$$\overbrace{P(X_1 = 1)} = \hat{\theta} e^{-\hat{\theta}} = 0.1134691.$$

**Esercizio 7.** Si ha:

$$E_\theta(T) = E_\theta(\sum_{i=1}^n a_i X_i) = \sum_{i=1}^n a_i E_\theta(X_i) = \mu \sum_{i=1}^n a_i$$

quindi la condizione affiche $T$ sia corretto è $\sum_{i=1}^n a_i = 1$.

Nel caso di $T$ non distorto,

$$MSE_\theta(T) = V_\theta(T) = V_\theta(\sum_{i=1}^n a_i X_i) = \sum_{i=1}^n a_i^2 V_\theta(X_i) = \sigma^2 \sum_{i=1}^n a_i^2.$$

Gli stimatori sono

$$T_1 = \bar{X} \qquad T_2 = \frac{(n-2)X_1 + X_2 + X_3}{n}$$

e soddisfano la condizione trovata al punto (a), quindi sono entrambi corretti. Gli errori quadratici medi sono

$$MSE_\theta(T_1) = \sigma^2 \sum_{i=1}^n \frac{1}{n^2} = \frac{\sigma^2}{n} \qquad MSE_\theta(T_2) = \sigma^2\left(\frac{(n-2)^2}{n^2} + \frac{1}{n^2} + \frac{1}{n^2}\right) = \sigma^2 \frac{n^2 - 4n + 6}{n^2}$$

e per $n \to \infty$ si ha

$$MSE_\theta(T_1) = \frac{\sigma^2}{n} \to 0 \qquad MSE_\theta(T_2) = \sigma^2 \frac{n^2 - 4n + 6}{n^2} \to \sigma^2.$$

Si vede quindi che $T_1$ è consistente, per $T_2$ sembrerebbe di no. In effetti per $n \to \infty$ si ha che $T_2 \to X_1$ e quindi $T_2$ non è consistente.

**Esercizio 8.** Si ha:

$$\begin{aligned}
L(\theta; \mathbf{x}_n) &= \theta^{-n} e^{-\frac{1}{\theta}\sum_{i=1}^n x_i^2} \\
\log L(\theta; \mathbf{x}_n) &= -n \log \theta - \frac{\sum_{i=1}^n x_i^2}{\theta} \\
\frac{\partial^2}{\partial \theta^2} \log L(\theta; \mathbf{x}_n) &= \frac{n}{\theta^2} - 2\frac{\sum_{i=1}^n x_i^2}{\theta^3} \\
-E\left(\frac{\partial^2}{\partial \theta^2} \log L(\theta; \mathbf{X}_n)\right) &= E\left(-\frac{n}{\theta^2} + 2\frac{\sum_{i=1}^n X_i^2}{\theta^3}\right) = -\frac{n}{\theta^2} + 2\frac{n\theta}{\theta^3} = \frac{n}{\theta^2}
\end{aligned}$$

quindi il limite inferiore di Cramer-Rao è $\frac{\theta^2}{n}$.

**Esercizio 9.** Sappiamo che $E_\theta(\bar{X}_1) = \mu$ e $V_\theta(\bar{X}_1) = \sigma^2/n_1$; analogamente, $E_\theta(\bar{X}_2) = \mu$ e $V_\theta(\bar{X}_2) = \sigma^2/n_2$. Per quanto riguarda $T$ abbiamo:

$$E_\theta(T) = E_\theta(q\bar{X}_1 + (1-q)\bar{X}_2) = qE_\theta(\bar{X}_1) + (1-q)E_\theta(\bar{X}_2) = q\mu + (1-q)\mu = \mu$$

e quindi $T$ è non distorto. L'errore quadratico medio è

$$MSE_\theta(T) = V_\theta(T) = V_\theta(q\bar{X}_1 + (1-q)\bar{X}_2) = q^2 V_\theta(\bar{X}_1) + (1-q)^2 V_\theta(\bar{X}_2) = \frac{q^2\sigma^2}{n_1} + \frac{(1-q)^2\sigma^2}{n_2}.$$

$$T_1 = \frac{n_1}{n_1+n_2}\bar{X}_1 + \frac{n_2}{n_1+n_2}\bar{X}_2 \qquad T_2 = \frac{1}{2}\bar{X}_1 + \frac{1}{2}\bar{X}_2$$

$$\begin{aligned}
MSE_\theta(T_1) &= \frac{n_1^2\sigma^2}{n_1(n_1+n_2)^2} + \frac{n_2^2\sigma^2}{n_2(n_1+n_2)^2} = \frac{\sigma^2}{n_1+n_2} \overset{n_1,n_2\to\infty}{\to} 0 \\
MSE_\theta(T_2) &= \frac{\sigma^2}{4n_1} + \frac{\sigma^2}{4n_2} \overset{n_1,n_2\to\infty}{\to} 0
\end{aligned}$$

quindi gli stimatori sono entrambi consistenti. NOTA: il fatto che la popolazione fosse normale non è servito a nulla.

**Esercizio 10.** Si ha

$$\begin{aligned}
L(\theta; \mathbf{x}_n) &= \theta^n e^{-\theta\sum_{i=1}^n x_i^2} \\
\log L(\theta; \mathbf{x}_n) &= n \log \theta - \theta \sum_{i=1}^n x_i^2 \\
\frac{\partial^2}{\partial \theta^2} \log L(\theta; \mathbf{x}_n) &= -\frac{n}{\theta^2} \\
-E\left(\frac{\partial^2}{\partial \theta^2} \log L(\theta; \mathbf{X}_n)\right) &= E\left(\frac{n}{\theta^2}\right) = \frac{n}{\theta^2}
\end{aligned}$$

quindi il limite inferiore di Cramer-Rao è $\frac{\theta^2}{n}$.

**Esercizio 11.** Si osservi che il modello considerato è una famiglia esponenziale per la quale $\sum_i^n X_i$ è statistica sufficiente e completa, stimatore non distorto di $n\theta$. Pertanto, per i teoremi di Rao-Blackwell e Lehmann-Scheffe', $\sum_i^n X_i$ è UMVUE. Si può, in alternativa, procedere utilizzando la disuguaglianza di Cramer-Rao. Indichiamo con $T = \sum_{i=1}^n X_i$, abbiamo che

$$E_\theta(T) = \sum_{i=1}^n E_\theta(X_i) = n\theta = \lambda$$

per cui $T$ è corretto. Valutiamone il MSE:

$$MSE_\theta(T) \quad = \quad V_\theta(T) = nV_\theta(X_i) = n.$$

Ora calcoliamo il limite inferiore di Cramer-Rao

$$\begin{aligned}
L(\theta; ) &= e^{-\frac{n}{2}(\bar{x}-\theta)^2} = e^{-\frac{n}{2}(\frac{T}{n}-\theta)^2} \\
L(\lambda, \mathbf{x}_n) &= e^{-\frac{n}{2}(\frac{T}{n}-\frac{\lambda}{n})^2} = e^{-\frac{n}{2n^2}(T-\lambda)^2} \\
\log L(\lambda, \mathbf{x}_n) &= -\frac{1}{2n}(T-\lambda)^2 \\
\frac{\partial^2}{\partial \lambda^2} \log L(\lambda, \mathbf{x}_n) &= -\frac{1}{n} \\
-E\left(\frac{\partial^2}{\partial \theta^2} \log L(\theta; \mathbf{X}_n)\right) &= \frac{1}{n}
\end{aligned}$$

e quindi l'estremo di Cramer-Rao è $n$ e coincide con $MSE_\theta(T)$. $T$ è UMVUE.

**Esercizio 12.** È del tutto analogo all'esercizio 9, con $q = \frac{1}{3}$. Lì avevamo l'ipotesi di normalità che però non usavamo (svolgendo i conti si verifica che il risultato è lo stesso). Quindi $T$ è non distorto, ha $MSE_\theta(T) = V_\theta(T) = \frac{\sigma^2}{9n_1} + \frac{4\sigma^2}{9n_2} \overset{n_1,n_2\to\infty}{\to} 0$ ed è consistente.

**Esercizio 13.** Ricordiamo che $E_\theta(X) = \theta$ e $V_\theta(X) = \theta(1-\theta)$. Studiamo prima $T_1$:

$$\begin{aligned}
E_\theta(T_1) &= E_\theta(X) = \theta \qquad T_1 \text{ è corretto} \\
B(T_1) &= 0 \\
MSE_\theta(T_1) &= V_\theta(T_1) = \frac{V_\theta(X)}{n} = \frac{\theta(1-\theta)}{n} \to 0 \qquad T_1 \text{ è consistente}
\end{aligned}$$

Consideriamo ora $T_2$.

$$\begin{aligned}
E_\theta(T_2) &= \frac{1+nE_\theta(X)}{n+2} = \frac{1+n\theta}{n+2} \\
B(T_2) &= \frac{1+n\theta}{n+2} - \theta = \frac{1-2\theta}{n+2} \to 0 \qquad T_2 \text{ è asintoticamente corretto} \\
MSE_\theta(T_2) &= V_\theta(T_2) + B(T_2)^2 = \frac{nV_\theta(X)}{(n+2)^2} + \frac{(1-2\theta)^2}{(n+2)^2} \to 0 \qquad T_2 \text{ è consistente.}
\end{aligned}$$

Si osservi che $E[T_2] = \theta$ per $\theta = 1/2$.

**Esercizio 14.**

$T_1 = X_{(n)}$ stimatore di massima verosimiglianza (MLE)
$T_2 = \frac{n+1}{n} X_{(n)}$ stimatore non distorto funzione di MLE (infatti $E_\theta(T_2) = \frac{n+1}{n} E_\theta(x_{(n)}) = \theta$)
$T_3 = 2\bar{X}$ stimatore dei momenti (perché $E_\theta(X) = \frac{\theta}{2}$) è non distorto.

$$
\begin{aligned}
MSE_\theta(T_1) &= V_\theta(T_1) + (E_\theta(T_1)-\theta)^2 = \frac{n\theta^2}{(n+1)^2(n+2)} + \left(\frac{-\theta}{n+1}\right)^2 = \frac{2\theta^2}{(n+1)(n+2)} \\
MSE_\theta(T_2) &= V_\theta(T_2) = \frac{(n+1)^2}{n^2}\cdot\frac{n\theta^2}{(n+1)^2(n+2)} = \frac{\theta^2}{n(n+2)} \\
MSE_\theta(T_3) &= V_\theta(T_3) = 4\cdot\frac{V_\theta(X)}{n} = \frac{\theta^2}{3n}.
\end{aligned}
$$

$$
eff(T_1/T_2) = \frac{MSE_\theta(T_2)}{MSE_\theta(T_1)} = \frac{\frac{\theta^2}{n(n+2)}}{\frac{2\theta^2}{(n+1)(n+2)}} = \frac{n+1}{2n} \le 1
$$

quindi $T_2$ è più efficiente di $T_1$

$$
eff(T_1/T_3) = \frac{MSE_\theta(T_3)}{MSE_\theta(T_1)} = \frac{\frac{\theta^2}{3n}}{\frac{2\theta^2}{(n+1)(n+2)}} = \frac{(n+1)(n+2)}{6n} \ge 1
$$

quindi $T_1$ è più efficiente di $T_3$. Di conseguenza anche $T_2$ è più efficiente di $T_3$.

Gli stimatori sono tutti e 3 consistenti perché i MSE tendono a 0 per $n \to \infty$.

Relativamente al campione osservato $T_1(\mathbf{x}_n) = 10$, $T_2(\mathbf{x}_n) = 11$ e $T_3(\mathbf{x}_n) = 8.4$.

**Esercizio 15.** Ricordiamo che $E_\theta(\bar{X}) = \theta$ e $V_\theta(\bar{X}) = \frac{\theta(1-\theta)}{n}$, per cui $T_1$ è non distorto, $MSE_\theta(T_1) = \frac{\theta(1-\theta)}{n} \to 0$ e quindi $T_1$ è consistente.

$$
\begin{aligned}
E_\theta(T_2) &= \frac{nE_\theta(\bar{x}) + \sqrt{n/4}}{n+\sqrt{n}} = \frac{n\theta + \sqrt{n/4}}{n+\sqrt{n}} \\
E_\theta(T_2) - \theta &= \frac{n\theta + \sqrt{n/4}}{n+\sqrt{n}} - \theta = \frac{\sqrt{n/4} - \theta\sqrt{n}}{n+\sqrt{n}} = \frac{\sqrt{n}(\frac{1}{2}-\theta)}{n+\sqrt{n}} \\
MSE_\theta(T_2) &= V_\theta(T_2) + (E_\theta(T_2)-\theta)^2 = \left(\frac{n}{n+\sqrt{n}}\right)^2 \frac{\theta(1-\theta)}{n} + \left(\frac{\sqrt{n}(\frac{1}{2}-\theta)}{n+\sqrt{n}}\right)^2 \\
&= \frac{n[\theta(1-\theta) + (\frac{1}{2}-\theta)^2]}{(n+\sqrt{n})^2} \to 0
\end{aligned}
$$

quindi $T_2$ è consistente.

Per $\theta = \frac{1}{2}$ si ha $E_\theta(T_2) = \theta$ (per ogni $n$, non solo se $n = 4$).

**Esercizio 16.** Per ogni $i = 1, \cdots, n$ si ha $E_\theta(X_i) = \theta$ e $V_\theta(X_i) = \sigma^2$, quindi

$$
\begin{aligned}
E_\theta(T) &= \frac{E_\theta(X_1) + \cdots + E_\theta(X_{n-1})}{n-1} + \frac{E_\theta(X_n)}{n} = \frac{(n-1)\theta}{n-1} + \frac{\theta}{n} = \theta(1+\frac{1}{n}) \\
E_\theta(T) - \theta &= \frac{\theta}{n} \\
MSE_\theta(T) &= V_\theta(T) + (E_\theta(T)-\theta)^2 = \frac{\sigma^2}{n-1} + \frac{\sigma^2}{n^2} + \frac{\theta^2}{n^2} \to 0 \qquad \text{per cui } T \text{ è consistente.}
\end{aligned}
$$

**Esercizio 17.** Caso particolare di quanto visto nell'esercizio 9, fissando rispettivamente

$$
q = 1 \text{ per } T_1 \qquad q = 0 \text{ per } T_2 \qquad q = \frac{1}{2} \text{ per } T_3 \qquad q = \frac{n_1}{n_1+n_2} \text{ per } T_4
$$

quindi tutti e quattro gli stimatori sono non distorti.

In generale $MSE_\theta(T) = \frac{q^2\sigma^2}{n_1} + \frac{(1-q)^2\sigma^2}{n_2}$, quindi

$$MSE_\theta(T_1) = \frac{\sigma^2}{n_1} \qquad MSE_\theta(T_2) = \frac{\sigma^2}{n_2} \qquad MSE_\theta(T_3) = \frac{\sigma^2}{4n_1} + \frac{\sigma^2}{4n_2} \qquad MSE_\theta(T_4) = \frac{\sigma^2}{n_1+n_2}$$

sono consistenti tutt e 4. Per l'efficiennza, $T_4$ è ovviamente più efficiente di $T_1$ e $T_2$. Quindi basta confrontare $T_4$ e $T_3$:

$$eff(T_3/T_4) = \frac{MSE_\theta(T_4)}{MSE_\theta(T_3)} = \frac{\frac{\sigma^2}{n_1+n_2}}{\frac{\sigma^2}{4n_1} + \frac{\sigma^2}{4n_2}} = \frac{4n_1 \cdot n_2}{(n_1+n_2)^2} \leq 1$$

quindi $T_4$ è più efficiente di $T_3$ e quindi il più efficiente di tutti.

**Esercizio 18.**
a)

$$B[T_a] = E[T_a] - \theta = \omega_n\theta + (1-\omega_n)a - \theta = -\theta(1-\omega_n) + (1-\omega_n)a = (1-\omega_n)(a-\theta) = \frac{1}{n+1}(a-\theta),$$

$$V[T_a] = \omega_n^2\frac{1}{n} = \frac{n^2}{(n+1)^2}\frac{1}{n} = \frac{n}{(n+1)^2}, \qquad \mathrm{MSE}[T_a] = \frac{4}{(n+1)^2}(a-\theta)^2 + \frac{n}{(n+1)^2} \to 0.$$

Quindi gli stimatori sono consistenti in media quadratica.

b)

$$T_0(\mathbf{X}_n) = \frac{n}{n+1}\ \bar{X}_n \qquad \mathrm{MSE}(T_0) = \frac{4}{(n+1)^2}\ \theta^2 + \frac{n}{(n+1)^2} \qquad \mathrm{MSE}(\bar{X}_n) = \frac{1}{n}$$

Quindi $T_0$ è migliore di $\bar{X}$ per

$$\theta^2\frac{4}{(n+1)^2} + \frac{n}{(n+1)^2} < \frac{1}{n}$$

**Esercizio 19.** a) Il modello statistico è:

$$\left((\mathbb{R}^+)^n,\ \frac{\theta^n}{[\prod_{i=1}^n(1+x_i)]^{(1+\theta)}}\prod_{i=1}^n I_{(0,+\infty)}(x_i), \mathbb{R}^+\right).$$

b)

$$L(\theta;\mathbf{x}_n) = \prod_{i=1}^n f_X(x_i;\theta) = \frac{\theta^n}{[\prod_{i=1}^n(1+x_i)]^{(1+\theta)}}\prod_{i=1}^n I_{(0,+\infty)}(x_i) = \frac{\prod_{i=1}^n I_{(0,+\infty)}(x_i)}{\prod_{i=1}^n(1+x_i)}\frac{\theta^n}{\prod_{i=1}^n(1+x_i)^\theta}.$$

da cui si evince che $L$ si fattorizza nel prodotto $h(\mathbf{x}_n)g(T(\mathbf{x}_n),\theta)$, dove il nucleo della funzione di verosim. è

$$g(T(\mathbf{x}_n),\theta) = \frac{\theta^n}{\prod_{i=1}^n(1+x_i)^\theta}$$

mentre $h(\mathbf{x}_n) = \frac{\prod_{i=1}^n I_{(0,+\infty)}(x_i)}{\prod_{i=1}^n(1+x_i)}$. Pertanto, per il teorema di fattorizzazione si ha che $T(\mathbf{x}_n) = \prod_{i=1}^n(1+x_i)$ è stat. suff. unidimensionale per il modello considerato.

c) Consideriamo la f.ne di log-verosimiglianza:

$$\ell(\theta;\mathbf{x}_n) = n\ln\theta - \theta\sum_{i=1}^n \ln(1+x_i).$$

Per trovare la stima di massima verosimiglianza consideriamo l'equazione di log-verosimiglianza:

$$\frac{d}{d\theta}\ell(\theta;\mathbf{x}_n) = 0 \qquad \text{ovvero} \qquad \frac{n}{\theta} - \sum_{i=1}^{n} \ln(1+x_i) = 0,$$

la cui unica radice è $\frac{n}{\sum_{i=1}^n \ln(1+x_i)}$. Poichè la derivata seconda di $\ell(\theta;\mathbf{x}_n)$ è $-n/\theta^2$, ovvero sempre negativa, la radice trovata è punto di massimo per $\ell$ (e per $L$) e quindi stima di massima verosimiglianza, $\hat{\theta}_{mv}(\mathbf{x}_n)$. Per il campione considerato si ha $\hat{\theta}_{mv}(\mathbf{x}_n) = 10/5 = 2$.

c) Si tratta di famiglia esponenziale. Infatti, osservando che

$$f_X(x;\theta) = \frac{1}{1+x}\exp\left(-\theta\ln(1+x) + \ln\theta\right),$$

le funzioni di densità per il modello considerato possono essere scritte nella forma generale

$$f_X(x;\theta) = h(x)\exp\left(\eta(\theta)T(x) - c(\theta)\right),$$

con

$$h(x) = \frac{1}{1+x}, \quad \eta(\theta) = -\theta, \quad T(x) = \ln(1+x), \quad c(\theta) = -\ln\theta.$$

**Esercizio 20.**

a) Si osservi che:

$$E_\theta[X] = m\theta, \qquad V_\theta[X] = m\theta(1-\theta).$$

Pertanto dall'equazione $m_1(\mathbf{X}_n) = E_\theta[X]$ discende che $\bar{X}_n = m\theta$, da cui

$$\hat{\theta}_M = \frac{\bar{X}_n}{m}.$$

b) $E_\theta[\hat{\theta}_M] = \frac{1}{m}E_\theta[\bar{X}_n] = \frac{1}{m}E_\theta[X] = \frac{m\theta}{m} = \theta, \quad \forall\theta$. Lo stimatore dei momenti è quindi non distorto ($B_\theta = 0$). Pertanto

$$MSE_\theta[\hat{\theta}_M] = V_\theta[\hat{\theta}_M] = \frac{1}{m^2}V_\theta[\bar{X}_n] = \frac{m\theta(1-\theta)}{nm^2} = \frac{\theta(1-\theta)}{nm}.$$

c) Poiché

$$\lim_{n\to+\infty} MSE_\theta[\hat{\theta}_M] = \lim_{n\to+\infty}\frac{\theta(1-\theta)}{nm} = 0$$

lo stimatore dei momenti è consistente in media quadratica.

d) Si ha che le funzioni di verosimiglianza e logverosimiglianza sono

$$L(\theta;\mathbf{x}) = \theta^{\sum_{i=1}^n x_i}(1-\theta)^{nm-\sum_{i=1}^n x_i}$$

e

$$\ell(\theta;\mathbf{x}) = (\sum_{i=1}^n x_i)\ln\theta + (nm - \sum_{i=1}^n x_i)\ln(1-\theta).$$

Derivando e risolvendo l'equazione di logveros. si ottiene che

$$\hat{\theta}_{MV} = \frac{\bar{X}_n}{m}.$$

(Si verifica facilmente che trattasi effettivamente di un punto di massimo). Lo stimatore di MV coincide con quello dei momenti e quindi i due stimatori hanno le stesse proprietà.

e) Si può rispondere al quesito in due modi.

I modo. Poiché il modello bernoulliano è una famiglia esponenziale (fatto noto, ma comunque semplicemente verificabile), uno stimatore non distorto di $\theta$, che sia anche funzione di una statistica sufficiente e completa, è UMVUE (teoremi di Rao-Blackwell e Lehmann-Scheffé). In questo caso $\sum_{i=1}^{n} X_i$ è stat. sufficiente e completa, e quindi $\bar{X}_n/m$, esssendo non distorto, è UMVUE.

II modo. E' sufficiente verificare che il limite inferiore di Cramer Rao risulta pari a $\frac{\theta(1-\theta)}{nm}$ e che quindi coincide con la varianza di $\bar{X}_n/m$.

**Esercizio 21.**

**1** – Osservando che il supporto della v.c. $X_i$ è costituito dai soli valori 0 e 1, si deduce che siamo in presenza di una variabile casuale bernoulliana con probabilità di successo

$$p = 2\theta - 1$$

per la quale è immediato verificare che

$$E[X_i] = p = 2\theta - 1$$

e che

$$\text{Var}\left[X_i\right] = p(1-p) = (2\theta - 1)(2 - 2\theta)$$

**2.** – Risolvendo in $\theta$ l'equazione che uguaglia il primo momento teorico (il valore atteso) di $X_i$ con il primo momento empirico (la media campionaria) si ottiene

$$2\theta - 1 = \bar{X} \Longleftrightarrow \theta = \frac{\bar{X}+1}{2}$$

e dunque la soluzione dell'equazione determina lo stimatore dei momenti

$$\hat{\theta}_M = \frac{\bar{X}+1}{2}$$

Dalla linearità dell'operatore valore atteso e dalla proprietà della media campionaria per cui $E[\bar{X}] = E[X_i] = 2\theta - 1$ si deduce che

$$\begin{aligned} E\left[\hat{\theta}_M\right] &= E\left[\frac{\bar{X}+1}{2}\right] = \frac{E\left[\bar{X}\right]+1}{2} \\ &= \frac{2\theta - 1 + 1}{2} = \theta \qquad \forall \theta \in \Theta \end{aligned}$$

e dunque siamo in presenza di uno stimatore non distorto per il parametro $\theta$ e quindi la distorsione

$$B_{\hat{\theta}_M}(\theta) = 0 \qquad\qquad \forall \theta \in \Theta.$$

In tal caso, essendo lo stimatore dei momenti non distorto, si ha che la sua varianza coincide con l'errore quadratico medio e si può calcolare come segue

$$\begin{aligned} MSE_{\hat{\theta}_M}(\theta) = \text{Var}_{\hat{\theta}_M}(\theta) &= \text{Var}\left[\frac{\bar{X}+1}{2}\right] \\ &= \frac{1}{4}\text{Var}\left[\bar{X}\right] = \frac{1}{4n}\text{Var}\left[X_i\right] \\ &= \frac{(2\theta-1)(2-2\theta)}{4n} \end{aligned}$$

grazie alle ben note proprietà della varianza di trasformazioni lineari, nonché alla proprietà della varianza della media campionaria, valida per tutti i modelli statistici.

Ora per verificare *se* lo stimatore è anche UMVUE sono possibili due strade alternative:

- avendo già verificato che lo stimatore è non distorto ed avendo calcolato il suo errore quadratico medio potrebbe verificarsi che la sua varianza raggiunge il LICR (Limite Inferiore di Cramer Rao). Nel caso in cui ciò si verificasse avremmo potuto dedurre la relazione richiesta. Nel caso però in cui invece il LICR non fosse ragiunto non potremmo dedurre alcunché.

- avendo già verificato che lo stimatore è non distorto si potrebbe cercare di verificare se tale stimatore è funzione di una statistica sufficiente e completa per il parametro $\theta$

La seconda strada sembra più agevole in quanto lo stimatore è funzione della statistica $S(X_1, ..., X_n) = \bar{X}$ che, in un modello bernoulliano, appartenente alla famiglia esponenziale, è noto essere una statistica sufficiente e completa per $p$. Quindi, dalle proprietà delle statistiche sufficienti e delle statistiche complete, se $S(X_1, ..., X_n)$ è sufficiente e completa per $p$ lo è anche per una funzione biunivoca del parametro $p$ nel nostro caso per

$$\theta = \frac{p+1}{2}.$$

**3.** – Dal momento che la media campionaria è consistente per $p = 2\theta - 1$

$$\bar{X} \longrightarrow p$$

sia in senso debole [convergenza in probabilità] che in senso forte [convergenza quasi certa], invocando il teorema di continuità per entrambe le forme di convergenza è immediato dedurre che una funzione continua della media aritmetica converge alla corrispondente funzione applicata nel limite $p = 2\theta - 1$ ovvero

$$\hat{\theta}_M = \frac{\bar{X}+1}{2} = g(\bar{X}) \longrightarrow g(p) = \frac{p+1}{2} = \frac{2\theta - 1 + 1}{2} = \theta \qquad \forall \theta \in \Theta.$$

Inoltre, considerando l'espressione precedente derivata dell'errore quadratico medio si ottiene che

$$MSE_{\hat{\theta}_M}(\theta) = \frac{(2\theta - 1)(2 - 2\theta)}{4n} \longrightarrow 0$$

e dunque lo stimatore è consistente anche in media quadratica.

**4.** – Avendo già osservato che il parametro usuale della distribuzione di Bernoulli $p \in (0, 1)$ è in relazione al parametro di interesse $\theta \in (\frac{1}{2}, 1)$ attraverso la relazione

$$p = 2\theta + 1$$

è vero anche il viceversa ovvero che il parametro di interesse $\theta$ è in relazione con $p$ attraverso la relazione inversa

$$\theta = g(p) = \frac{p+1}{2}$$

e dunque è immediato argomentare che dalla nota proprietà di invarianza (equivarianza) dello stimatore di massima verosimiglianza e dal fatto che $\hat{p}_{MV} = \bar{X}$ si ha che

$$\hat{\theta}_{MV} = g(\hat{p}_{MV}) = \frac{\hat{p}_{MV}+1}{2} = \frac{\bar{X}+1}{2}.$$

**Esercizio 22.** Il modello statistico-probabilistico è:

$$\left(\mathcal{X}^n = (R_+)^n, \;\; f_n(\mathbf{x}; \theta) = \frac{\prod x_i}{2^n \theta^3} \exp\left\{-\sum x_i/\theta\right\} \theta^n, \;\; \Theta = R_+\right)$$

La distribuzione delle popolazione $X$ appartiene alla famiglia esponenziale e può essere scritta come:

$$f_X(x;\theta) = \frac{x^2}{2}\exp\{-x/\theta - 3\log 2\theta\},$$

con: $h(x) = \frac{x^2}{2}$, $\psi(\theta) = -1/\theta$, $T(x) = x$, $B(\theta) = -3\log 2\theta$. Si ottiene quindi che una statistica sufficiente è data da $T(\mathbf{x}_{oss}) = \sum_{i=1}^n T(x_i) = \sum x_i$.

**2.** Le funzioni di verosimiglianza, di log-verosimiglianza e le derivate prima e seconda della funzione di log-verosimiglianza sono:

$$L_{\mathbf{x}_{oss}}(\theta) = \frac{1}{\theta^{3n}}e^{-\sum x_i/\theta}; \qquad \ell(\theta) = -3n\log\theta - \frac{\sum x_i}{\theta}; \qquad \ell'(\theta) = -\frac{3n}{\theta} + \frac{\sum x_i}{\theta^2}; \qquad \ell''(\theta) = \frac{3n}{\theta^2} - \frac{2\sum x_i}{\theta^3}.$$

Da $\ell'(\theta) = 0$, si ottiene $-3n\theta + \sum x_i = 0$ da cui $\hat\theta_{MLE} = \bar X/3$ è lo stimatore di massima verosimiglianza. Infatti

$$\ell''(\hat\theta_{MLE}) = 9(3n/\bar x_n^2) - 27(2n/\bar x_n^2) = -27n/\bar x_n^2 < 0.$$

Per la proprietà di invarianza, lo stimatore di massima verosimiglianza per $\psi$ è:

$$\hat\psi_{MLE} = \hat\theta_{MLE} + \hat\theta^2_{MLE} = \frac{\bar X_n}{3} + \frac{\bar X_n^2}{9}.$$

**3.** I momenti primo e secondo della media campionaria sono:

$$\mathbb{E}[\,\bar X_n\,] = \frac{1}{n}\sum_{i=1}^n X_i = 3\,\theta; \;\; \mathbb{E}[\,\bar X_n^2\,] = \mathbb{V}[\,\bar X_n\,] + \mathbb{E}\big([\,\bar X_n\,]\big)^2 = \frac{1}{n}\,\mathbb{V}[\bar X_n] + 9\,\theta^2 = \frac{1}{n}3\,\theta^2 + 9\,\theta^2 = 3\,\theta^2\frac{1+3n}{n}.$$

**4.** La media dello stimatore $\hat\psi_{MLE}$ è:

$$\mathbb{E}[\,\hat\psi_{MLE}\,] = \mathbb{E}[\hat\theta_{MLE} + \hat\theta^2_{MLE}\,] = \mathbb{E}\left[\frac{\bar X}{3} + \frac{\bar X^2}{9}\right] = \frac{1}{3}\,\mathbb{E}[\,\bar X_n\,] + \frac{1}{9}\,\mathbb{E}[\,\bar X_n^2\,] = \theta + \frac{1+3n}{3n}\,\theta^2 \neq \theta + \theta^2$$

Quindi lo stimatore $\hat\psi_{MLE}$ è distorto. La sua distorsione è data da:

$$D(\hat\psi_{MLE}) = \mathbb{E}[\,\hat\psi\,] - \theta - \theta^2 = \theta^2/3n,$$

e questo implica che lo stimatore è asintoticamente corretto perché la sua distorsione tende a 0 al crescere di $n$.

**Esercizio 23.**

**1.** La funzione di verosimiglianza è:

$$L_{\mathbf{x}_{oss}}(\theta) = \prod_{i=1}^n \theta^{-1} x_i^{-\frac{1}{\theta}-1} = \theta^{-n}\left(\prod_{i=1}^n x_i\right)^{-\frac{1}{\theta}-1}.$$

La funzione di log-verosimiglianza e le sue derivate sono:

$\ell(\theta) = -n\log\theta + (-\frac{1}{\theta} - 1)\sum_{i=1}^n \log x_i \qquad \ell'(\theta) = -\frac{n}{\theta} + \frac{1}{\theta^2}\sum_{i=1}^n \log x_i \qquad \ell''(\theta) = \frac{n}{\theta^2} - \frac{2\sum_{i=1}^n \log x_i}{\theta^3}$.

Si ottiene quindi: $\ell'(\theta) = 0 \Rightarrow \hat\theta_{MLE} = \frac{\sum_{i=1}^n \log x_i}{n}$, infatti $\ell''(\hat\theta_{MLE}) = \frac{n}{\hat\theta^2_{MLE}} - \frac{2\sum_{i=1}^n \log x_i}{\hat\theta^3_{MLE}} = -\frac{n^3}{(\sum_{i=1}^n \log x_i)^2} < 0$.

**2.** L'informazione osservata di Fisher è $I(\hat\theta_{MLE}) = -\ell''(\hat\theta_{MLE}) = \frac{n^3}{(\sum_{i=1}^n \log x_i)^2}$. Nell'intorno di $\hat\theta$ si può approssimare la funzione di verosimiglianza di un qualsiasi modello statistico con una distribuzione normale di media $\hat{\theta}$ e di varianza $I^{-1}(\hat{\theta})$ e l'intervallo di verosiniglianza approssimato di livello $q$ si ottiene come l'intervallo di verosimiglianza di livello $q$ di una distribuzione normale.

**Parte facoltativa.** L'intervallo di verosimiglianza di livello $q$ della distribuzione normale $L_{\mathbf{x}_{oss}}(\theta) \approx N(\hat{\theta}_{MLE}, I^{-1}(\hat{\theta}_{MLE}))$ che approssima $L_{\mathbf{x}_{oss}}(\theta)$ è dato da: $\left(\hat{\theta}_{MLE} \pm \sqrt{I^{-1}(\hat{\theta}_{MLE})}\ \ \sqrt{-2\ \log q}\right)$ quindi, sostituendo i valori trovati per $\hat{\theta}_{MLE}$ e $I(\hat{\theta}_{MLE})$, si ottiene che: $I_q \approx \left(\frac{\sum_{i=1}^n \log x_i}{n} \pm \frac{(\sum_{i=1}^n \log x_i)^2}{n^3}\sqrt{-2\ \log q}\right)$.

**3.** La media dello stimatore di massima verosimiglianza è

$$\mathbb{E}[\hat{\theta}_{MLE}] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n \log X_i\right] = \frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[\log X_i\right] = \frac{1}{n}\sum_{i=1}^n \theta = \theta.$$

E quindi $\hat{\theta}_{MLE}$ è uno stimatore corretto. Si ha che

$$MSE(\hat{\theta}_{MLE}) = \mathbb{V}[\hat{\theta}_{MLE}] = \mathbb{V}[\frac{1}{n}\sum_{i=1}^n \log X_i] = \frac{1}{n^2}\sum_{i=1}^n \mathbb{V}\left[\log X_i\right] = \frac{1}{n^2}\sum_{i=1}^n \theta^2 = \frac{\theta^2}{n}.$$

Ciò significa che $\hat{\theta}_{MLE}$ è uno stimatore consistente perché per $n \to \infty$, si ha $MSE(\hat{\theta}_{MLE}) \to 0$.

**4.** Il limite inferiore di Cramer Rao per la varianza di uno stimatore corretto è dato dall'inverso dell'informazione attesa di Fisher. L'informazione attesa di Fisher risulta essere:

$$\begin{aligned} I(\theta) \quad = \quad & -n\,\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2}\left[\log f_X(x;\theta)\right]\right] = -n\,\mathbb{E}\left\{\frac{\partial^2}{\partial\theta^2}\left[-\log\theta - \left(\frac{1}{\theta}+1)\right)\log x\right]\right\} = \\ & -n\,\mathbb{E}\left[\frac{1}{\theta^2} - \frac{2\log x}{\theta^3}\right] = -n\left[\frac{1}{\theta^2} - \frac{2\theta}{\theta^3}\right] = \frac{n}{\theta^2}. \end{aligned}$$

L'informazione attesa di Fisher è pari all'inverso della varianza di $\hat{\theta}_{MLE}$ che risulta quindi essere uno stimatore UMVUE.

**Esercizio 24. 1.**

$$\mathbb{E}[X] = \int_0^\theta x\frac{3}{\theta^3}x^2 dx = \frac{3}{\theta^3}\left[\frac{x^4}{4}\right]_0^\theta = \frac{3}{4}\theta$$

$$\mathbb{E}[X^2] = \int_0^\theta x^2\frac{3}{\theta^3}x^2 dx = \frac{3}{\theta^3}\left[\frac{x^5}{5}\right]_0^\theta = \frac{3}{5}\theta^2$$

$$\mathbb{V}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{3}{5}\theta^2 - \left(\frac{3}{4}\theta\right)^2 = \frac{3}{80}\theta^2$$

Per determinare lo stimatore dei momenti, si eguaglia la media campionaria al valore atteso:

$$\bar{X}_n = \mathbb{E}[X] \Longleftrightarrow \bar{X}_n = \frac{3}{4}\theta \Longleftrightarrow \hat{\theta}_M = \frac{4}{3}\bar{X}_n.$$

**2.**

Si verifica facilmente che lo stimatore dei momenti è non distorto, quindi l'errore quadratico medio è uguale alla varianza:

$$MSE(\hat{\theta}_M) = \mathbb{V}(\hat{\theta}_M) = \mathbb{V}(\frac{4}{3}\bar{X}_n) = \frac{16}{9}\frac{\mathbb{V}(X)}{n} = \frac{16}{9}\frac{1}{n}\frac{3}{80}\theta^2 = \frac{\theta^2}{15n}$$

Poiché l'errore quadratico medio tende a 0 per $n \to \infty$ possiamo concludere che lo stimatore dei momenti è consistente.

**3.**

Ricordando che $\bar{X}_n \approx N\left(\mathbb{E}[X], \frac{\mathbb{V}[X]}{n}\right)$ (Teorema del Limite Centrale), si ricava facilmente la distribuzione asintotica dello stimatore dei momenti:

$$\hat{\theta}_M = \frac{4}{3}\bar{X}_n \approx N\left(\frac{4}{3}\mathbb{E}[X], \frac{16}{9}\frac{\mathbb{V}[X]}{n}\right) \approx N\left(\theta, \frac{\theta^2}{15n}\right)$$

**4.**

La stima dei momenti è

$$\hat{\theta}_M = \frac{4}{3}\bar{x}_n = \frac{4}{3}\frac{\sum_{i=1}^n x_i}{n} = \frac{4}{3}\frac{3}{2}\frac{1}{36} = \frac{1}{18} = 0.056.$$

**Esercizio 25.**

**1.**

$$\mathbb{E}[X] = \int_0^\theta x\frac{1}{2\sqrt{\theta}}\frac{1}{\sqrt{x}}dx = \frac{1}{2\theta^{\frac{1}{2}}}\left[\frac{x^{\frac{3}{2}}}{\frac{3}{2}}\right]_0^\theta = \frac{\theta}{3}$$

$$\mathbb{E}[X^2] = \int_0^\theta x^2\frac{1}{2\sqrt{\theta}}\frac{1}{\sqrt{x}}dx = \frac{1}{2\theta^{\frac{1}{2}}}\left[\frac{x^{\frac{5}{2}}}{\frac{5}{2}}\right]_0^\theta = \frac{\theta^2}{5}$$

$$\mathbb{V}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{\theta^2}{5} - \left(\frac{\theta}{3}\right)^2 = \frac{4}{45}\theta^2.$$

Per determinare lo stimatore dei momenti, si eguaglia la media campionaria al valore atteso:

$$\bar{X}_n = \mathbb{E}[X] \iff \bar{X}_n = \frac{\theta}{3} \iff \hat{\theta}_M = 3\bar{X}_n.$$

**2.** Si verifica facilmente che lo stimatore dei momenti è non distorto, quindi l'errore quadratico medio è uguale alla varianza:

$$MSE(\hat{\theta}_M) = \mathbb{V}(\hat{\theta}_M) = \mathbb{V}(3\bar{X}_n) = 9\frac{\mathbb{V}(X)}{n} = 9\frac{4}{45}\frac{1}{n}\theta^2 = \frac{4\theta^2}{5n}$$

Poiché l'errore quadratico medio tende a 0 per $n \to \infty$ possiamo concludere che lo stimatore dei momenti è consistente.

**3.** Ricordando che $\bar{X}_n \approx N\left(\mathbb{E}[X], \frac{\mathbb{V}[X]}{n}\right)$ (Teorema del Limite Centrale), si ricava facilmente la distribuzione asintotica dello stimatore dei momenti:

$$\hat{\theta}_M = 3\bar{X}_n \approx N\left(3\mathbb{E}[X], 9\frac{\mathbb{V}[X]}{n}\right) \approx N\left(\theta, \frac{4\theta^2}{5n}\right).$$

**4.** La stima dei momenti

$$\hat{\theta}_M = 3\bar{x}_n = 3\frac{\sum_{i=1}^n x_i}{n} = 3\frac{1}{3}\frac{1}{36} = \frac{1}{36} = 0.028.$$

**Esercizio 26.**

1. Lo stimatore è non distorto in quanto, per le proprietà del valore atteso, si ha:

$$\mathbb{E}[T] = \frac{\mathbb{E}[\bar{X}_1] + \mathbb{E}[\bar{X}_2]}{2} = \frac{\theta_1 + \theta_2}{2} = \theta, \qquad \forall \theta \in \mathbb{R}.$$

   Dall'ipotesi di indipendenza dei due campioni discende che

$$\mathbb{V}\left[\frac{\bar{X}_1 + \bar{X}_2}{2}\right] = \frac{1}{4}(\mathbb{V}[\bar{X}_1] + \mathbb{V}[\bar{X}_2]) = \frac{1}{4}\left(\frac{1}{n_1} + \frac{1}{n_2}\right).$$

   Pertanto, essendo $T$ uno stimatore non distorto di $\theta$, si ha che

$$MSE[T] = \mathbb{V}[T] = \frac{1}{4}\left(\frac{1}{n_1} + \frac{1}{n_2}\right).$$

   Lo stimatore è consistente poichè, al divergere di $n_1$ e $n_2$, $MSE[T] \to 0, \qquad \forall \theta \in \mathbb{R}$.

2. Per le proprietà delle medie campionarie nei modelli normali, si ha che

$$\bar{X}_i | \theta_i \sim N\left(\theta_i, \frac{1}{n_i}\right), \qquad i = 1, 2.$$

   Ricordando che una combinazione lineare di v.a. normali ha distribuzione normale, abbiamo che:

$$\frac{\bar{X}_1 + \bar{X}_2}{2} \sim N\left(\theta, \frac{1}{4}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right).$$

**Esercizio 27.**

1. Per le propr. di valore atteso e varianza e dal suggerimento dato, si ha che:

$$\mathbb{E}[X_{(n)}] = \frac{n}{n+1}\theta, \qquad \mathbb{V}[X_{(n)}] = \frac{n}{(n+1)^2(n+2)}\theta^2.$$

   Si ha quindi che, $\forall \theta \in (0, 1)$:

   - $\mathbb{E}[T_1] = \theta$;
   - $\mathbb{V}[T_1] = MSE[T_1] = \frac{\theta^2}{n(n+2)}$;
   - $\mathbb{E}[T_2] = \frac{n(n+2)}{(n+1)^2}\theta$;
   - $\mathbb{V}[T_2] = \frac{n(n+2)}{(n+1)^4}\theta^2$;
   - $MSE[T_2] = (\mathbb{E}[T_2] - \theta)^2 + \mathbb{V}[T_2] = \ldots = \frac{\theta^2}{(n+1)^2}$.

   Pertanto:

   - $T_1$ è stimatore non distorto di $\theta$ e funzione di statistica sufficiente e completa, $X_{(n)}$. Si tratta quindi dello stimatore non distorto di minima varianza (UMVUE).
   - $T_1$ è consistente in errore quadratico medio, dal momento che, $\forall \theta \in (0, 1)$,

$$\lim_{n \to +\infty} MSE[T_1] = \lim_{n \to +\infty} \frac{\theta^2}{n(n+2)} = 0.$$

- Lo stimatore $T_2$ è stimatore distorto di $\theta$ con distorsione negativa e pari a

$$B[T_2] = -\frac{\theta}{(n+1)^2}.$$

Tuttavia lo stimatore risulta consistente (e dunque asintoticamente corretto), in quanto $\forall\theta \in (0,1)$,

$$\lim_{n\to+\infty} MSE[T_2] = \lim_{n\to+\infty} \frac{\theta^2}{(n+1)^2} = 0.$$

- Lo stimatore $T_2$ è più efficiente di $T_1$ poichè:

$$MSE[T_2] < MSE[T_1] \Leftrightarrow (n+1)^2 > n(n+2) \Leftrightarrow 1 > 0,$$

condizione che risulta essere ovviamente verificata per ogni valore di $n$ e di $\theta$.

Da quanto verificato si evince che, in base al criterio delle errore quadratico medio, lo stimatore $T_2$ benchè distorto, è migliore dello stimatore non distorto $T_1$ (peraltro anche UMVUE).

2. Poichè il primo momento di $X$ è $\mu_1(\theta) = \mathbb{E}[X] = \frac{\theta}{2}$ e il primo momento campionario è $m_1(\mathbf{X}_n) = \bar{X}_n$, l'equazione dei momenti($\mu_1(\theta) = m_1(\mathbf{X}_n)$) diventa:

$$\frac{\theta}{2} = \bar{X}_n,$$

da cui si ottiene lo stimatore dei momenti:

$$\hat{\theta}_M = 2\bar{X}_n.$$

Per le proprietà della media campionaria e dal momento che $X_1, \ldots, X_n$ sono v.a. i.i.d. con valore atteso e varianza finiti vale il teorema centrale di convergenza e si ha che $\bar{X}_n$ ha distribuzione asintotica normale. Pertanto, osservando che

$$\mathbb{E}[\hat{\theta}_M] = \mathbb{E}[2\bar{X}_n] = \theta, \qquad \mathbb{V}[\hat{\theta}_M] = \mathbb{V}[2\bar{X}_n] = 4\mathbb{V}[\bar{X}_n] = 4\frac{\theta^2}{12n} = \frac{\theta^2}{3n},$$

possiamo affermare che, per un qualsiasi $\theta$,

$$\frac{\hat{\theta}_M - \theta}{\sqrt{\mathbb{V}[\hat{\theta}_M]}} \approx N(0,1)$$

ovvero che

$$\hat{\theta}_M = 2\bar{X}_n \approx N\left(\theta, \frac{\theta^2}{3n}\right).$$

3. Sostituendo i valori indicati, si ottiene:

$$T_1 = \frac{21}{20} * 0.9, \quad T_2 = \frac{22}{21} * 0.9, \quad \hat{\theta}_M = 2 * 0.4 = 0.8.$$

Si noti che il dato fornito relativamente alla varianza campionaria $S_n^2$ risulta inutile ai fini della soluzione al quesito posto.

## III PARTE: INTERVALLI DI CONFIDENZA E TEST

**Esercizio 1*.** Sia $\mathbf{X}_n = (X_1, \dots, X_n)$ un campione casuale dalla popolazione $N(0, \theta)$.

a) Verificare che la statistica $S_0^2 = \sum_{i=1}^n X_i^2/n$ è uno stimatore non distorto di $\theta$.

b) Determinare la generica espressione di un intervallo di confidenza per $\theta$, in funzione della statistica $S_0^2$.

c) In un campione di dimensione $n = 15$ si è osservato che lo stimatore $S_0^2$ assume il valore $3.23^2$. Determinare l'intervallo di confidenza per $\theta$ al 95%.

**Esercizio 2.** In una fabbrica di generi alimentari si vuole determinare il valore medio di "grasso totale" (in grammi) in una confezione regolare di patatine. Si analizzano $n = 101$ confezioni e si ottengono i seguenti risultati: $\bar{x}_n = 18.2\ g$, $s_n^2 = 0.56\ g^2$. Assumendo che le osservazioni ottenute siano i valori osservati di un campione casuale da una popolazione normale di media $\mu$ e varianza $\sigma^2$ entrambe incognite, determinare:

a) l'intervallo di confidenza al 90% per $\mu$;

b) l'intervallo di confidenza al 90% per $\sigma^2$.

**Esercizio 3*.** Un campione di 100 transistor viene estratto da una grossa fornitura e sottoposto a controllo di qualità. Il risultato è che 80 pezzi superano il controllo. Si determini un intervallo di confidenza al 95% per la percentuale $p$ di transistor idonei nel lotto considerato.

**Esercizio 4*.** In un processo di controllo di qualità emerge che, su 371 pezzi controllati, 18 sono difettosi. Trattando il campione considerato come un campione casuale,

a) determinare la funzione di verosimiglianza e calcolare il valore della stima di massima verosimiglianza di $p$, proporzione dei pezzi difettosi nella popolazione da cui proviene il campione;

b) determinare un intervallo di confidenza al 90% approssimato per il parametro $p$;

c) stabilire, alla luce dei dati osservati, se risulta più verosimile per $p$ il valore 0.07 oppure il valore 0.03?

**Esercizio 5*.** Si consideri un campione casuale $\mathbf{X}_n = (X_1, \dots, X_n)$ da una popolazione bernoulliana di parametro incognito, $p$, e le due ipotesi

$$H_0 : \ p = 0.2 \qquad H_1 : \ p = 0.4.$$

a) Verificare che la generica regione di rifiuto del test basato sul rapporto delle verosimiglianze risulta essere:

$$R = \{\mathbf{x}_n : \ y = \sum_{i=1}^n x_i \geq k\}.$$

b) Determinare il valore della probabilità di errore di prima specie, $\alpha$, che si ottiene ponendo $n = 4$ e $k = 2$.

**Esercizio 6.** Sia $\mathbf{X}_n = (X_1, \dots, X_n)$ un campione casuale dalla popolazione con distribuzione di probabilità

$$f_X(x; \theta, \alpha) = \frac{\Gamma(\alpha + x)}{\Gamma(x+1)\Gamma(\alpha)}\, \theta^\alpha (1-\theta)^x, \qquad x = 0, 1, 2, \dots, \quad \theta \in (0,1), \quad \alpha > 0.$$

Si supponga che $\alpha = 1$ e si consideri il campione $\mathbf{x}_n = (3, 2, 7, 3, 5)$.

a) Determinare la funzione di verosimiglianza di $\theta$ per il campione considerato.

b) Verificare se, alla luce dei dati osservati, risulta più verosimile per $\theta$ il valore $\theta_1 = 0.3$ oppure il valore $\theta_2 = 0.6$.

**Esercizio 7*.** Viene estratto un campione casuale di dimensione $n = 16$ da una popolazione normale di parametri incogniti $\mu$ e $\sigma$. Sapendo che la media campionaria è $\bar{x}_n = 27.9$ e che la varianza campionaria è $s_n^2 = 3.23^2$, determinare l'intervallo di confidenza al 90% per i parametri incogniti.

**Esercizio 8*.** Il seguente campione si suppone proveniente da una popolazione con distribuzione normale con varianza $\sigma = 16$ e valore atteso $\mu$ incognito:

$$23,\ 32,\ 22,\ 31,\ 27,\ 25,\ 21,\ 24,\ 20,\ 18.$$

a) Sottoporre a verica le seguenti ipotesi:

$$H_0: \ \mu = 20 \qquad H_1: \ \mu > 20$$

al livello $\alpha = 0.05$.

b) Supponendo che il valore vero di $\mu$ sia pari a 25, calcolare la potenza del test.

**Esercizio 9*.** Si consideri un campione casuale $\mathbf{X}_n = (X_1, \ldots, X_n)$ da una popolazione con funzione di densità

$$f_X(x;\theta) = \theta e^{-\theta x}, \qquad x > 0, \quad \theta > 0.$$

Si considerino, per il parametro incognito $\theta$, le ipotesi

$$H_0: \ \theta = 1 \qquad H_1: \ \theta = 2.$$

a) Verificare che la generica regione di rifiuto del test basato sul rapporto delle verosimiglianze risulta essere:

$$R = \{\mathbf{x}_n : \ n\bar{x} = \sum_{i=1}^{n} x_i < k\}.$$

b) Determinare il valore della probabilità di errore di prima specie, $\alpha$, che si ottiene ponendo $n = 36$ e $k = 2/3$. [Sugg.: Si ricordi che la somma di v.a. esponenziali indipendenti ha distribuzione gamma. In alternativa, utilizzare l'approssimazione normale, ricordando che $E_\theta(X) = 1/\theta$ e che $\mathrm{V}_\theta(X) = 1/\theta^2$. ].

**Esercizio 10*.** Si suppone che il tempo di azione di un anestetico sia una variabile aleatoria normale di valore atteso $\mu$ e varianza $\sigma^2$, entrambi incogniti. In un esperimento, i valori osservati per un campione casuale di dimensione $n = 10$ della media e della varianza campionaria sono pari a $\bar{x} = 9.28$ min e $s_n^2 = 0.1619^2$ min$^2$.

a) Determinare un intervallo di confidenza all' 80% per il parametro incognito $\mu$.

b) Si sottoponga a verifica il seguente sistema di ipotesi, assumendo un livello di significatività $\alpha = 0.01$:

$$H_0: \ \mu = \mu_0 = 10.5 \text{ min}$$
$$H_1: \ \mu < \mu_0 = 10.5 \text{ min}$$

Ripetere assumendo $\alpha = 0.025$.

c) Determinare graficamente il valore $p$ per il test di cui al punto precedente.

**Esercizio 11*.** Si suppone che la dimensione delle ali di un insetto tropicale sia una variabile aleatoria normale di valore atteso $\mu$ e varianza $\sigma^2$, entrambi incogniti. In un esperimento, i valori osservati per un campione casuale di dimensione $n = 20$ della media e della varianza campionaria sono pari a $\bar{x} = 3.23$ mm e $s_n^2 = 0.214^2$ mm$^2$.

a) Determinare un intervallo di confidenza al 90% per il parametro incognito $\mu$.

b) Si sottoponga a verifica il seguente sistema di ipotesi, assumendo un livello di significatività $\alpha = 0.05$:

$$H_0: \ \mu = \mu_0 = 3.14 \text{ mm}$$

$$H_1: \ \mu > \mu_0 = 3.14 \text{ mm}$$

c) Determinare graficamente il valore $p$ per il test di cui al punto precedente.

**Esercizio 12.** Si suppone che la dimensione delle ali di un insetto tropicale sia una variabile aleatoria normale di valore atteso $\mu$ e varianza $\sigma^2$, entrambi incogniti. In un esperimento, i valori osservati per un campione casuale di dimensione $n = 20$ della media e della varianza campionaria sono pari a $\bar{x} = 3.23$ mm e $S^2 = 0.214^2$ mm$^2$.

Costruire con il metodo del rapporto delle massime verosimiglianze il test per la verifica del seguente sistema di ipotesi, assumendo un livello di significatività $\alpha = 0.05$:

$$H_0: \ \mu = \mu_0 = 3.14 \text{ mm}$$

$$H_1: \ \mu \neq \mu_0 = 3.14 \text{ mm}$$

**Esercizio 13*.** Sia $X_1, \ldots, X_n$ un campione casuale di dimensione $n$ dalla popolazione di Poisson di parametro incognito $\theta$, con distribuzione di probabilità:

$$f_X(x; \theta) = e^{-\theta} \, \frac{\theta^x}{x!}, \qquad x = 0, 1, 2, \ldots \quad \theta > 0.$$

Si consideri il sistema di ipotesi:

$$H_0: \ \theta = \theta_0 \qquad H_1: \ \theta = \theta_1.$$

Assumendo $\theta_0 > \theta_1$, la regione di accettazione del test basato sul rapporto delle verosimiglianze risulta essere

$$A = \{\mathbf{x}_n : \ \bar{x}_n > k\}.$$

a) Verificare che, per $n$ sufficientemente elevato, la variabile aleatoria $\bar{X}_n$ ha, approssimativamente, distribuzione $N(\theta, \theta/n)$.

b) Calcolare la potenza del test, $1 - \beta$, che si ottiene ponendo $\theta_1 = 5$, $n = 25$ e $k = 5.5$.

(**Sugg.**: utilizzare l'approssimazione normale per la distribuzione di $\bar{X}_n$)

c) Verificare che, effettivamente, la regione (1) definisce la regione di accettazione del test di Neyman-Pearson.

**Esercizio 14*.** Sia $X_1, \ldots X_n$ un campione casuale da una popolazione uniforme in $(0, \theta)$, $\theta > 0$. Si consideri la statistica campionaria $X_{(n)} = \max\{X_1, \ldots, X_n\}$, per la quale è noto che, $\forall \theta > 0$,

$$E[X_{(n)}] = \frac{n}{n+1} \, \theta \qquad \text{e} \qquad V[X_{(n)}] = \frac{n}{(n+1)^2(n+2)} \, \theta^2.$$

a) Sapendo che un intervallo di confidenza di livello $1-\alpha$ per il parametro $\theta$ è fornito dal seguente intervallo aleatorio:

$$IC(\mathbf{X}_n) = \left( X_{(n)}, \frac{1}{\alpha^{1/n}} \, X_{(n)} \right),$$

verificare che il valore atteso della lunghezza aleatoria di $IC(\mathbf{X}_n)$, indicata con $\mathcal{L}_\alpha(\mathbf{X}_n)$, risulta essere

$$E[\mathcal{L}_\alpha(\mathbf{X}_n)] = \frac{n}{n+1} \left[ \frac{1-\alpha^{1/n}}{\alpha^{1/n}} \right] \theta.$$

b) Supponendo di avere osservato il campione di dati $\mathbf{x}_n = (1, 9, 3, 4, 5, 3, 2, 0, 10, 5)$, determinare la stima per intervallo di $\theta$ (assumere $1-\alpha = 0.95$).

**Esercizio 15*.** Sia $X_1, \ldots, X_n$ un campione casuale di dimensione $n$ da una popolazione bernoulliana di parametro incognito $\theta$, con distribuzione di probabilità:

$$f_X(x;\theta) = \theta^x (1-\theta)^{1-x}, \qquad x = 0, 1 \qquad \theta \in (0,1) > 0.$$

Si consideri il sistema di ipotesi:

$$H_0 : \; \theta = \theta_0 = 0.5 \qquad H_1 : \; \theta = \theta_1 = 0.35.$$

La regione di rifiuto del test basato sul rapporto delle verosimiglianze risulta essere

$$R = \{\mathbf{x}_n : \; \sum_{i=1}^{n} X_i < k\}. \tag{1}$$

a) Verificare che, per $n$ sufficientemente elevato, la variabile aleatoria $\sum_{i=1}^n X_i$ ha, approssimativamente, distribuzione $N(n\theta, n\theta(1-\theta))$.

b) Calcolare le probabilità di errore di I e di II specie che si ottengono ponendo $n = 144$ e $k = 60$. (**Sugg.**: utilizzare l'approssimazione normale per la distribuzione di $\sum_{i=1}^n X_i$)

c) Verificare che, effettivamente, la regione (1) definisce la regione di rifiuto del test di Neyman-Pearson.

**Esercizio 16*.** Si considerino due campioni indipendenti di numerosità $n_1 = 14$ e $n_2 = 12$, estratti da due popolazioni normali di varianze note, $\sigma_1^2 = 40$ e $\sigma_2^2 = 100$ e valori attesi incogniti $\mu_1$ e $\mu_2$. Nel primo campione il valore osservato della media campionaria è pari a 44, nel secondo è pari a 50. Calcolare un intervallo di confidenza con livello di confidenza pari a 0.95 per la differenza $\mu_1 - \mu_2$. In base al risultato ottenuto è possibile escludere che $\mu_1$ e $\mu_2$ siano uguali tra loro?

**Esercizio 17*.** Sia $X_1, \ldots, X_n$ un campione casuale da una popolazione esponenziale di parametro $\theta$, per la quale $E[X_i] = \theta$ e $V[X_i] = \theta^2$. Si consideri il seguente intervallo di confidenza di livello $1-\alpha$ per $\theta$:

$$\left( \frac{2\sum_{i=1}^n X_i}{\chi^2_{1-\frac{\alpha}{2},2n}}, \frac{2\sum_{i=1}^n X_i}{\chi^2_{\frac{\alpha}{2},2n}}. \right)$$

a) Per la v.a. $L_\alpha(\mathbf{X}_n)$, lunghezza dell'intervallo considerato, determinare il valore atteso e la varianza.

b) Supponendo di avere osservato un campione di $n = 10$ osservazioni dalla popolazione considerata, in cui la somma campionaria è pari a 1740, calcolare gli estremi dell'intervallo di confidenza di livello 0.95 e la lunghezza osservata.

**Esercizio 18*.** Si suppone che, in una città, il consumo di acqua al giorno per abitazione sia una v.a. normale di media $\mu$ e varianza $\sigma^2$ incogniti. Per un campione casuale di dimensione $n = 20$ si è rilevato che $\bar{x}_n = 353.8$ galloni e $S = 21.85$ galloni. Sottoporre a verifica le seguenti ipotesi, ponendo $\alpha = 0.05$:

$$H_0 : \ \mu = 350 \qquad H_1 : \ \mu \neq 350.$$

Si calcoli infine il $p$-value, lo si rappresenti graficamente e si commenti il risultato.

**Esercizio 19*.** Si suppone che l'altezza dei cittadini maschi di un centro urbano sia una v.a. normale di media $\mu$ e varianza $\sigma^2 = 300$ cm$^2$. Per un campione casuale di dimensione $n = 20$ si è rilevato che $\bar{x}_n = 177.5$ cm.

a) Si calcoli un intervallo di confidenza di livello $1 - \alpha = 0.95$ per $\mu$.

b) Sottoporre a verifica le seguenti ipotesi, ponendo $\alpha = 0.05$:

$$H_0 : \ \mu = 175 \qquad H_1 : \ \mu > 175.$$

c) Si calcoli infine il $p$-value, lo si rappresenti graficamente e si commenti il risultato.

**Esercizio 20*.** Una casa farmaceutica è interessata a stabilire la riduzione di peso corporeo in grammi legata alla somministrazione di un determinato farmaco. Supponendo che la "riduzione di peso" sia una variabie aleatoria con varianza pari a $900^2$ $g^2$, determinare il numero di individui che devono essere considerati nello studio, al fine di ottenere un intervallo di confidenza di livello $\alpha = 0.01$, la cui semi-lunghezza non sia superiore a 200 g.

**Esercizio 21*.** Sia $X_1, \ldots, X_n$ un campione casuale di dimensione $n = 20$ da una popolazione $N(\mu, \sigma^2)$ con $\sigma^2 = 300$. Per il sistema di ipotesi

$$H_0 : \ \mu = \mu_0 = 175$$

$$H_1 : \ \mu > \mu_0,$$

a) determinare la funzione di potenza del test di livello $1 - \alpha$;

b) calcolare il valore della funzione di potenza in $\mu = 178$, ponendo $1 - \alpha = 0.95$.

**Esercizio 22*.** Si ipotizza che il tempo medio di attesa dei clienti di una banca nell'ora di punta sia una variabile aleatoria normale di varianza nota pari a 4. Per un campione casuale di 16 clienti si è rilevato un tempo medio di attesa di 5.5 minuti.

a) Sottoporre a verifica l'ipotesi che il tempo medio di attesa $\theta$ sia inferiore 5 minuti contro l'ipotesi che $\theta$ sia maggiore di 5 minuti, ad un livello $\alpha = 0.01$ e rappresentare graficamente la regione di rifiuto del test.

b) Determinare il valore $p$ e rappresentarlo graficamente.

**Esercizio 23*.** Sia $X_1, \ldots, X_n$ un campione casuale da una popolazione di Poisson di parametro $\theta$. Per la quantità $g(\theta) = e^{-\theta}$:

a) determinare lo stimatore di massima verosimiglianza, $\widehat{g(\theta)}$;

b) verificare che

$$\widehat{g(\theta)} \ \dot{\sim} \ \mathrm{N}(e^{-\theta}, \mathrm{V}[\widehat{g(\theta)}]), \qquad \text{dove} \qquad \mathrm{V}[\widehat{g(\theta)}] = e^{-2\theta}\frac{\theta}{n};$$

c) determinare una stima per $V[\widehat{g(\theta)}]$, varianza asintotica di $\widehat{g(\theta)}$;

d) determinare un intervallo di confidenza approssimato al 95%.;

e) determinare la stima di massima verosimiglianza e gli estremi dell'intervallo di confidenza al 95% per il seguente campione osservato:

$$x_1 = \ldots = x_5 = 2, \quad x_6 = \ldots = x_{10} = 1, \quad x_{11} = \ldots = x_{15} = 3, \quad x_{16} = \ldots = x_{25} = 4.$$

**Esercizio 24*.** Un prodotto viene commercializzato con due confezioni diverse, A e B. Si assume che, in un periodo di riferimento, il ricavo delle vendite (in migliaia di euro) del prodotto con confezione $j$ $(j = A, B)$ sia una v.a. normale di parametri $(\theta_j, \sigma^2)$. Si considerano due campioni casuali e indipendenti tra loro di ricavi relativi a $n = 9$ punti vendita. Indichiamo con

$$X_1^A, \ldots X_i^A, \ldots X_n^A \qquad \text{e} \qquad X_1^B, \ldots X_i^B, \ldots X_n^B$$

i due campioni casuali considerati. I valori osservati dei ricavi nei due campioni sono riportati nella seguente tabella.

| Punto Vendita | $x_i^A$ | $x_i^B$ |
|---|---|---|
| 1 | 1.72 | 1.17 |
| 2 | 0.50 | 1.73 |
| 3 | 1.01 | 1.42 |
| 4 | 1.14 | 2.20 |
| 5 | 1.13 | 1.21 |
| 6 | 1.55 | 1.11 |
| 7 | 2.32 | 0.84 |
| 8 | 0.71 | 1.51 |
| 9 | 0.94 | 1.75 |

a) Determinare un intervallo di confidenza al 95% per il parametro incognito $\Delta = \theta_A - \theta_B$, assumendo che $\sigma^2 = 0.16$. Sulla base dell'intervallo che si ottiene con i dati assegnati, è lecito supporre che il diverso confezionamento non determina una differenza statisticamente significativa nei ricavi?

b) Si sottoponga a verifica il seguente sistema di ipotesi, assumendo un livello di significatività pari a $\alpha = 0.05$:

$$H_0 : \ \Delta = 0$$
$$H_1 : \ \Delta \neq 0$$

e commentare l'esito del test.

**Esercizio 25*.** Sia $X_1, \ldots, X_n$ un campione casuale di dimensione $n = 3$ proveniente da una popolazione di Poisson di parametro incognito $\theta$, con distribuzione di probabilità:

$$f_X(x; \theta) = e^{-\theta} \frac{\theta^x}{x!}, \qquad x = 0, 1, 2, \ldots \quad \theta > 0.$$

Si consideri il sistema di ipotesi:

$$H_0 : \ \theta = \theta_0 = 2 \qquad H_1 : \ \theta = \theta_1 = 1.$$

a) Verificare che la regione di rifiuto del test basato sul rapporto delle verosimiglianze risulta essere

$$R = \{\mathbf{x}_n : \ x_1 + x_2 + x_3 \leq k\}. \tag{2}$$

b) Calcolare la probabilità di errore di I specie, $\alpha$, e la potenza del test, $1-\beta$, che si ottengono ponendo $k=1$.

[**Sugg.** Ricordare che se $X_1, \ldots, X_n$ sono $n$ v.a. i.i.d. distribuite secondo la legge di Poisson di parametro $\theta$, la loro somma è una v.a. di Poisson di parametro $n\theta$.]

**Esercizio 26*.** Si consideri un campione casuale di dimensione $n$ proveniente da una popolazione bernoulliana di parametro $\theta$ incognito. Siamo interessati a effettuare inferenza sul logaritmo naturale della varianza di tale popolazione, ovvero sul parametro $g(\theta) = \ln[\theta(1-\theta)]$.

a) Determinare lo stimatore di massima verosimiglianza di $g(\theta)$, $\widehat{g(\theta)}$.

b) Determinare la varianza asintotica di $\widehat{g(\theta)}$ e uno stimatore di tale quantità.

c) Determinare l'approssimazione normale della distribuzione campionaria dello stimatore di massima verosimiglianza di $g(\theta)$ e l'espressione generica di un intervallo di confidenza approssimato per $g(\theta)$.

d) Supponendo di avere osservato in un campione di $n = 100$ osservazioni un valore della media campionaria pari a 0.6, determinare un intervallo di confidenza al 95% per $g(\theta)$.

**Esercizio 27*.** Dato un campione casuale di $n$ osservazioni estratto da una popolazione $X$ con funzione di densità:

$$f_X(x;\theta) = \theta\,(1-x)^{\theta-1}, \qquad x \in (0,1), \qquad \theta > 0,$$

lo stimatore di massima verosimiglianza per $\theta$ è $\hat{\theta}_{MLE} = -\frac{n}{\sum_{i=1}^{n}\log(1-X_i)}$.

1. Si utilizzi il test di Neyman-Pearson per il confronto tra le ipotesi

$$H_0 : \theta = \theta_0 \qquad H_1 : \theta = \theta_1 \qquad (\theta_0 < \theta_1)$$

e si mostri che la regione di accettazione è

$$RA = \{\mathbf{X} \in \mathcal{X}^n : \hat{\theta}_{\text{MLE}} < K\}. \tag{3}$$

2. Verificare che asintoticamente si ha $\hat{\theta}_{MLE} \approx N(\theta, \frac{\theta^2}{n})$.

3. Ottenere l'espressione dell'intervallo di confidenza approssimato per $\theta$ di livello 0.80 e calcolarne gli estremi supponendo di avere osservato un campione di $n = 225$ unità, per il quale $\hat{\theta}_{MLE} = 1.5$.

4. Si consideri il campione osservato di cui al punto precedente. Si vogliono confrontare le ipotesi:

$$H_0 : \theta_0 = 1.3 \qquad H_1 : \theta_1 = 1.6.$$

Calcolare il valore approssimato della potenza del test (1), ottenuto ponendo $K = 1.464$.

**Esercizio 28*.** Si consideri un campione casuale di $n$ osservazioni estratte da una popolazione $X$ con distribuzione di densità di probabilità:

$$f_X(x;\theta) = (\theta+2)\,x^{\theta+1}, \qquad 0 \le x \le 1, \qquad \theta > -2.$$

1. Verificare che la distribuzione campionaria asintotica di $\hat{\theta}_{MLE}$ è $\hat{\theta}_{MLE} \approx N\left(\theta, \dfrac{(\theta+2)^2}{n}\right)$.

**Parte facoltativa.** È necessario conoscere esplicitamente l'espressione di $\hat{\theta}_{MLE}$? Spiegare perchè.

2. Ottenere un intervallo di confidenza approssimato di livello 0.95 per $\theta$ assumendo che in un campione osservato di numerosità $n = 144$, il valore della stima di massima verosimiglianza ottenuto sia $\hat{\theta}_{MLE} = 1.6$.

3. Per il confronto tra le ipotesi

$$H_0 : \theta = \theta_0 \text{ vs } H_1 : \theta < \theta_0,$$

si consideri il test con la seguente regione di rifiuto

$$RR = \{\mathbf{x} \in \mathcal{X}^n : \hat{\theta}_{MLE} < k\}$$

(test di Wald). Utilizzando l'approssimazione normale per $\hat{\theta}_{\text{MLE}}$, si verifichi che la regione di rifiuto per un test con probabilità di errore di prima specie pari ad $\alpha$ si ottiene ponendo $k = z_\alpha \dfrac{\hat{\theta}_{MLE} + 2}{\sqrt{n}} + \theta_0$, dove $z_\alpha$ indica il percentile di livello $\alpha$ della distribuzione normale standardizzata.

4. Per il campione osservato della domanda **2**, si esegua il test per $\theta_0 = 1.5$ a livello $\alpha = 0.05$.

**Esercizio 29*.** La tabella seguente riporta i punteggi medi ($\bar{x}$ e $\bar{y}$) riportati da due docenti (A e B) valutati da due campioni distinti di studenti, di dimensione rispettivamente pari a $n = 229$ e $m = 243$. Sono anche riportati gli scarti quadratici medi dei punteggi nei due gruppi ($s_x$ e $s_y$).

| A | B |
|---|---|
| $n = 229$ | $m = 243$ |
| $\bar{x}_n = 2.14$ | $\bar{y} = 4.21$ |
| $s_x = 0.94$ | $s_y = 0.83$ |

Si supponga che i due campioni siano indipendenti e provenienti da popolazioni con uguale varianza, $\sigma^2$.

1. Determinare la stima congiunta di $\sigma^2$ basata sui dati campionari disponibili.

2. Determinare stima puntuale e stima per intervallo (al 95 %) per il parametro "differenza tra i punteggi dei due docenti".

3. Verificare se è più plausibile l'ipotesi che i punteggi siano uguali o se il punteggio del primo docente risulta inferiore a quello del secondo.

4. Stabilire se, nel problema trattato, è necessario ipotizzare che i dati siano realizzazioni di variabili aleatorie normali.

**Esercizio 30*.** [3] Sia $X_1, \ldots, X_n$ un campione casuale proveniente dalla popolazione con funzione di densità di probabilità

$$f_X(x;\theta) = \frac{3}{\theta^3}x^2, \qquad 0 < x < \theta, \qquad \theta > 0.$$

1. Verificare che

$$\mathbb{E}[X] = \frac{3}{4}\theta, \qquad \mathbb{V}[X] = \frac{3}{80}\theta^2$$

e che lo stimatore dei momenti di $\theta$ è

$$\hat{\theta}_M = \frac{4}{3}\,\bar{X}_n$$

[3] Parte di questo esercizio è stato proposto nella sezione sulle proprietà degli stimatori.

2. Determinare l'errore quadratico medio dello stimatore dei momenti e studiarne la consistenza.

3. Determinare la distribuzione asintotica dello stimatore $\hat{\theta}_M$ e l'intervallo di confidenza asintotico per $\theta$, assumendo come stima di $\mathbb{V}[X]$ la quantità ottenuta sostituendo a $\theta$ lo stimatore dei momenti.

4. Determinare la stima dei momenti e l'intervallo di confidenza asintotico di livello $1-\alpha = 0.95$ supponendo di avere osservato un campione di dimensione $n = 36$ per il quale $\sum_{i=1}^{n} x_i = \frac{3}{2}$.

**Esercizio 31*.** Sia $X_1, \ldots, X_n$ un campione casuale proveniente dalla popolazione con funzione di densità di probabilità

$$f_X(x;\theta) = \theta^2 x e^{-\theta x}, \qquad x > 0, \qquad \theta > 0,$$

con

$$\mathbb{E}[X] = \frac{2}{\theta} \qquad\qquad \mathbb{V}[X] = \frac{2}{\theta^2}.$$

Si consideri il sistema di ipotesi

$$H_0: \quad \theta = \theta_0, \qquad H_1: \quad \theta = \theta_1, \qquad\qquad (\theta_0 > \theta_1).$$

1. Verificare che la regione di accettazione del test di Neyman-Pearson risulta essere l'insieme

$$A = \{\mathbf{x} \in \mathcal{X}^n : \quad \bar{x}_n < k\}, \qquad k > 0$$

2. Determinare la distribuzione asintotica di $\bar{X}_n$.

3. Verificare che, utilizzando la distribuzione asintotica di $\bar{X}_n$, il valore di $k$ per il quale la probabilità di errore di I specie del test (vedi il punto 1.) è pari ad $\alpha$ è

$$k_\alpha = \frac{2}{\theta_0} + \sqrt{\frac{2}{n\theta_0^2}} z_{1-\alpha}.$$

   Determinare il valore di $k_\alpha$ per $\alpha = 0.05$, $\theta_0 = 2$ e $n = 25$.

4. Determinare la potenza del test considerato per $\theta_1 = 1$.

5. Stabilire se, per un campione osservato di dimensione $n = 25$ e con media campionaria pari a 1, l'ipotesi nulla viene accettata o rifiutata.

**Esercizio 32*.** [4] Sia $X_1, \ldots, X_n$ un campione casuale proveniente dalla popolazione con funzione di densità di probabilità

$$f_X(x;\theta) = \frac{1}{2\sqrt{\theta}} \frac{1}{\sqrt{x}}, \qquad 0 < x < \theta, \qquad \theta > 0.$$

1. Verificare che

$$\mathbb{E}[X] = \frac{1}{3}\theta, \qquad \mathbb{V}[X] = \frac{4}{45}\theta^2$$

   e che lo stimatore dei momenti di $\theta$ è

$$\hat{\theta}_M = 3\bar{X}_n$$

2. Determinare l'errore quadratico medio dello stimatore dei momenti e studiarne la consistenza.

[4] Parte di questo esercizio è stato proposto nella sezione sulle proprietà degli stimatori.

3. Determinare la distribuzione asintotica dello stimatore $\hat{\theta}_M$ e l'intervallo di confidenza asintotica per $\theta$, assumendo come stima di $\mathbb{V}[X]$ la quantità ottenuta sostituendo a $\theta$ lo stimatore dei momenti.

4. Determinare la stima dei momenti e l'intervallo di confidenza asintotico di livello $1-\alpha = 0.95$ supponendo di avere osservato un campione di dimensione $n = 36$ per il quale $\sum_{i=1}^{n} x_i = \frac{1}{3}$.

**Esercizio 33*.** Sia $X_1, \ldots, X_n$ un campione casuale proveniente dalla popolazione con funzione di densità di probabilità

$$f_X(x;\theta) = \theta e^{-\theta x}, \qquad x > 0, \qquad \theta > 0, \quad \text{dove} \quad \mathbb{E}[X] = \frac{1}{\theta}, \quad \mathbb{V}[X] = \frac{1}{\theta^2}.$$

Si consideri il sistema di ipotesi

$$H_0: \quad \theta = \theta_0, \qquad H_1: \quad \theta = \theta_1, \qquad\qquad (\theta_0 > \theta_1).$$

1. Verificare che la regione di accettazione del test di Neyman-Pearson risulta essere l'insieme

$$A = \{\mathbf{x} \in \mathcal{X}^n : \quad \bar{x}_n < k\}, \qquad k > 0$$

2. Determinare la distribuzione asintotica di $\bar{X}_n$.

3. Verificare che, utilizzando la distribuzione asintotica di $\bar{X}_n$, il valore di $k$ per il quale la probabilit di errore di I specie del test (vedi il punto 1.) è pari ad $\alpha$ è

$$k_\alpha = \frac{1}{\theta_0} + \sqrt{\frac{1}{n\theta_0^2}} z_{1-\alpha}.$$

Determinare il valore di $k_\alpha$ per $\alpha = 0.05$, $\theta_0 = 2$ e $n = 25$.

4. Determinare la potenza del test considerato, assumendo $\theta_1 = 1$.

5. Stabilire se, per un campione osservato di dimensione $n = 25$ e con media campionaria pari a 1.5, l'ipotesi nulla viene accettata o rifiutata.

**Esercizio 34*.** [5] Siano $X_1^1, \ldots, X_{n_1}^1$ e $X_1^2, \ldots, X_{n_2}^2$ due campioni casuali indipendenti, rispettivamente di ampiezza $n_1$ e $n_2$, provenienti da distribuzioni $N(\theta_1, 1)$ e $N(\theta_2, 1)$. Si consideri il parametro incognito

$$\theta = \frac{\theta_1 + \theta_2}{2}$$

e lo stimatore

$$T(\mathbf{X}_n) = \frac{\bar{X}_1 + \bar{X}_2}{2},$$

dove $\bar{X}_1$ e $\bar{X}_2$ sono le medie campionarie dei due campioni.

1. Verificare che lo stimatore $T(\mathbf{X}_n)$ è non distorto e consistente per il parametro $\theta$.

2. Determinare la distribuzione campionaria di $T(\mathbf{X}_n)$.

3. Sulla base del precedente risultato, determinare un intervallo di confidenza di livello 0.95 per il parametro $\theta$ avendo a disposizione due campioni di dimensione $n_1 = n_2 = 10$ per i quali si ha $\bar{x}_1 = 2$ e $\bar{x}_2 = 3$.

[5] Parte di questo esercizio è stato proposto nella sezione sulle proprietà degli stimatori.

**Esercizio 35*.** [6] Sia $X_1, \dots, X_n$ un campione casuale proveniente da una distribuzione uniforme nell'intervallo $[0, \theta]$.

1. Si confrontino e si discutano le proprietà inferenziali dei seguenti due stimatori per campioni di ampiezza $n$ fissata:

$$\begin{aligned} T_1(\mathbf{X}_n) &= \frac{n+1}{n} X_{(n)}, \\ T_2(\mathbf{X}_n) &= \frac{n+2}{n+1} X_{(n)}. \end{aligned}$$

   *Suggerimento*: si ricordi che

$$\begin{aligned} \mathbb{E}\left[\frac{X_{(n)}}{\theta}\right] &= \frac{n}{n+1} \\ \mathbb{V}\left[\frac{X_{(n)}}{\theta}\right] &= \frac{n}{(n+1)^2(n+2)} \end{aligned}$$

2. Determinare $T_3(\mathbf{X}_n) = \hat{\theta}_M$ lo stimatore dei momenti di $\theta$ e la sua distribuzione asintotica. Attraverso quest'ultima si ottenga il corrispondente intervallo di confidenza asintotico per il parametro $\theta$ al generico livello $1 - \alpha$.

3. Dato un campione osservato di dimensione $n = 20$ in cui

$$x_{(1)} = 0.2, \qquad \bar{x}_n = 0.4, \qquad S_n^2 = 0.1, \qquad x_{(n)} = 0.9$$

   determinare le tre stime puntuali e l'intervallo di confidenza approssimato ponendo $1 - \alpha = 0.90$

**Esercizio 36*.** Sia $X$ una v.a. assolutamente continua con funzione di densità

$$f_X(x; \theta) = \theta\, x^{\theta - 1}, \qquad x \in (0, 1), \qquad \theta > 0.$$

1. Determinare la funzione di verosimiglianza e lo stimatore di massima verosimiglianza.

2. Si consideri il sistema di ipotesi

$$H_0 : \ \theta \leq 1 \qquad \text{vs.} \qquad H_1 : \ \theta > 1.$$

   ed il test con regione di rifiuto basata su una singola osservazione ($n = 1$) e definita da:

$$R = \left\{ x \in \mathcal{X} : x > \frac{1}{2} \right\}.$$

   Determinare la funzione di potenza del test e calcolarne il valore in $\theta = 2$.

[6] Parte di questo esercizio è stato proposto nella sezione sulle proprietà degli stimatori.

## SOLUZIONI

Intervalli di Confidenza e Test

**Esercizio 1.** Poiché $E(X) = 0$, abbiamo che $E(X^2) = Var(X) = \theta$ e anche $E(X_i^2) = \theta$. Quindi

$$E(T) = E(\frac{\sum_{i=1}^{n} X_i^2}{n}) = \frac{\sum_{i=1}^{n} E(X_i^2)}{n} = \theta.$$

Sappiamo che se $X \sim N(\mu, \sigma^2)$ allora $\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2} \sim \chi_n^2$; quindi nel nostro caso

$$\frac{nT}{\theta} \sim \chi_n^2$$

è una quantità pivotale. Possiamo usare questa per costruire l'intervallo di confidenza per un livello fissato $1 - \alpha$:

$$1 - \alpha = \mathbb{P}(q_1 \leq \frac{nT}{\theta} \leq q_2) = \mathbb{P}(\chi^2_{\alpha/2;n} \leq \frac{nT}{\theta} \leq \chi^2_{1-\alpha/2;n}) = \mathbb{P}(\frac{nT}{\chi^2_{1-\alpha/2;n}} \leq \theta \leq \frac{nT^2}{\chi_{\ \alpha/2;n}}).$$

Per $n = 15$ e un valore osservato di $T = 3.23^2$, l'intervallo al 95% per $\theta$ è

$$\left[\frac{156.49}{27.49}, \frac{156.49}{6.26}\right] = [5.69, 24.99] .$$

**Esercizio 2.** Possiamo utilizzare la quantità pivotale $\frac{\bar{X}-\mu}{S_n/\sqrt{n}} \sim t_{n-1}$ per costruire l'intervallo per $\mu$ e $\frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2_{n-1}$ per $\sigma^2$. Gli intervalli sono dati dalle relazioni

$$\begin{aligned} 1 - \alpha &= \mathbb{P}\left(t_{\alpha/2;n-1} \leq \frac{\bar{X}-\mu}{S_n/\sqrt{n}} \leq t_{1-\alpha/2;n-1}\right) = \mathbb{P}\left(\bar{X} + t_{\alpha/2;n-1}S_n/\sqrt{n} \leq \mu \leq \bar{X} + t_{1-\alpha/2;n-1}S_n/\sqrt{n}\right) \\ &= \mathbb{P}\left(\bar{X} - t_{1-\alpha/2;n-1}S_n/\sqrt{n} \leq \mu \leq \bar{X} + t_{1-\alpha/2;n-1}S_n/\sqrt{n}\right) \end{aligned}$$

(l'ultima uguaglianza è giustificata dalla simmetria della distribuzione $t$) e

$$1 - \alpha = \mathbb{P}\left(\chi^2_{\alpha/2;n-1} \leq \frac{(n-1)S_n^2}{\sigma^2} \leq \chi^2_{1-\alpha/2;n-1}\right) = \mathbb{P}\left(\frac{(n-1)S_n^2}{\chi^2_{1-\alpha/2;n-1}} \leq \sigma^2 \leq \frac{(n-1)S_n^2}{\chi^2_{\alpha/2;n-1}}\right).$$

Dalle tavole si ha: $t_{1-\alpha/2;n-1} = t_{0.95,100} = 1.66$, $\chi^2_{1-\alpha/2;n-1} = \chi^2_{0.95;100} = 124.3$ e $\chi^2_{\alpha/2;n-1} = \chi^2_{0.05;100} = 77.9$. Gli intervalli al 90% relativi al campione osservato sono

$$\text{per } \mu : [18.08, 18.32] \qquad \text{per } \sigma^2 : [0.45, 0.72].$$

Si noti che, in quest'esempio, la numerosità campionaria è elevata. Si può pertanto ricorrere all'approssimazione normale delle v.a. $t$ e $\chi^2$.

**Esercizio 3.** Non abbiamo una quantità pivotale per $p$, pero' possiamo utilizzare l'approssimazione normale per la distribuzione dello stimatore di massima verosimiglianza. Sappiamo che lo stimatore MLE di $p$ è $\hat{p} = \sum_{i=1}^{n} X_i/n$ (la frazione di successi). Sappiamo anche che, per $n$ grande $\hat{p}$ ha distribuzione approssimativamente normale con valore atteso $p$ (il parametro incognito) e varianza $1/I_n(p)$ (il limite inferiore di Cramer-Rao). Quindi possiamo costruire l'intervallo di confidenza usando la quantità pivotale

$$\frac{\hat{p} - p}{\sqrt{1/I_n(p)}} \sim N(0, 1)$$

e l'intervallo al livello $1-\alpha$ è

$$z_{\alpha/2} \leq \frac{\hat{p}-p}{\sqrt{1/I_n(p)}} \leq z_{1-\alpha/2}.$$

In questo caso la varianza è

$$\frac{1}{I_n(p)} = \frac{1}{-E\left(\frac{\partial^2}{\partial p^2}\log L(p;\mathbf{X}_n)\right)} = \cdots = \frac{p(1-p)}{n}$$

che rende difficile esplicitare l'intervallo rispetto a $p$. In questi casi si fa una ulteriore approssimazione considerando l'informazione di Fisher osservata invece di quella attesa, quindi

$$\frac{1}{I_n^{oss}(\hat{p})} = \frac{1}{\left[\frac{\partial^2}{\partial p^2}\log L(p;\mathbf{X}_n)\right]_{p=\hat{p}}} = \cdots = \frac{\hat{p}(1-\hat{p})}{n}.$$

L'intervallo approssimato è (considerando anche che $z_{\alpha/2} = -z_{1-\alpha/2}$)

$$-z_{1-\alpha/2} \leq \frac{\hat{p}-p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \leq z_{1-\alpha/2} \qquad \text{e quindi} \qquad \hat{p} - z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.$$

Relativamente al campione osservato (con $n = 100$ e $\hat{p} = 0.8$), l'intervallo al 95% è

$$\left[0.8 - 1.96\sqrt{\frac{0.16}{100}}, 0.8 + 1.96\sqrt{\frac{0.16}{100}}\right] \quad = \quad [0.7216, 0.8784].$$

**Esercizio 4.** Si ha che: $L(p;\mathbf{x}_n) = p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}$; relativamente al campione osservato (con $n = 371$ e $\sum_{i=1}^n x_i = 18$), la funzione di verosimiglianza è $L(p;\mathbf{x}_n) = p^{18}(1-p)^{353}$. Per l'intervallo di confidenza le argomentazioni sono quelle fatte all'esercizio 3, e l'intervallo è

$$\hat{p} - z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.$$

Il valore della stima di massima verosimiglianza è $\hat{p} = \frac{18}{371} = 0.048$ e quindi l'intervallo al 90% è $[0.029, 0.066]$. Inoltre, $L(0.07, \mathbf{x}_n) = 1.22 \cdot 10^{-32}$, $\quad L(0.03, \mathbf{x}_n) = 8.29 \cdot 10^{-33}$. Quindi il valore 0.07 è più verosimile. Infatti il rapporto tra le verosimiglianze è $\frac{L(0.07,x')}{L(0.03,x')} \simeq 1.47$, da cui si vede che 0.07 ha verosimiglianza più alta.

**Esercizio 5.** Il rapporto delle verosimiglianze:

$$\frac{L(p_0,\mathbf{x}_n)}{L(p_1,\mathbf{x}_n)} \;=\; \frac{p_0^{\sum_{i=1}^n x_i}(1-p_0)^{n-\sum_{i=1}^n x_i}}{p_1^{\sum_{i=1}^n x_i}(1-p_1)^{n-\sum_{i=1}^n x_i}} = \left(\frac{p_0}{p_1}\right)^{\sum_{i=1}^n x_i}\left(\frac{1-p_0}{1-p_1}\right)^{n-\sum_{i=1}^n x_i} = \left(\frac{p_0}{p_1}\right)^{T}\left(\frac{1-p_0}{1-p_1}\right)^{n-T}$$

avendo posto $T = \sum_{i=1}^n x_i$. Tale rapporto è monotono rispetto a $T$. Infatti, essendo $p_0 < p_1$ e $1-p_0 > 1-p_1$, si ha:

$$\left(\frac{p_0}{p_1}\right)^{T} \qquad \text{decresce al crescere di } T$$

$$\left(\frac{1-p_0}{1-p_1}\right)^{n-T} \qquad \text{decresce al crescere di } T$$

e quindi il rapporto di verosimiglianze è decrescente rispetto a $T$. Di conseguenza la regione di rifiuto

$$\left\{\mathbf{x}_n \in \mathcal{X}^n : \frac{L(p_0, \mathbf{x}_n)}{L(p_1, \mathbf{x}_n)} \leq s\right\} = \{\mathbf{x}_n \in \mathcal{X}^n : T \geq g(s)\}$$

è del tipo $\{\mathbf{x}_n \in \mathcal{X}^n : T \geq k\}$. Se fissiamo $n = 4$ e $k = 2$, la regione di rifiuto diventa $R = \left\{\mathbf{x}_4 \in \mathcal{X}^4 : \sum_{i=1}^4 x_i \geq 2\right\}$. Ricordando che $\sum_{i=1}^n X_i \sim \text{Binom}(4, p)$, la probabilità di errore di prima specie è

$$\begin{aligned} Pr(R|p_0) &= \mathbb{P}(\sum_{i=1}^4 X_i \geq 2|p_0) = \mathbb{P}(\text{Binom}(4, p_0) \geq 2) \\ &= \binom{4}{2} p_0^2(1-p_0)^2 + \binom{4}{3} p_0^3(1-p_0)^1 + \binom{4}{4} p_0^4(1-p_0)^0 = 0.1808. \end{aligned}$$

**Esercizio 6.** Per $\alpha = 1$ si ha $\Gamma(\alpha) = 1$ e $\Gamma(\alpha + x) = \Gamma(x+1)$, quindi

$$f_X(x; \theta, \alpha = 1) = \theta(1-\theta)^x.$$

La funzione di verosimiglianza è $L(\theta; \mathbf{x}_n) = \theta^n(1-\theta)^{\sum_{i=1}^n x_i}$ che nel campione osservato diventa $L(\theta; (3,2,7,3,5)) = \theta^5(1-\theta)^{20}$. Rispetto ai due valori da confrontare, abbiamo

$$L(0.3; (3,2,7,3,5)) = (0.3)^5(0.7)^{20} > L(0.6; (3,2,7,3,5)) = (0.6)^5(0.4)^{20}.$$

e quindi $\theta_1 = 0.3$ è più verosimile.

**Esercizio 7.** La costruzione è analoga a quella dell'esercizio 2. Questa volta, però, la numerosità non è molto alta e quindi dobbiamo necessariamente usare le tavole della v.a. $t_{15}$ e della v.a. $\chi^2_{15}$ senza ricorrere all'approssimazione normale. Gli intervalli di confidenza al livello $1-\alpha$ sono

$$\begin{aligned} &\text{per } \mu: && \left(\bar{X} - t_{1-\alpha/2;n-1} S_n/\sqrt{n} \leq \mu \leq \bar{X} + t_{1-\alpha/2;n-1} S_n/\sqrt{n}\right) \\ &\text{per } \sigma^2: && \left(\frac{(n-1)S_n^2}{\chi^2_{1-\alpha/2;n-1}} \leq \sigma^2 \leq \frac{(n-1)S_n^2}{\chi^2_{\alpha/2;n-1}}\right) \end{aligned}$$

Relativamente al campione osservato gli intervalli sono

$$\text{per } \mu\text{: } [26.48, 29.32] \qquad \text{per } \sigma^2\text{: } [6.26, 21.55].$$

**Esercizio 8.** Trattiamo prima il test con due ipotesi semplici

$$H_0 : \mu = 20 \qquad \text{contro} \qquad H_1 : \mu = \mu_1 \quad (\text{con } \mu_1 > 20).$$

La funzione di verosimiglianza è

$$L(\mu, \mathbf{x}_n) = e^{-\frac{n}{2\sigma^2}(\bar{x}-\mu)^2}$$

quindi il rapporto delle verosimiglianze è

$$\frac{L(\mu_0, \mathbf{x}_n)}{L(\mu_1, \mathbf{x}_n)} = e^{-\frac{n}{2\sigma^2}[(\bar{x}-\mu_0)^2-(\bar{x}-\mu_1)^2]} = e^{-\frac{n}{2\sigma^2}[(\bar{x}(\mu_1-\mu_0)+\mu_0^2-\mu_1^2]}.$$

Essendo $\mu_1 > \mu_0$, il rapporto è decrescente rispetto a $\bar{x}$, quindi la regione di rifiuto è

$$R = \left\{\mathbf{x}_n \in \mathcal{X}^n : \frac{L(\mu_0, \mathbf{x}_n)}{L(\mu_1, \mathbf{x}_n)} \leq k\right\} = \{\mathbf{x}_n \in \mathcal{X}^n : \bar{x} \geq g(k)\}.$$

Troviamo $g(k)$ attraverso la condizione sull'errore di prima specie, ricordando che $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$:

$$0.05 = \mathbb{P}(R|\mu_0) = \mathbb{P}(\bar{X} \geq g(k)|\mu_0) = \mathbb{P}(\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} \geq \frac{g(k)-\mu_0}{\sigma/\sqrt{n}})$$

quindi $g(k)$ è tale che $\frac{g(k)-\mu_0}{\sigma/\sqrt{n}} = 1.65$ (e quindi $g(k) = \mu_0 + 1.65\sigma/\sqrt{n}$).
La regione di rifiuto è:

$$R = \left\{ \mathbf{x}_n \in \mathcal{X}^n : \bar{X} \geq \mu_0 + 1.65\sigma/\sqrt{n} \right\}.$$

La regione di rifiuto non dipende dalla scelta di $\mu_1$ (purché rimanga $\mu_1 > \mu_0$). Quindi possiamo considerare la stessa regione di rifiuto per il test con ipotesi composte (il test risultante è UMP)

$$H_0 : \mu = 20 \qquad \text{contro} \qquad H_1 : \mu > 20.$$

In particolare rifiuteremo (con $\alpha = 5\%$) se $\bar{x} \geq 20 + 1.65 \cdot 4/\sqrt{10} = 22.08$. Nel campione osservato la media campionaria è $24.3 > 22.08$ e quindi RIFIUTIAMO $H_0$.
La funzione potenza è

$$\mathbb{P}(R|\mu) \;=\; \mathbb{P}(\bar{X} \geq 22.08|\mu) = \mathbb{P}(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \geq \frac{22.08-\mu}{4/\sqrt{10}}) = 1 - \Phi(\frac{22.08-\mu}{4/\sqrt{10}}),$$

dove $\Phi(\cdot)$ è la funzione di ripartizione della v.a. normale standardizzata. In $\mu = 25$ la funzione potenza vale $1 - \Phi(\frac{22.08-25}{4/\sqrt{10}}) = 1 - \Phi(-2.30) = 0.99$.

**Esercizio 9.** Il rapporto delle verosimiglianze è

$$\frac{L(\theta_0, \mathbf{x}_n)}{L(\theta_1, \mathbf{x}_n)} \;=\; \frac{\theta_0^n e^{-\theta_0 \sum_{i=1}^n x_i}}{\theta_1^n e^{-\theta_1 \sum_{i=1}^n x_i}} = \left(\frac{\theta_0}{\theta_1}\right)^n e^{-(\theta_0-\theta_1)\sum_{i=1}^n x_i}.$$

Poiché nel nostro test $\theta_0 - \theta_1 < 0$, il rapporto di verosimiglianze è funzione crescente di $\sum_{i=1}^n x_i$, la regione di rifiuto è

$$R = \left\{ \mathbf{x}_n \in \mathcal{X}^n : \frac{L(\theta_0, \mathbf{x}_n)}{L(\theta_1, \mathbf{x}_n)} \leq k \right\} = \left\{ \mathbf{x}_n \in \mathcal{X}^n : \sum_{i=1}^n x_i \leq k \right\}.$$

Dobbiamo calcolare $\mathbb{P}(R|\theta = 1)$ avendo fissato $k = \frac{2}{3}$ e supponendo che $n = 36$. Ricordando che $\sum_{i=1}^n X_i \sim \text{Gamma}(\lambda = \theta, \nu = n)$ (e quindi sotto l'ipotesi nulla $\sum_{i=1}^n X_i \sim \text{Gamma}(\lambda = 1, \nu = 36)$) abbiamo

$$\mathbb{P}(R|\theta = 1) \;=\; \mathbb{P}(\sum_{i=1}^n x_i \leq \frac{2}{3}|\theta = 1) = F_{\text{Gamma}(1,36)}(\frac{2}{3}) = 6.43 \cdot 10^{-49} \simeq 0$$

Utilizzando l'approssimazione normale alla distribuzione di $\bar{X}$ (con valore atteso $E(X) = 1/\theta$ e varianza $\frac{V(X)}{n} = \frac{1}{n\theta^2}$, abbiamo

$$\begin{aligned} \mathbb{P}(R|\theta = 1) &= \mathbb{P}(\sum_{i=1}^n X_i \leq \frac{2}{3}|\theta = 1) = \mathbb{P}(\bar{X} \leq \frac{2}{3 \cdot 36}|\theta = 1) \\ &= \mathbb{P}(\frac{\bar{X}-1}{\sqrt{1/36}} \leq \frac{\frac{2}{108}-1}{\sqrt{1/36}}) = \mathbb{P}(Z \leq -5.89) \simeq 0. \end{aligned}$$

**Esercizio 10.** a) L'intervallo di confidenza (basato sulla quantità pivotale $\frac{\bar{X}_n-\mu}{S_n/\sqrt{n}}$) al livello $1-\alpha$ è

$$\bar{x} - t_{1-\alpha/2;n-1}\frac{S_n}{\sqrt{n}} \quad \leq \mu \leq \quad \bar{x} + t_{1-\alpha/2;n-1}\frac{S_n}{\sqrt{n}}.$$

Per $\alpha = 0.2$ e $n = 10$, si ha $t_{1-\alpha/2;n-1} = 1.38$ e quindi l'intervallo diventa (con i dati osservati) $[9.20, 9.35]$.

b) Ricordiamo brevemente di seguito le regioni critiche dei test con ipotesi nulla semplice (sulla media), e ipotesi alternativa unilaterale, sia nel caso di varianza nota che incognita.

Varianza nota: $H_0 : \mu = \mu_0$ contro $H_1 : \mu > \mu_0$ (vale anche se $H_0 : \mu \leq \mu_0$)

$$R = \left\{ x \in \mathcal{X}^n : \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \geq z_{1-\alpha} \right\}.$$

$H_0 : \mu = \mu_0$ contro $H_1 : \mu < \mu_0$ (vale anche se $H_0 : \mu \geq \mu_0$)

$$R = \left\{ x \in \mathcal{X}^n : \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \leq z_{\alpha} \right\}.$$

Varianza ignota: $H_0 : \mu = \mu_0$ contro $H_1 : \mu > \mu_0$ (vale anche se $H_0 : \mu \leq \mu_0$)

$$R = \left\{ x \in \mathcal{X}^n : \frac{\bar{x} - \mu_0}{s_n/\sqrt{n}} \geq t_{1-\alpha;n-1} \right\}.$$

$H_0 : \mu = \mu_0$ contro $H_1 : \mu < \mu_0$ (vale anche se $H_0 : \mu \geq \mu_0$)

$$R = \left\{ x \in \mathcal{X}^n : \frac{\bar{x} - \mu_0}{s_n/\sqrt{n}} \leq t_{\alpha;n-1} \right\}.$$

Il test che ci interessa in questo esercizio è il quarto. Con $n = 10$ e $\alpha = 0.01$, quindi la soglia è $t_{0.01;9} = -2.82$;

$$\frac{\bar{x} - \mu_0}{S_n/\sqrt{n}} = \frac{9.28 - 10.5}{0.1619/\sqrt{10}} = -23.82$$

che fa rifiutare decisamente l'ipotesi nulla.

In questo caso il $p$-value è $\mathbb{P}(t_9 < 23.82) \simeq 0$.

**Esercizio 11.** L'intervallo di confidenza (basato sulla quantità pivotale $\frac{\bar{X}_n - \mu}{S_n/\sqrt{n}}$) al livello $1 - \alpha$ è

$$\bar{X}_n - t_{1-\alpha/2;n-1}\frac{S_n}{\sqrt{n}} \quad \leq \mu \leq \quad \bar{X}_n + t_{1-\alpha/2;n-1}\frac{S_n}{\sqrt{n}}.$$

Per $\alpha = 0.05$ e $n = 20$, si ha $t_{0.975;19} = 2.093$ e quindi l'intervallo diventa (con i dati osservati) $[3.130, 3.330]$.

Il test è del terzo tipo tra quelli elencati all'esercizio precedente, quindi la regione critica è

$$R = \left\{ \mathbf{x}_n \in \mathcal{X}^n : \frac{\bar{x} - \mu_0}{S_n/\sqrt{n}} \geq t_{1-\alpha;n-1} \right\}.$$

Per $n = 20$ e $\alpha = 0.05$ la soglia è $t_{0.95;19} = 1.73$; la statistica test $\dfrac{3.23 - 3.14}{0.214/\sqrt{20}} = 1.88$ è più grande della soglia e quindi si rifiuta l'ipotesi nulla.

Per $n = 20$ e $\alpha = 0.025$ la soglia è $t_{0.975;19} = 2.09$; la statistica test $\dfrac{3.23 - 3.14}{0.214/\sqrt{20}} = 1.88$ è più piccola della soglia e quindi si accetta l'ipotesi nulla.

Il $p$-value è uguale a $\mathbb{P}(t_{19} > 1.881) = 0.037 < 0.05$.

**Esercizio 12.** La funzione di verosimiglianza è

$$L(\mu, \sigma^2; \mathbf{x}_n) = (\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu)^2}.$$

Se $\mu = \mu_0$, la verosimiglianza ha massimo se $\sigma^2 = s_0^2 = \frac{1}{n}\sum_{i=1}^n (x_i - \mu_0)^2$, quindi al numeratore avremo $\sup_{\mu=\mu_0,\sigma^2>0} L(\mu, \sigma^2; \mathbf{x}_n) = L(\mu_0, s_0^2; \mathbf{x}_n)$.

Sappiamo che la verosimiglianza è massima per $(\mu = \bar{x}, \sigma^2 = s^2)$ (dove $s^2$ è la varianza campionaria non corretta), quindi al denominatore metteremo $\sup_{\mu\in\mathbb{R},\sigma^2>0} L(\mu, \sigma^2; \mathbf{x}_n) = L(\bar{x}, s^2; \mathbf{x}_n)$. Ricordando che $s_0^2 = s^2 + (\bar{x} - \mu)^2$, il rapporto delle massime verosimiglianze è

$$\begin{aligned}\frac{\sup_{\mu=\mu_0,\sigma^2>0} L(\mu, \sigma^2; \mathbf{x}_n)}{\sup_{\mu\in\mathbb{R},\sigma^2>0} L(\mu, \sigma^2; \mathbf{x}_n)} &= \frac{L(\mu_0, s_0^2; \mathbf{x}_n)}{L(\bar{x}, s^2; \mathbf{x}_n)} = \frac{s_0^{2^{-\frac{n}{2}}} e^{-\frac{1}{2s_0^2}\sum_{i=1}^n (x_i-\mu_0)^2}}{s^{2^{-\frac{n}{2}}} e^{-\frac{1}{2s^2}\sum_{i=1}^n (x_i-\bar{x})^2}} = \frac{s_0^{2^{-\frac{n}{2}}} e^{-\frac{n}{2}}}{s^{2^{-\frac{n}{2}}} e^{-\frac{n}{2}}} = \left(\frac{s_0^2}{s^2}\right)^{-\frac{n}{2}} \\ &= \left(1 + \frac{(\bar{x}-\mu)^2}{s^2}\right)^{-\frac{n}{2}}\end{aligned}$$

che è funzione decrescente di $\left(\frac{\bar{x}-\mu}{s}\right)^2$. La regione critica quindi sarà del tipo

$$R = \left\{\mathbf{x}_n \in \mathcal{X}^n : \left(\frac{\bar{x}-\mu}{s}\right)^2 \geq k\right\} = \left\{\mathbf{x}_n \in \mathcal{X}^n : \frac{\bar{x}-\mu}{s} \leq -\sqrt{k} \text{ oppure } \frac{\bar{x}-\mu}{s} \geq \sqrt{k}\right\}.$$

Per trovare le soglie $\pm\sqrt{k}$ usiamo la condizione sulla probabilità di errore di prima specie $Pr(R|H_0) = \alpha$. Non conosciamo la distribuzione di $\frac{\bar{X}-\mu}{S_n}$, però sappiamo che $\frac{\bar{X}-\mu}{S_n/\sqrt{n}} \sim t_{n-1}$. Poichè $\frac{\bar{X}-\mu}{S_n} = \frac{\bar{X}-\mu}{S_n/\sqrt{n}} \cdot \frac{1}{\sqrt{n-1}}$, la regione critica può essere espressa in termini di $\frac{\bar{x}-\mu}{S_n/\sqrt{n}}$:

$$R = \left\{\mathbf{x}_n \in \mathcal{X}^n : \frac{\bar{x}_n-\mu}{s/\sqrt{n}} \leq -k' \text{ oppure } \frac{\bar{x}_n-\mu}{s/\sqrt{n}} \geq k'\right\}$$

(dove $k' = \sqrt{k}\,\sqrt{n-1}$). Per la condizione $Pr(R|H_0) = \alpha$, $k'$ non può che essere $t_{1-\alpha/2;n-1}$ e quindi la regione di rifiuto

$$R = \left\{\mathbf{x}_n \in \mathcal{X}^n : \frac{\bar{x}-\mu}{s/\sqrt{n}} \leq -t_{1-\alpha/2;n-1} \text{ oppure } \frac{\bar{x}_n-\mu}{s/\sqrt{n}} \geq t_{1-\alpha/2;n-1}\right\}.$$

Con i dati dell'esercizio, $\frac{\bar{x}-\mu}{s_n/\sqrt{n}} = 1.88$, mentre la soglia $t_{0.975;19} = 2.093$. Quindi accettiamo l'ipotesi nulla per $\alpha = 0.05$.

**Esercizio 13.**
a) Segue dal teorema centrale di convergenza, osservando che $X_1, \ldots, X_n\ldots$ sono i.i.d. e che: $E(\bar{X}_n) = \theta$, $V(\bar{X}_n) = \theta/n$.
b) $1 - \beta = \mathbb{P}(\mathbf{x}_n \in R|\theta = \theta_1) = \mathbb{P}(\bar{X}_n < 5.5|\theta = 5) \simeq \mathbb{P}(Z < (5.5-5)/\sqrt{5/25}) = \Phi(1.12) = 0.87$.
c) Posto $y = \sum_{i=1}^n x_i$, si ha che $\lambda(\mathbf{x}_n) = e^{-n(\theta_0-\theta_1)}(\theta_0/\theta_1)^y$. Poichè $\theta_0 > \theta_1$, $\lambda$ è una funzione monotona di $y$ e quindi, per l'arbitrarietà di $k$, si ha: $A = \{\mathbf{x}_n \in \mathcal{X}^n : \ y > k\}$.

**Esercizio 14.**
a) Si ha che:

$$\mathcal{L}(\mathbf{X}_n) = U_n(\mathbf{X}_n) - L_n(\mathbf{X}_n) = (\frac{1}{\alpha^{1/n}} - 1)X_{(n)} = \frac{1-\alpha^{1/n}}{\alpha^{1/n}}X_{(n)}.$$

Pertanto, per la linearità del valore atteso,

$$E[\mathcal{L}(\mathbf{X}_n)] = \frac{1-\alpha^{1/n}}{\alpha^{1/n}}E[X_{(n)}] = \frac{1-\alpha^{1/n}}{\alpha^{1/n}}\frac{n+1}{n}\theta.$$

b) Per $x_{10} = 10$ si ha: $IC(\mathbf{x}_{(n)}) = (10, \frac{10}{(0.05)^{1/10}}) = (10, 13.7)$.

**Esercizio 15.**
a) Segue dal teorema centrale di convergenza, osservando che $X_1, \ldots, X_n$... sono i.i.d. e che: $E(\sum_{i=1}^n X_i) = n\theta$, $V(\sum_{i=1}^n X_i) = n\theta(1-\theta)$.
b)

$$\begin{aligned} \alpha &= \mathbb{P}(\sum_{i=1}^n X_i < k | \theta = \theta_0) \simeq \mathbb{P}(Z < \frac{k - n\theta_0}{\sqrt{n\theta_0(1-\theta_0)}}) \\ &\simeq \Phi\left(\frac{60-72}{\sqrt{36}}\right) = \Phi(-2) = 0.02. \end{aligned}$$

Analogamente,

$$\begin{aligned} 1-\beta &= \mathbb{P}(\sum_{i=1}^n X_i < k | \theta = \theta_1) \simeq \mathbb{P}(Z < \frac{k - n\theta_1}{\sqrt{n\theta_1(1-\theta_1)}}) \\ &\simeq \Phi\left(\frac{60 - 144(0.35)}{\sqrt{144(0.35)(1-0.35)}}\right) = 0.953. \end{aligned}$$

Pertanto: $\beta = 0.047$.
c) Sia $y = \sum_{i=1}^n x_i$. Il rapporto delle verosimiglianze è: $\lambda(\mathbf{x}_n) = \left(\frac{\theta_0}{\theta_1}\right)^y \left(\frac{1-\theta_0}{1-\theta^1}\right)^{n-y}$. Tale funzione è monotona in $y$. Pertanto: $\lambda(\mathbf{x}_n) < k \Leftrightarrow \sum_{i=1}^n x_i < k'$.

**Esercizio 16.**
Per l'indipendenza e la normalità dei due campioni casuali, si ha che la v.a. $\bar{X}_{n_1} - \bar{X}_{n_2} \sim N(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})$. Da quanto scritto discende che $\frac{\bar{X}_{n_1} - \bar{X}_{n_2} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$ è una quantità pivotale per il parametro $\mu_1 - \mu_2$. Pertanto

$$\left(\bar{X}_{n_1} - \bar{X}_{n_2} - z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}},\ \bar{X}_{n_1} - \bar{X}_{n_2} + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$$

è un intervallo di confidenza al livello $1-\alpha$ per il parametro considerato. Sostituendo i valori campionari si ottiene: $(-12.56, 0.55)$. Poichè tale intervallo contiene lo zero, al livello di confidenza considerato non possiamo escludere che $\mu_1$ e $\mu_2$ siano uguali.

**Esercizio 17.**
a) Si ha che

$$E[\mathcal{L}(\mathbf{X}_n)] = U_n(\mathbf{X}_n) - L_n(\mathbf{X}_n) = 2\sum_{i=1}^n X_i \left(\frac{1}{\chi^2_{1-\frac{\alpha}{2};2n}} - \frac{1}{\chi^2_{\frac{\alpha}{2};2n}}\right) = c \times \bar{X}_n,$$

con $c = 2n\left(\frac{1}{\chi^2_{1-\frac{\alpha}{2};2n}} - \frac{1}{\chi^2_{\frac{\alpha}{2};2n}}\right)$. Quindi:

$$E[\mathcal{L}(\mathbf{X}_n)] = c\theta, \qquad V[\mathcal{L}(\mathbf{X}_n)] = c^2\frac{\theta^2}{n}.$$

b) Per il campione osservato, si ha (utilizzando le tavole della v.a. $\chi^2$): $IC(\mathbf{x}_n) = (\frac{2\times 1740}{34.170}, \frac{2\times 1740}{9.591}) = (101.84, 322.84)$.

**Esercizio 18.** Vedere soluzione esercizio 10 b).

a) Qui si ha: $t_{n-1;1-\frac{\alpha}{2}} = t_{19;0.975} = 2.09$. Per il campione assegnato, la statistica test assume il valore $t_{oss} = \frac{\bar{x}_n - \mu_0}{s/\sqrt{n}} = \frac{353.8-350}{21.85/\sqrt{20}} = 0.78$. Si accetta pertanto $H_0$.
b) Il $p$-value è qui uguale a $2\,\mathbb{P}(t_{n-1} > |t_{oss}|) = 2[1 - T_{n-1}(|t_{oss}|)] = 2[1 - T_{19}(t_{oss})] = 0.445$, dove $T_{n-1}$ indica qui la funzione di ripartizione di una v.a. $t$ di Student con $n-1$ gradi di libertà. Il $p$-value è elevato: i dati a disposizione forniscono una evidenza forte a favore di $H_0$.

**Esercizio 19.**

a) IC risulta: $\bar{X}_n \pm z_{1-\alpha/2}\sigma/\sqrt{n}$. Per il campione osservato e per l'$\alpha$ considerato si ha: $[169.9, 185.1]$.

b) $R = \{\mathbf{x}_n : \sqrt{n}(\bar{x}_n - \mu)/\sigma = t_{oss} > z_{1-\alpha/2}\}$. In questo caso $t_{oss} = (177.5 - 175)/\sqrt{300/20} = 0.645 < 1.644 = z_{0.95}$. Si accetta quindi $H_0$.

c) Il $p$-value è $\mathbb{P}\{Z > t_{oss} = 0.645\} = 1 - \Phi(0.645) = 0.259$. E' un valore elevato che indica una evidenza sostanziale a favore dell'ipotesi nulla.

**Esercizio 20.** La semilunghezza dell'IC è $\ell = z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$. Si ha che $\ell > 200 \Leftrightarrow n > \left(z_{1-\alpha/2}\frac{\sigma}{200}\right)^2 = \left(2.57 \times \frac{900}{200}\right)^2$, da cui si ottiene che $n^* = 134$.

**Esercizio 21.** La funzione di potenza è $\beta(\mu) = \mathbb{P}(\frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}} > z_{1-\alpha}) = \mathbb{P}(\bar{X}_n > z_{1-\alpha}\frac{\sigma}{\sqrt{n}} + \mu_0)$. In generale, $\bar{X}_n \sim N(\mu, \frac{\sigma}{\sqrt{n}})$. Pertanto, $\beta(\mu) = \mathbb{P}(Z > \frac{\sqrt{n}(\mu_0 - \mu)}{\sigma} + z_{1-\alpha}) = 1 - \Phi(\frac{\sqrt{n}(\mu_0-\mu)}{\sigma} + z_{1-\alpha})$. Si ha in questo caso che $\beta(178) = 1 - \Phi(\sqrt{20}\frac{(175-178)}{\sqrt{300}} + 1.644) = 0.19$.

**Esercizio 22.**

a) Il test è del tipo:

$$H_0: \quad \theta < \theta_0 = 5, \qquad H_1: \quad \theta > \theta_0 = 5,$$

da cui,

$$R = \{\mathbf{x}_n : T(\mathbf{x}_n) > z_{1-\alpha}\}; \qquad \alpha = 0.01, \qquad z_{0.99} = 2.32$$

$$T(\mathbf{x}_n) = \frac{\bar{x}_n - \theta_0}{\sigma/\sqrt{n}} = \frac{5.5-5}{2/4} = \frac{0.5}{1/2} = 1 < z_{1-\alpha} = 2.32.$$

Si accetta quindi $H_0$.

b) Il valore-p si ottiene calcolando:

$$Pr\{T(\mathbf{X}_n) > 1 | \theta = \theta_0\} = 1 - \Phi(1) = 1 - 0.84 = 0.16.$$

**Esercizio 23.**

a) Per l'equivariana SMV:

$$\hat{g(\theta)} = g(\hat{\theta}) = e^{-\bar{X}}$$

b) Per il Delta method:

$$\hat{g(\theta)} \approx N\bigg(g(\theta), [g'(\theta)]^2 I^{-1}(\theta)\bigg).$$

Dato che

$$I(\theta) = \frac{n}{\theta}, \qquad g'(\theta) = -e^{-\theta}, \qquad [g'(\theta)]^2 = e^{-2\theta},$$

si ha

$$\hat{g(\theta)} \approx N\bigg(e^{-\theta}, e^{-2\theta}\ \frac{\theta}{n}\bigg)$$

c)

$$V[\hat{g(\theta)}] \approx e^{-2\bar{X}}\ \frac{\bar{X}}{n}$$

d)

$$IC_{(1-\alpha)}(\mathbf{X}) \approx g(\hat{\theta}) \pm z_{1-\frac{\alpha}{2}} \sqrt{\widehat{V[g(\hat{\theta})]}} = e^{-\bar{x}} \pm 1.96 e^{-2\bar{x}} \sqrt{\frac{\bar{x}}{n}}.$$

Dato che $\bar{x} = 2.8$, l'intervallo è: $IC = (0.058, 0.063)$.

**Esercizio 24.**

a) Per i noti risultati sulle statistiche medie campionarie per popolazioni normali, si ha che

$$\bar{X}^A_{n_A} - \bar{X}^B_{n_B} \sim N\left(\theta_A - \theta_B, 2\frac{\sigma^2}{n}\right).$$

Si ottiene così la seguente quantità pivotale :

$$\frac{\bar{X}^A_{n_A} - \bar{X}^B_{n_B} - \Delta}{\sigma\sqrt{\frac{2}{n}}} \sim N(0,1).$$

L'intervallo di confidenza per $\Delta$ è quindi:

$$[X^A_{n_A} - \bar{X}^B_{n_B} - z_{1-\frac{\alpha}{2}}\sigma\sqrt{\frac{2}{n}}, X^A_{n_A} - \bar{X}^B_{n_B} + z_{1-\frac{\alpha}{2}}\sigma\sqrt{\frac{2}{n}}],$$

dove $z_{1-\frac{\alpha}{2}}$ indica il percentile di livello $1 - \frac{\alpha}{2}$ della v.a. $N(0,1)$. Assumendo $1 - \alpha = 0.95$ si ha $z_{0.975} = 1.96$. Poiché $x^A_{n_A} = 1.22$ e $\bar{x}^B_{n_B} = 1.44$, si ottiene che l'intervallo osservato risulta essere $[-0.59, 0.15]$. Dal momento che tale intervallo include lo zero (valore che indica l'uguaglianza dei due parametri $\theta_A$ e $\theta_B$), sulla base dei dati disponibili non possiamo concludere che il confezionamento determini una differenza statisticamente significativa nei ricavi.

b) Si può rispondere al quesito in due modi.

I modo. L'intervallo di confidenza osservato contiene tutti i valori che corrispondono a ipotesi nulle accettate dal test bilaterale di livello $\alpha$, condotto con i dati a disposizione. In altre parole, se l'intervallo di confidenza di livello $1 - \alpha$ osservato contiene lo zero, l'ipotesi nulla $\Delta = 0$ contro l'alternativa $\Delta \neq 0$ viene accettata nel test di livello $\alpha$ condotto con gli stessi dati che si sono usati per determinare l'intervallo. (IMPORTANTE: notare la corrsipondenza tra livello di confidenza $(1 - \alpha)$ e livello del test $(\alpha)$).

II modo. Il valore osservato della statistica test per il confronto tra un'ipotesi nulla semplice e un'ipotesi alternativa bilaterale è

$$\frac{(x^A_{n_A} - \bar{x}^B_{n_B}) - 0}{\sigma\sqrt{\frac{2}{n}}} = \frac{0.22}{\sqrt{(0.16 \times 2)/9}} = 1.17 < 1.96 = z_{1-\frac{\alpha}{2}}.$$

Si accetta pertanto $H_0$.

**Esercizio 25.**

a) La funzione di verosimiglianza per il modello di Poisson è

$$L(\theta; \mathbf{x}_n) = e^{-n\theta}\theta^{\sum_{i=1}^n x_i},$$

da cui si ottiene il rapporto delle verosimiglianze

$$\lambda(\mathbf{x}_n) = e^{-n(\theta_0 - \theta_1)}\left(\frac{\theta_0}{\theta_1}\right)^{\sum_{i=1}^n x_i}.$$

Per il Lemma di Neyman e Pearson, osservando che $\lambda$ è funzione crescente di $\sum_{i=1}^{n} x_i$ (dal momento che $\theta_0 = 2 > 1 = \theta_1$), si ha che

$$R = \left\{ \mathbf{x}_n : \lambda(\mathbf{x}_n) \le k' \right\} = \{ \mathbf{x}_n : \bar{x}_n \le k \} .$$

Per $n = 3$ si ottiene quanto cercato.

b) Osserviamo innanzitutto che $Y_3 = X_1 + X_2 + X_3 \sim \text{Poisson}(3\theta)$. Per definizione, $\alpha = \mathbb{P}\{R; \theta = \theta_0\}$ e $1 - \beta = \mathbb{P}\{R; \theta = \theta_1\}$. Assumendo $k = 1$, si ha

$$\alpha = \mathbb{P}\{Y_3 \le 1; \theta = 2\} = \mathbb{P}\{Y_3 = 0; \theta = 2\} + \mathbb{P}\{Y_3 = 1; \theta = 2\} = e^{-6} + 6e^{-6} = 0.017,$$

$$1 - \beta = \mathbb{P}\{Y_3 \le 1; \theta = 1\} = \mathbb{P}\{Y_3 = 0; \theta = 1\} + \mathbb{P}\{Y_3 = 1; \theta = 1\} = e^{-3} + 3e^{-3} = 0.2.$$

In questo problema la distribuzione della statistica test è nota, e non è quindi necessario ricorrere alla distribuzione asintotica che, per $n$ sufficientemente elevato, risulta $N(n\theta, n\theta)$. Si noti inoltre che, in questo caso, essendo $n = 3$ molto piccolo, i valori che si ottengono utilizzando l'approssimazione normale sono piuttosto imprecisi. Infatti

$$\alpha \approx \Phi\left( \frac{1 - n\theta_0}{\sqrt{n\theta_0}} \right) = \Phi(-5/\sqrt{6}) = 0.0207,$$

$$1 - \beta \approx \Phi\left( \frac{1 - n\theta_1}{\sqrt{n\theta_1}} \right) = \Phi(-2/\sqrt{3}) = 0.124.$$

**Esercizio 26.**

a) Per la proprietà di equivarianza degli stimatori di MV si ha

$$\widehat{g(\theta)} = g(\hat{\theta}_{MV}) = \ln \bar{X}_n(1 - \bar{X}_n).$$

b) Per il *Delta Method* (le cui condizioni di applicabilità sono qui soddisfatte) si ha che $V_\theta[\widehat{g(\theta)}] \approx [g'(\theta)]^2 I_n(\theta)^{-1}$, dove, nel problema qui considerato, $g'(\theta) = \frac{d}{d\theta} \ln \theta(1-\theta) = \frac{1-2\theta}{\theta(1-\theta)}$ e $I_n(\theta) = \frac{n}{\theta(1-\theta)}$ è l'informazione attesa di Fisher per il modello bernoulliano. Si ha quindi che

$$V_\theta[\ln \bar{X}_n(1 - \bar{X}_n)] \approx \frac{(1 - 2\theta)^2}{n\theta(1 - \theta)}.$$

Inoltre, ricordando che l'informazione osservata risulta $I_n^{oss}(\mathbf{x}_n) = -\frac{d}{d\theta} \ln L(\theta; \mathbf{x}_n)|_{\theta=\hat{\theta}_{MV}} = \frac{n}{\bar{x}_n(1-\bar{x}_n)}$, si ha

$$\widehat{V_\theta[\widehat{g(\theta)}]} \approx [g'(\hat{\theta}_{MV})]^2 [I_n^{oss}(\mathbf{x}_n)]^{-1} = \frac{(1 - 2\bar{X}_n)^2}{n\bar{X}_n(1 - \bar{X}_n)}.$$

c) Per il *Delta Method* si ha che $\widehat{g(\theta)}$ ha distribuzione approssimativamente

$$N\left( g(\theta), [g'(\theta)]^2 I_n(\theta)^{-1}, \right)$$

da cui si ottiene l'intervallo di confidenza di livello $1 - \alpha$ approssimato

$$\widehat{g(\theta)} \pm z_{1-\frac{\alpha}{2}} \sqrt{[g'(\hat{\theta}_{MV})]^2 [I_n^{oss}(\mathbf{x}_n)]^{-1}}$$

In questo esercizio si ha quindi che

$$\ln \bar{X}_n(1 - \bar{X}_n) \sim N\left( \ln \theta(1 - \theta), \frac{|1 - 2\theta|}{\sqrt{n\theta(1 - \theta)}} \right)$$

e che l'intervallo approssimato è

$$\ln \bar{X}_n(1-\bar{X}_n) \pm z_{1-\frac{\alpha}{2}} \frac{|1-2\bar{x}_n|}{\sqrt{n\bar{x}_n(1-\bar{x}_n)}}.$$

Per i dati a disposizione si ottiene quindi:

$$\ln(0.6 * 0.4) \pm 1.96 \frac{0.2}{\sqrt{100 * 0.6 * 0.4}},$$

da cui si ottiene l'intervallo $[-1.51, -1.35]$.

**Esercizio 27.**
**1.** Per il test di Neyman-Pearson, la regione di accettazione è: $RA = \{\mathbf{x} \in \mathcal{X}^n : \lambda(\mathbf{x}) > k\}$, dove $\lambda(\mathbf{x}) = f_n(\mathbf{x}; \theta = \theta_0)/f_n(\mathbf{x}; \theta = \theta_1) = L_{\mathbf{x}}(\theta_0)/L_{\mathbf{x}}(\theta_1)$, e dove $k$ è tale che $\mathbb{P}(\mathbf{X} \in RR|H_o) = \alpha$, con $\alpha$ fissato. Dato che

$$\lambda(\mathbf{x}) = \frac{\theta_0^n \left[\prod(1-x_i)\right]^{\theta_0-1}}{\theta_1^n \left[\prod(1-x_i)\right]^{\theta_1-1}} = \left(\frac{\theta_0}{\theta_1}\right)^n \left[\prod_{i=1}^n (1-x_i)\right]^{\theta_0-\theta_1}$$

si ha che $\lambda(\mathbf{x}) > k$ quando:
$n(\log\theta_0 - \log\theta_1) + (\theta_0-\theta_1)\sum_{i=1}^n \log(1-x_i) > \log k$ e, dato che $\theta_0 - \theta_1 < 0$, si ottiene
$\sum_{i=1}^n \log(1-x_i) < \frac{\log k - n(\log\theta_0 - \log\theta_1)}{\theta_0-\theta_1}$, da cui: $\frac{n}{\sum_{i=1}^n \log(1-x_i)} > \frac{n(\theta_0-\theta_1)}{\log k - n(\log\theta_0-\log\theta_1)}$.
Quindi $RA = \{\mathbf{X} \in \mathcal{X}^n : \hat{\theta}_{\text{MLE}} < K\}$, con $K = -\frac{n(\theta_0-\theta_1)}{\log k - n(\log\theta_0-\log\theta_1)}$.

**2.** $\hat{\theta}_{MLE} \approx N(\theta, I^{-1}(\theta))$, dove $I(\theta) = n\mathbb{E}\big[-\frac{\partial^2}{\partial\theta^2}\log f_X(x;\theta)\big]$ è l'informazione attesa di Fisher. Dato che $\log f_X(x;\theta) = \log\theta + (\theta-1)\log(1-x)$ e $\frac{\partial^2}{\partial\theta^2}\log f_X(x;\theta) = -\theta^{-2}$, si ottiene che $I^{-1}(\theta) = \frac{\theta^2}{n}$ e quindi $\hat{\theta}_{MLE} \approx N(\theta, \frac{\theta^2}{n})$.

**3.** L'intervallo di confidenza approssimato di livello $1-\alpha$ si ottiene utilizzando la distribuzione asintotica di $\hat{\theta}_{MLE}$ dove il valore di $\theta$ nell'espressione della varianza è sostituito da $\hat{\theta}_{MLE}$. Nel caso di distribuzione campionaria $N(\theta, V^2)$, per uno stimatore $T(\mathbf{X})$, l'intervallo di confidenza è dato da:

$$IC = (T(\mathbf{x}_{oss}) - z_{1-\alpha/2}\, V \;,\; T(\mathbf{x}_{oss}) + z_{1-\alpha/2}\, V).$$

In questo caso $T(\mathbf{X}) = \hat{\theta}_{MLE}$ con distribuzione campionaria asintotica è $N(\theta, \hat{\theta}^2_{MLE}/n)$ ed il valore di $z_{\alpha/2} = z_{.9} = 1.28$. Quindi l'intervallo di confidenza approssimato è:

$$\widetilde{IC} \approx \big(\hat{\theta}_{MLE} - z_{.9}\,(\hat{\theta}_{MLE}/\sqrt{n})\,,\, \hat{\theta}_{MLE} - z_{.9}\,(\hat{\theta}_{MLE}/\sqrt{n})\big) = \big(1.5 - 1.28(0.1)\;,\; 1.5 + 1.28(0.1)\big) = (1.372\;,\; 1.628).$$

**4.** Il valore approssimato della potenza del test è:

$$\begin{aligned} 1-\beta &= \mathbb{P}(RR;\theta_1) = \mathbb{P}(\hat{\theta}_{MLE} > K;\theta_1) = \mathbb{P}\bigg(\frac{\hat{\theta}_{MLE}-\theta_1}{\hat{\theta}_{MLE}/\sqrt{n}} > \frac{K-\theta_1}{\hat{\theta}_{MLE}/\sqrt{n}}\bigg) \\ &\approx \mathbb{P}\bigg(Z > \frac{1.464-1.6}{0.1}\bigg) = \mathbb{P}(Z > -1.36) = 1 - \Phi(-1.36) = 0.913. \end{aligned}$$

**Esercizio 28.**
**1.** $\hat{\theta}_{MLE} \approx N(\theta, I^{-1}(\theta))$, dove $I(\theta) = n\mathbb{E}\Big[-\frac{\partial^2}{\partial\theta^2}\log f_X(x;\theta)\Big]$ è l'informazione attesa di Fisher. Dato che

$$\log f_X(x;\theta) = \log(\theta+2) + (\theta+1)\log x\;, \qquad \text{e che} \qquad \frac{\partial^2}{\partial\theta^2}\log f_X(x;\theta) = -\frac{1}{(\theta+2)^2},$$

si ha $I(\theta) = \dfrac{n}{(\theta+2)^2}$, quindi $\hat{\theta}_{MLE} \approx N\left(\theta, \dfrac{(\theta+2)^2}{n}\right)$.

**Parte facoltativa.** No, non è necessario. Perchè l'approssimazione campionaria asintotica dello stimatore di massima verosimiglianza non dipende dalla sua espressione matematica. Dipende solo dall'espressione matematica dell'informazione attesa di Fisher, il cui inverso è la varianza asintotica.

**2.** Per costruire un intervallo di confidenza approssimato per $\hat{\theta}_{MLE}$ si deve usare la distribuzione campionaria asintotica ottenuta al punto precedente, dove il valore di $\theta$ nell'espressione della varianza è sostituito da $\hat{\theta}_{MLE}$.
L'intervallo è dato da

$$\widetilde{IC}_{(1-\alpha)} = \left(\hat{\theta} \pm z_{1-\dfrac{\alpha}{2}} \frac{\hat{\theta}+2}{\sqrt{n}}\right) = \left(1.6 \pm 1.96\ \frac{3.6}{12}\right) = (1.6 \pm 0.588) = (1.012\ ,\ 2.188)$$

**3.** Dato che, come al punto **2**, la distribuzione campionaria asintotica di $\hat{\theta}$ è: $\hat{\theta} \approx N\left(\theta, \dfrac{(\hat{\theta}+2)^2}{n}\right)$, si ha:

$$\mathbb{P}\left\{RR|\theta=\theta_0\right\} = \mathbb{P}\{\hat{\theta} < k|\theta=\theta_0\} = \mathbb{P}\left\{\frac{\hat{\theta}-\theta_0}{(\hat{\theta}+2)/\sqrt{n}} < \frac{k-\theta_0}{(\hat{\theta}+2)/\sqrt{n}}\right\} \approx \Phi\left(\frac{k-\theta_0}{(\hat{\theta}+2)/\sqrt{n}}\right) = \alpha.$$

Da cui si ottiene che $\dfrac{k-\theta_0}{(\hat{\theta}+2)/\sqrt{n}} = z_\alpha$ e $k = z_\alpha \dfrac{\hat{\theta}+2}{\sqrt{n}} + \theta_0$

**4.** Dato che $\hat{\theta} = 1.6$ e che $z_\alpha = z_{0.05} = -1.64$, si ottiene che $k = -1.64\dfrac{3.6}{12} + 1.5 = -1.64(.3) + 1.5 = -0.49 + 1.5 = 1.01$ si ha dunque che $\hat{\theta} = 1.6 > k = 1.01$, il valore di $\hat{\theta}$ non appartiene alla regione di rifiuto e si accetta $H_0$.

**Esercizio 29.**

1. Dalla ben nota formula della varianza congiunta $\hat{\sigma}_P^2$ (*pooled variance*)

$$\hat{\sigma}_P^2 = \frac{s_x^2(n_x-1) + s_y^2(n_y-1)}{n_x+n_y-2}$$

si ottiene numericamente $\hat{\sigma}_P^2 = 0.78$

```
> (0.94^2*(229-1)+0.83^2*(243-1))/(229+243-2)
[1] 0.7833502
```

2. Per la stima puntuale si ottiene immediatamente che

$$\hat{\Delta} = \bar{X} - \bar{Y} = \bar{x} - \bar{y} = 2.14 - 4.29 = -2.15$$

Ricordando che per determinare la stima per intervallo per il parametro $\Delta = \mu_x - \mu_y$ si utilizza la statistica pivotale

$$\frac{\bar{X} - \bar{Y} - \Delta}{\sqrt{\left(\frac{1}{n_x} + \frac{1}{n_y}\right)\hat{\sigma}_P^2}} \sim T_{n_x+n_y-2}$$

e che la distribuzione T di Student, per valori elevati dei gradi di libertà è ben approssimata da una distribuzione normale

$$Pr\left\{-z_{1-\frac{\alpha}{2}} \leq \frac{\bar{X}-\bar{Y}-\Delta}{\sqrt{\left(\frac{1}{n_x}+\frac{1}{n_y}\right)\hat{\sigma}_P^2}} \leq z_{1-\frac{\alpha}{2}}\right\}$$

da cui si ricava

$$\left[\bar{X}-\bar{Y}-z_{1-\frac{\alpha}{2}}\sqrt{\left(\frac{1}{n_x}+\frac{1}{n_y}\right)\hat{\sigma}_P^2}, \bar{X}-\bar{Y}+z_{1-\frac{\alpha}{2}}\sqrt{\left(\frac{1}{n_x}+\frac{1}{n_y}\right)\hat{\sigma}_P^2}\right]$$

e quindi per i valori campionari osservati

$$\left[\hat{\Delta}_{inf}, \hat{\Delta}_{sup}\right] = [-1.99, -2.31]$$

```
> (2.14-4.29)+1.96*sqrt(0.78*(1/229+1/243))
[1] -1.990576
> (2.14-4.29)-1.96*sqrt(0.78*(1/229+1/243))
[1] -2.309424
```

3. Per rispondere al quesito ricorriamo ad una procedura formale di verifica di ipotesi del sistema

$$\begin{cases} H_0: & \Delta = 0 \\ H_1: & \Delta < 0 \end{cases}$$

Per la costruzione della regione di rifiuto, ci si affida alla statistica test

$$T(X_1, ..., X_{n_x}, Y_1, ...Y_{n_y}) = \frac{\bar{X}-\bar{Y}}{\sqrt{\left(\frac{1}{n_x}+\frac{1}{n_y}\right)\hat{\sigma}_P^2}}$$

di cui è noto che, quando $\Delta = 0$ (ovvero sotto $H_0$), si distribuisce come una distribuzione di tipo $T$ di Student e precisamente $T_{n_x+n_y-2}$. La regione di rifiuto è data da

$$R = \left\{(X_1, ..., X_{n_x}, Y_1, ...Y_{n_y}) : T(X_1, ..., X_{n_x}, Y_1, ...Y_{n_y}) < t_{n_x+n_y-2,\alpha}\right\}$$

Dal momento che $n_x + n_y - 2 = 470 > 100$ la distribuzione T di Student $n_x + n_y - 2$ con gradi di libertà è molto ben approssimata da una distribuzione normale standard $Z \sim N(0, 1)$ considereremo i seguenti quantili $t_{n_x+n_y-2,\alpha} = z_\alpha = -1.64$. Dal campione osservato si ha

```
> t.oss=(2.14-4.29)/sqrt(0.78*(1/229+1/243))
> t.oss
[1] -26.43267
```

per cui $T = t_{oss} = -26.43267 < -1.64$ e dunque, ad un livello di significatività $\alpha = 0.05$, concludiamo in favore dell'ipotesi alternativa che il punteggio medio del primo docente è inferiore a quello del secondo.

4. Da quanto già argomentato è evidente che l'elevata numerosità campionaria e il teorema centrale del limite consentono di utilizzare la distribuzione normale come approssimazione della distribuzione delle media campionarie $\bar{X}$ e $\bar{Y}$ e dunque è possibile utilizzare la stessa quantità pivotale e la stessa statistica test con gli stessi quantili anche in assenza di esplicita ipotesi di normalità delle singole osservazioni campionarie.

**Esercizio 30.**

**1.**

$$\mathbb{E}[X] = \int_0^\theta x \frac{3}{\theta^3} x^2 dx = \frac{3}{\theta^3} \left[\frac{x^4}{4}\right]_0^\theta = \frac{3}{4}\theta$$

$$\mathbb{E}[X^2] = \int_0^\theta x^2 \frac{3}{\theta^3} x^2 dx = \frac{3}{\theta^3} \left[\frac{x^5}{5}\right]_0^\theta = \frac{3}{5}\theta^2$$

$$\mathbb{V}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{3}{5}\theta^2 - \left(\frac{3}{4}\theta\right)^2 = \frac{3}{80}\theta^2$$

Per determinare lo stimatore dei momenti, si eguaglia la media campionaria al valore atteso:

$$\bar{X}_n = \mathbb{E}[X] \iff \bar{X}_n = \frac{3}{4}\theta \iff \hat{\theta}_M = \frac{4}{3}\bar{X}_n.$$

**2.**

Si verifica facilmente che lo stimatore dei momenti è non distorto, quindi l'errore quadratico medio è uguale alla varianza:

$$MSE(\hat{\theta}_M) = \mathbb{V}(\hat{\theta}_M) = \mathbb{V}(\frac{4}{3}\bar{X}_n) = \frac{16}{9}\frac{\mathbb{V}(X)}{n} = \frac{16}{9}\frac{1}{n}\frac{3}{80}\theta^2 = \frac{\theta^2}{15n}$$

Poiché l'errore quadratico medio tende a 0 per $n \to \infty$ possiamo concludere che lo stimatore dei momenti è consistente.

**3.**

Ricordando che $\bar{X}_n \approx N\left(\mathbb{E}[X], \frac{\mathbb{V}[X]}{n}\right)$ (Teorema del Limite Centrale), si ricava facilmente la distribuzione asintotica dello stimatore dei momenti:

$$\hat{\theta}_M = \frac{4}{3}\bar{X}_n \approx N\left(\frac{4}{3}\mathbb{E}[X], \frac{16}{9}\frac{\mathbb{V}[X]}{n}\right) \approx N\left(\theta, \frac{\theta^2}{15n}\right)$$

Assumendo come stima della varianza $\frac{\hat{\theta}_M^2}{15n}$, possiamo trovare l'intervallo di confidenza asintotico per $\theta$:

$$\tilde{IC}_{1-\alpha}(\theta) = \hat{\theta}_M \pm z_{\frac{1-\alpha}{2}}\sqrt{\frac{\hat{\theta}_M^2}{15n}} = \frac{4}{3}\bar{X}_n \pm z_{\frac{1-\alpha}{2}}\sqrt{\frac{(\frac{4}{3}\bar{X}_n)^2}{15n}}$$

**4.**

La stima dei momenti è

$$\hat{\theta}_M = \frac{4}{3}\bar{x}_n = \frac{4}{3}\frac{\sum_{i=1}^n x_i}{n} = \frac{4}{3}\frac{3}{2}\frac{1}{36} = \frac{1}{18} = 0.056$$

e l'intervallo di confidenza asintotico a livello 0.95:

$$\tilde{IC}_{1-\alpha}(\theta) = \frac{1}{18} \pm 1.96\sqrt{\frac{(1/18)^2}{15 \cdot 36}} = [0.051; 0.06]$$

**Esercizio 31.**

**1.**

Applicando il Lemma di Neyman-Pearson, la regione di accettazione sarà del tipo

$$A = \{\mathbf{x} \in \mathcal{X}^n : \quad \frac{L(\theta_0; \mathbf{x})}{L(\theta_1; \mathbf{x})} \geq c\}.$$

Per il modello considerato la verosimiglianza risulta essere

$$L(\theta;\mathbf{x}) = \prod_{i=1}^{n} \left(\theta^2 x_i e^{-\theta x_i}\right) \propto \theta^{2n} exp\left\{-\theta \sum_{i=1}^{n} x_i\right\}$$

e quindi si ha:

$$\frac{L(\theta_0;\mathbf{x})}{L(\theta_1;\mathbf{x})} = \frac{\theta_0^{2n} exp\{-\theta_0 \sum_{i=1}^{n} x_i\}}{\theta_1^{2n} exp\{-\theta_1 \sum_{i=1}^{n} x_i\}} = \left(\frac{\theta_0}{\theta_1}\right)^{2n} exp\left\{-(\theta_0-\theta_1)\sum_{i=1}^{n} x_i\right\}.$$

Dal momento che $-(\theta_0 - \theta_1) < 0$ il rapporto delle verosimiglianze è funzione decrescente di $\sum_{i=1}^{n} x_i$ (o equivalentemente di $\bar{x}_n$) e dunque possiamo concludere che la regione di accettazione è del tipo

$$A = \{\mathbf{x} \in \mathcal{X}^n : \quad \sum_{i=1}^{n} x_i < k'\} = \{\mathbf{x} \in \mathcal{X}^n : \quad \bar{x}_n < k\}$$

**2.**

$$\bar{X}_n \approx N\left(\mathbb{E}[X], \frac{\mathbb{V}[X]}{n}\right) \approx N\left(\frac{2}{\theta}, \frac{2}{n \cdot \theta^2}\right)$$

**3.**
Fissando $\alpha$ e utilizzando la regione di accettazione di cui al punto 1 si ha

$$\alpha = P(\mathbf{X} \in R = A^c | H_0) = P(\bar{X}_n > k | \theta = \theta_0).$$

Sotto $H_0$ vale la distribuzione asintotica di cui al punto 2, e standardizzando si ha:

$$\alpha = P\left(\frac{\bar{X}_n - \frac{2}{\theta_0}}{\sqrt{\frac{2}{n\theta_0^2}}} > \frac{k - \frac{2}{\theta_0}}{\sqrt{\frac{2}{n\theta_0^2}}}\right) = 1 - P\left(Z < \frac{k - \frac{2}{\theta_0}}{\sqrt{\frac{2}{n\theta_0^2}}}\right) \Longleftrightarrow 1 - \alpha = \Phi\left(\frac{k - \frac{2}{\theta_0}}{\sqrt{\frac{2}{n\theta_0^2}}}\right)$$

Applicando la funzione inversa alla funzione di ripartizione si ottiene:

$$\frac{k - \frac{2}{\theta_0}}{\sqrt{\frac{2}{n\theta_0^2}}} = z_{1-\alpha}$$

da cui si ricava

$$k_\alpha = \frac{2}{\theta_0} + z_{1-\alpha}\sqrt{\frac{2}{n\theta_0^2}}.$$

Per $\alpha = 0.05$, $\theta_0 = 2$ e $n = 25$, si ha

$$k_\alpha = 1 + 1.64\sqrt{\frac{2}{25 \cdot 4}} = 1.23.$$

**4.** La potenza, in corrispondenza dell'ipotesi alternativa semplice $\theta_1 = 1$, risulta essere

$$1 - \beta = P(\mathbf{X} \in R = A^c | H_1) = P(\bar{X}_n > k_\alpha | \theta = \theta_1) = 1 - \Phi\left(\frac{1.23 - \frac{2}{1}}{\sqrt{\frac{2}{25}}}\right) = 1 - \Phi(-2.72) = 0.997$$

**5.** Poiché il valore osservato della media campionaria $\bar{x}_{25} = 1 < k_\alpha = 1.23$ cade nella regione di accettazione, non si può rifiutare l'ipotesi nulla a livello $\alpha = 0.05$.

**Esercizio 32.**

**1.**

$$\mathbb{E}[X] = \int_0^\theta x \frac{1}{2\sqrt{\theta}} \frac{1}{\sqrt{x}} dx = \frac{1}{2\theta^{\frac{1}{2}}} \left[ \frac{x^{\frac{3}{2}}}{\frac{3}{2}} \right]_0^\theta = \frac{\theta}{3}$$

$$\mathbb{E}[X^2] = \int_0^\theta x^2 \frac{1}{2\sqrt{\theta}} \frac{1}{\sqrt{x}} dx = \frac{1}{2\theta^{\frac{1}{2}}} \left[ \frac{x^{\frac{5}{2}}}{\frac{5}{2}} \right]_0^\theta = \frac{\theta^2}{5}$$

$$\mathbb{V}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{\theta^2}{5} - \left(\frac{\theta}{3}\right)^2 = \frac{4}{45}\theta^2$$

Per determinare lo stimatore dei momenti, si eguaglia la media campionaria al valore atteso:

$$\bar{X}_n = \mathbb{E}[X] \iff \bar{X}_n = \frac{\theta}{3} \iff \hat{\theta}_M = 3\bar{X}_n$$

**2.** Si verifica facilmente che lo stimatore dei momenti è non distorto, quindi l'errore quadratico medio è uguale alla varianza:

$$MSE(\hat{\theta}_M) = \mathbb{V}(\hat{\theta}_M) = \mathbb{V}(3\bar{X}_n) = 9\frac{\mathbb{V}(X)}{n} = 9\frac{4}{45}\frac{1}{n}\theta^2 = \frac{4\theta^2}{5n}$$

Poiché l'errore quadratico medio tende a 0 per $n \to \infty$ possiamo concludere che lo stimatore dei momenti è consistente.

**3.** Ricordando che $\bar{X}_n \approx N\left(\mathbb{E}[X], \frac{\mathbb{V}[X]}{n}\right)$ (Teorema del Limite Centrale), si ricava facilmente la distribuzione asintotica dello stimatore dei momenti:

$$\hat{\theta}_M = 3\bar{X}_n \approx N\left(3\mathbb{E}[X], 9\frac{\mathbb{V}[X]}{n}\right) \approx N\left(\theta, \frac{4\theta^2}{5n}\right)$$

Assumendo come stima della varianza $\frac{4\hat{\theta}_M^2}{5n}$, possiamo trovare l'intervallo di confidenza asintotico per $\theta$:

$$\tilde{IC}_{1-\alpha}(\theta) = \hat{\theta}_M \pm z_{\frac{1-\alpha}{2}}\sqrt{\frac{4\hat{\theta}_M^2}{5n}} = 3\bar{X}_n \pm z_{\frac{1-\alpha}{2}}\sqrt{\frac{4(3\bar{X}_n)^2}{5n}} =$$

**4.** La stima dei momenti

$$\hat{\theta}_M = 3\bar{x}_n = 3\frac{\sum_{i=1}^n x_i}{n} = 3\frac{1}{3}\frac{1}{36} = \frac{1}{36} = 0.028$$

e l'intervallo di confidenza asintotico a livello 0.95:

$$\tilde{IC}_{1-\alpha}(\theta) = \frac{1}{36} \pm 1.96\sqrt{4\frac{(1/36)^2}{5 \cdot 36}} = [0.0197; 0.0359]$$

**Esercizio 33.**

**1.** Applicando il Lemma di Neyman-Pearson, la regione di accettazione sarà del tipo

$$A = \{\mathbf{x} \in \mathcal{X}^n : \quad \frac{L(\theta_0; \mathbf{x})}{L(\theta_1; \mathbf{x})} \geq c\}.$$

Per il modello considerato la verosimiglianza risulta essere

$$L(\theta;\mathbf{x})=\prod_{i=1}^{n}\left(\theta e^{-\theta x_i}\right)\propto \theta^n exp\left\{-\theta\sum_{i=1}^{n}x_i\right\}$$

e quindi si ha:

$$\frac{L(\theta_0;\mathbf{x})}{L(\theta_1;\mathbf{x})}=\frac{\theta_0^n exp\{-\theta_0\sum_{i=1}^{n}x_i\}}{\theta_1^n exp\{-\theta_1\sum_{i=1}^{n}x_i\}}=\left(\frac{\theta_0}{\theta_1}\right)^n exp\left\{-(\theta_0-\theta_1)\sum_{i=1}^{n}x_i\right\}.$$

Dal momento che $-(\theta_0-\theta_1)<0$ il rapporto delle verosimiglianze è funzione decrescente di $\sum_{i=1}^{n}x_i$ (o equivalentemente di $\bar{x}_n$) e dunque possiamo concludere che la regione di accettazione è del tipo

$$A=\{\mathbf{x}\in\mathcal{X}^n:\quad \sum_{i=1}^{n}x_i<k'\}=\{\mathbf{x}\in\mathcal{X}^n:\quad \bar{x}_n<k\}$$

**2.**

$$\bar{X}_n\approx N\left(\mathbb{E}[X],\frac{\mathbb{V}[X]}{n}\right)\approx N\left(\frac{1}{\theta},\frac{1}{n\cdot\theta^2}\right)$$

**3.** Fissando $\alpha$ e utilizzando la regione di accettazione di cui al punto 1 si ha

$$\alpha=P(\mathbf{X}\in R=A^c|H_0)=P(\bar{X}_n>k|\theta=\theta_0).$$

Sotto $H_0$ vale la distribuzione asintotica di cui al punto 2, e standardizzando si ha:

$$\alpha=P\left(\frac{\bar{X}_n-\frac{1}{\theta_0}}{\sqrt{\frac{1}{n\theta_0^2}}}>\frac{k-\frac{1}{\theta_0}}{\sqrt{\frac{1}{n\theta_0^2}}}\right)=1-P\left(Z<\frac{k-\frac{1}{\theta_0}}{\sqrt{\frac{1}{n\theta_0^2}}}\right)\Longleftrightarrow 1-\alpha=\Phi\left(\frac{k-\frac{1}{\theta_0}}{\sqrt{\frac{1}{n\theta_0^2}}}\right)$$

Applicando la funzione inversa alla funzione di ripartizione si ottiene:

$$\frac{k-\frac{1}{\theta_0}}{\sqrt{\frac{1}{n\theta_0^2}}}=z_{1-\alpha}$$

da cui si ricava

$$k_\alpha=\frac{1}{\theta_0}+z_{1-\alpha}\sqrt{\frac{1}{n\theta_0^2}}.$$

Per $\alpha=0.05$, $\theta_0=2$ e $n=25$, si ha

$$k_\alpha=\frac{1}{2}+1.64\sqrt{\frac{1}{25\cdot 4}}=0.664.$$

**4.** La potenza, in corrispondenza dell'ipotesi alternativa semplice $\theta_1=1$, risulta essere

$$1-\beta=P(\mathbf{X}\in R=A^c|H_1)=P(\bar{X}_n>k_\alpha|\theta=\theta_1)=1-\Phi\left(\frac{0.664-1}{\sqrt{\frac{1}{25}}}\right)=1-\Phi(-1.68)=0.9535$$

**5.** Poiché il valore osservato della media campionaria $\bar{x}_{25}=1.5>k_\alpha=0.664$ cade nella regione di rifiuto, si rifiuta l'ipotesi nulla a livello $\alpha=0.05$.

**Esercizio 34.**

1. Lo stimatore è non distorto in quanto, per le proprietà del valore atteso, si ha:
$$\mathbb{E}[T] = \frac{\mathbb{E}[\bar{X}_1] + \mathbb{E}[\bar{X}_2]}{2} = \frac{\theta_1 + \theta_2}{2} = \theta, \qquad \forall \theta \in \mathbb{R}.$$
Dall'ipotesi di indipendenza dei due campioni discende che
$$\mathbb{V}\left[\frac{\bar{X}_1 + \bar{X}_2}{2}\right] = \frac{1}{4}(\mathbb{V}[\bar{X}_1] + \mathbb{V}[\bar{X}_2]) = \frac{1}{4}\left(\frac{1}{n_1} + \frac{1}{n_2}\right).$$
Pertanto, essendo $T$ uno stimatore non distorto di $\theta$, si ha che
$$MSE[T] = \mathbb{V}[T] = \frac{1}{4}\left(\frac{1}{n_1} + \frac{1}{n_2}\right).$$
Lo stimatore è consistente poichè, al divergere di $n_1$ e $n_2$, $MSE[T] \to 0, \qquad \forall \theta \in \mathbb{R}$.

2. Per le proprietà delle medie campionarie nei modelli normali, si ha che
$$\bar{X}_i|\theta_i \sim N\left(\theta_i, \frac{1}{n_i}\right), \qquad i = 1, 2.$$
Ricordando che una combinazione lineare di v.a. normali ha distribuzione normale, abbiamo che:
$$\frac{\bar{X}_1 + \bar{X}_2}{2} \sim N\left(\theta, \frac{1}{4}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right).$$

3. Per la normalità dello stimatore $\frac{\bar{X}_1+\bar{X}_2}{2}$ si ha che, per $\alpha \in (0, 1)$,
$$1 - \alpha = \mathbb{P}\left(-z_{1-\frac{\alpha}{2}} < \frac{\frac{\bar{X}_1+\bar{X}_2}{2} - \theta}{\sqrt{\frac{1}{4}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} < z_{1-\frac{\alpha}{2}}\right),$$
dove $z_{1-\frac{\alpha}{2}}$ è il percentile di livello $1 - \frac{\alpha}{2}$ della v.a. $N(0, 1)$. Con gli usuali passaggi si può quindi facilmente verificare che l'intervallo aleatorio
$$\left(\frac{\bar{X}_1 + \bar{X}_2}{2} - z_{1-\frac{\alpha}{2}}\frac{1}{2}\sqrt{\frac{n_1 + n_2}{n_1 n_2}}, \frac{\bar{X}_1 + \bar{X}_2}{2} + z_{1-\frac{\alpha}{2}}\frac{1}{2}\sqrt{\frac{n_1 + n_2}{n_1 n_2}}\right),$$
è un intervallo di confidenza di livello $1 - \alpha$ per $\theta$. Per $\alpha = 0.05$ si ha inoltre che $z_{1-\frac{\alpha}{2}} = z_{0.975} = \texttt{qnorm}(0.975) = 1.96$. L'intervallo osservato risulta essere:
$$\left(\frac{2+3}{2} - 1.96\frac{1}{2}\sqrt{\frac{2}{10}}, \frac{2+3}{2} + 1.96\frac{1}{2}\sqrt{\frac{2}{10}}\right) = (2.19, 2.80).$$

**Esercizio 35.**

1. Per le propr. di valore atteso e varianza e dal suggerimento dato, si ha che:
$$\mathbb{E}[X_{(n)}] = \frac{n}{n+1}\theta, \qquad \mathbb{V}[X_{(n)}] = \frac{n}{(n+1)^2(n+2)}\theta^2.$$
Si ha quindi che, $\forall \theta \in (0, 1)$:
   - $\mathbb{E}[T_1] = \theta$;

- $\mathbb{V}[T_1] = MSE[T_1] = \frac{\theta^2}{n(n+2)}$;
- $\mathbb{E}[T_2] = \frac{n(n+2)}{(n+1)^2}\theta$;
- $\mathbb{V}[T_2] = \frac{n(n+2)}{(n+1)^4}\theta^2$;
- $MSE[T_2] = (\mathbb{E}[T_2] - \theta)^2 + \mathbb{V}[T_2] = \ldots = \frac{\theta^2}{(n+1)^2}$.

Pertanto:

- $T_1$ è stimatore non distorto di $\theta$ e funzione di statistica sufficiente e completa, $X_{(n)}$. Si tratta quindi dello stimatore non distorto di minima varianza (UMVUE).
- $T_1$ è consistente in errore quadratico medio, dal momento che, $\forall\theta \in (0,1)$,

$$\lim_{n\to+\infty} MSE[T_1] = \lim_{n\to+\infty} \frac{\theta^2}{n(n+2)} = 0.$$

- Lo stimatore $T_2$ è stimatore distorto di $\theta$ con distorsione negativa e pari a

$$B[T_2] = -\frac{\theta}{(n+1)^2}.$$

  Tuttavia lo stimatore risulta consistente (e dunque asintoticamente corretto), in quanto $\forall\theta \in (0,1)$,

$$\lim_{n\to+\infty} MSE[T_2] = \lim_{n\to+\infty} \frac{\theta^2}{(n+1)^2} = 0.$$

- Lo stimatore $T_2$ è più efficiente di $T_1$ poichè:

$$MSE[T_2] < MSE[T_1] \Leftrightarrow (n+1)^2 > n(n+2) \Leftrightarrow 1 > 0,$$

  condizione che risulta essere ovviamente verificata per ogni valore di $n$ e di $\theta$.

Da quanto verificato si evince che, in base al criterio delle errore quadratico medio, lo stimatore $T_2$ benchè distorto, è migliore dello stimatore non distorto $T_1$ (peraltro anche UMVUE).

2. Poichè il primo momento di $X$ è $\mu_1(\theta) = \mathbb{E}[X] = \frac{\theta}{2}$ e il primo momento campionario è $m_1(\mathbf{X}_n) = \bar{X}_n$, l'equazione dei momenti($\mu_1(\theta) = m_1(\mathbf{X}_n)$) diventa:

$$\frac{\theta}{2} = \bar{X}_n,$$

da cui si ottiene lo stimatore dei momenti:

$$\hat{\theta}_M = 2\bar{X}_n.$$

Per le proprietà della media campionaria ed il Dal momento che $X_1, \ldots, X_n$ sono v.a. i.i.d. con valore atteso e varianza finiti vale il teorema centrale di convergenza e si ha che $\bar{X}_n$ ha distribuzione asintotica normale. Pertanto, osservando che

$$\mathbb{E}[\hat{\theta}_M] = \mathbb{E}[2\bar{X}_n] = \theta, \qquad \mathbb{V}[\hat{\theta}_M] = \mathbb{V}[2\bar{X}_n] = 4\mathbb{V}[\bar{X}_n] = 4\frac{\theta^2}{12n} = \frac{\theta^2}{3n},$$

possiamo affermare che, per un qualsiasi $\theta$,

$$\frac{\hat{\theta}_M - \theta}{\sqrt{\mathbb{V}[\hat{\theta}_M]}} \approx N(0,1)$$

ovvero che

$$\hat{\theta}_M = 2\bar{X}_n \approx N\left(\theta, \frac{\theta^2}{3n}\right).$$

Da quanto appena determinato discende che, per un qualsiasi $\alpha \in (0,1)$,

$$1-\alpha \approx \mathbb{P}\left(-z_{1-\frac{\alpha}{2}} < \frac{\hat{\theta}_M - \theta}{\sqrt{\mathbb{V}[\hat{\theta}_M]}} < z_{1-\frac{\alpha}{2}}\right),$$

dove $z_{1-\frac{\alpha}{2}}$ indica il percentile della v.a. $N(0,1)$. Pertanto, sostituendo a $\mathbb{V}[\hat{\theta}_M] = \frac{\theta^2}{3n}$ il suo stimatore,

$$\hat{\mathbb{V}}[\hat{\theta}_M] = \frac{\hat{\theta}_M^2}{3n} = \frac{4\bar{X}_n^2}{3n},$$

l'intervallo di confidenza asintotico per $\theta$ è

$$\hat{\theta}_M \pm z_{1-\frac{\alpha}{2}}\sqrt{\hat{\mathbb{V}}[\hat{\theta}_M]},$$

ovvero

$$2\bar{X}_n \pm z_{1-\frac{\alpha}{2}}\frac{2}{\sqrt{3n}}\bar{X}_n.$$

3. Sostituendo i valori indicati, si ottiene:

$$T_1 = \frac{21}{20} * 0.9, \quad T_2 = \frac{22}{21} * 0.9, \quad \hat{\theta}_M = 2 * 0.4 = 0.8.$$

Poichè $z_{1-\frac{\alpha}{2}} = z_{0.95} =$ `qnorm`$(0.95) = 1.65$, l'intervallo osservato risulta essere:

$$0.8 \pm 1.645\,\frac{2 * 0.4}{\sqrt{3 * 20}} = (0.66, 0.94).$$

Si noti che il dato fornito relativamente alla varianza campionaria $S_n^2$ risulta inutile ai fini della soluzione al quesito posto.

**Esercizio 36.**

1. La funzione di verosimiglianza di $\theta$ è

$$L(\theta) = \prod_{i=1}^{n}\left(\theta x_i^{-1} x_i^{\theta}\right) \propto \theta^n \left(\prod_{i=1}^{n} x_i\right)^{\theta}.$$

La funzione di log-verosimiglianza risulta quindi essere:

$$\ell(\theta) = \ln L(\theta) = c + n\ln\theta + \theta\sum_{i=1}^{n}\ln x_i.$$

L'equazione di log-verosimiglianza

$$\ell'(\theta) = 0$$

è quindi:

$$\frac{n}{\theta} + \sum_{i=1}^{n}\ln x_i = 0,$$

da cui si ottiene

$$\hat{\theta}_{MV} = -\frac{n}{\sum_{i=1}^{n}\ln x_i}.$$

(NB: si tratta proprio del punto di massimo in quanto $\ell''(\theta) = -n\theta^{-2} < 0, \forall\theta > 0$).

2. La funzione di potenza, per definizione, è:

$$\eta(\theta) = \mathbb{P}[R;\theta] = \mathbb{P}[X > 1/2;\theta] = \int_{1/2}^{1} \theta x^{\theta-1} dx = \left[x^{\theta}\right]_{1/2}^{1} = 1 - \frac{1}{2^{\theta}}.$$

Si ha quindi che:

$$\eta(2) = 1 - \frac{1}{2^2} = 1 - \frac{1}{4} = \frac{3}{4}.$$