

Esercitazioni di probabilità e statistica

Estimatori puntuali, stima intervallare

Correttezza stimatore:

$$E\{\hat{\theta}\} = \theta$$

Correttezza asintotica stimatore:

$$\lim_{n \rightarrow \infty} E\{\hat{\theta}\} = \theta$$

Consistenza stimatore:

$$\lim_{n \rightarrow \infty} Var\{\hat{\theta}\} = 0$$

Esercizio 4.1:

Sia X una v.a. uniforme su $(0, a)$ con a incognita e sia $X_1 \dots X_n$ un campione casuale di X .

- Data la seguente v.a. $Z = \max\{X_1 \dots X_n\}$ determinarne la densità di probabilità.
- Verificare se Z è uno stimatore corretto di a .
- Verificare se Z è uno stimatore consistente di a .

Soluzione: Per prima cosa dobbiamo costruirci la cumulativa della variabile uniforme X :

$$F_X(x) = \begin{cases} 0 & x \leq 0 \\ \frac{x}{a} & 0 < x < a \\ 1 & x \geq a \end{cases}$$

Per prima cosa ci viene chiesto di valutare la densità $f_Z(z)$ della variabile aleatoria Z .

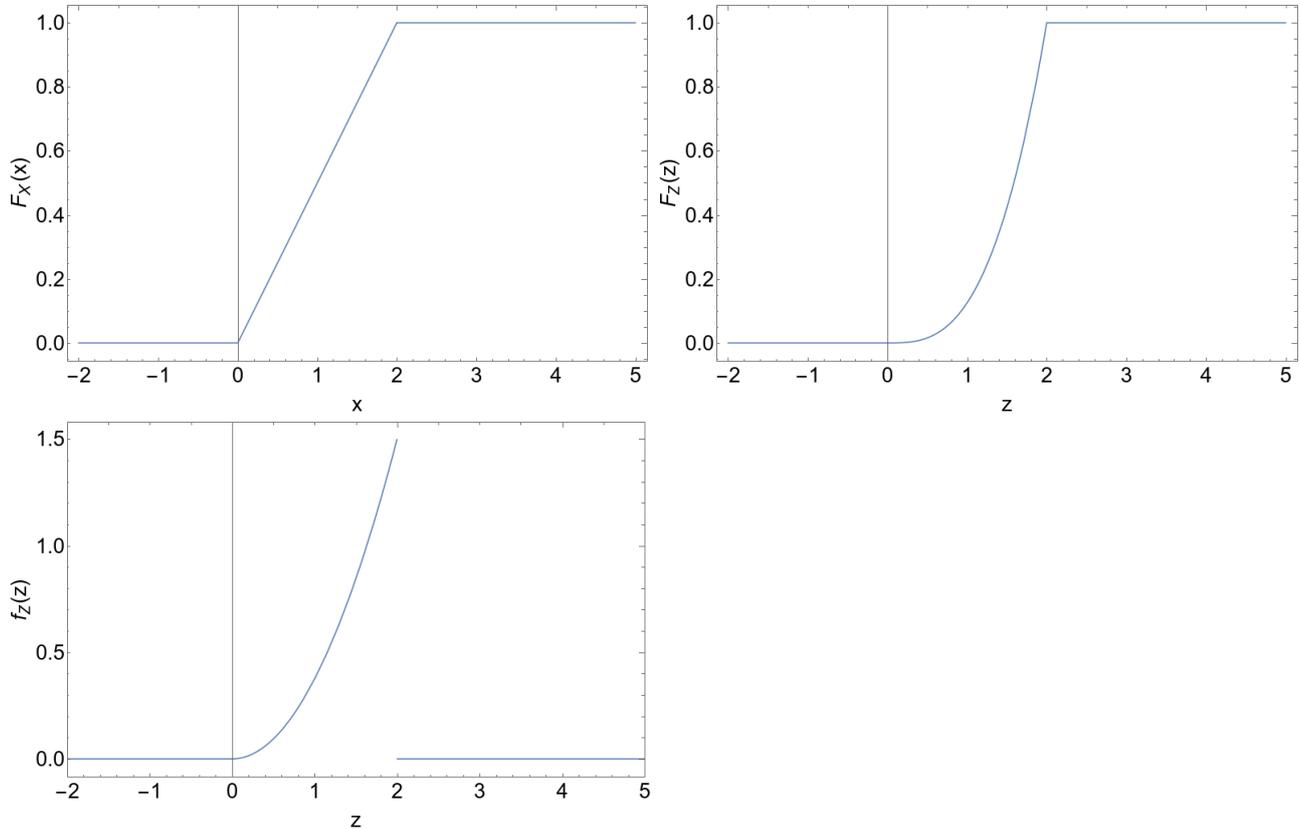
$$\begin{aligned} F_Z(z) &= \Pr\{Z \leq z\} \\ &= \Pr\{(X_1 \leq z) \cap \dots \cap (X_n \leq z)\} \\ &= F_X(z)^n \end{aligned}$$

quindi per quanto riguarda la v.a. Z abbiamo:

$$F_Z(z) = \begin{cases} 0 & z \leq 0 \\ \left(\frac{z}{a}\right)^n & 0 < z < a \\ 1 & z \geq a \end{cases}$$

Quindi infine la densità può essere scritta come:

$$f_Z(z) = \begin{cases} 0 & z \leq 0 \\ \frac{nz^{n-1}}{a^n} & 0 < z < a \\ 0 & z \geq a \end{cases}$$



b) Controlliamo se è uno stimatore corretto:

$$\begin{aligned} E\{Z\} &= \int_0^a z \frac{nz^{n-1}}{a^n} dz \\ &= \frac{n}{a^n} \int_0^a z^n dz \\ &= \frac{n}{a^n} \frac{a^{n+1}}{n+1} \\ &= \frac{n}{n+1} a \end{aligned}$$

non è corretto.

c) Verifichiamo se è consistente calcolando la varianza: Per calcolare la varianza di uno stimatore si usa la formula:

$$\begin{aligned} Var\{\hat{\theta}\} &= E\{(\hat{\theta} - E\{\hat{\theta}\})^2\} \\ &= E\{\hat{\theta}^2\} - E\{\hat{\theta}\}^2 \end{aligned}$$

Quindi iniziamo dal momento secondo dello stimatore:

$$\begin{aligned}
 E\{Z^2\} &= \int_0^a z^2 \frac{n z^{n-1}}{a^n} dz \\
 &= \frac{n}{a^n} \int_0^a z^n + 1 dz \\
 &= \frac{n}{a^n} \frac{a^{n+2}}{n+2} \\
 &= \frac{n}{n+2} a^2
 \end{aligned}$$

e quindi la varianza è:

$$\begin{aligned}
 Var\{Z\} &= \frac{n}{n+2} a^2 - \frac{n^2}{(n+1)^2} a^2 \\
 &\rightarrow_{n \rightarrow \infty} 0
 \end{aligned}$$

quindi Z è consistente.

Esercizio 4.2:

Sia (X_1, \dots, X_n) un campione casuale di una v.a. gaussiana X avente parametri μ σ ignoti. La seguente v.a. è uno stimatore di μ :

$$\hat{X} = \frac{X_1}{2} + \frac{1}{2(n-1)} \sum_{j=2}^n X_j$$

- Dire se lo stimatore è corretto.
- Che tipo di variabile aleatoria è \hat{X} ?
- Dire se lo stimatore è consistente.

Soluzione: Valutiamo la media dello stimatore e vediamo cosa otteniamo:

$$\begin{aligned}
 E\{\hat{X}\} &= \frac{1}{2} E\{X_1\} + \frac{1}{2(n-1)} \sum_{j=2}^n E\{X_j\} \\
 &= \frac{\mu}{2} + \frac{1}{2(n-1)} (n-1)\mu \\
 &= \frac{\mu}{2} + \frac{\mu}{2} \\
 &= \mu
 \end{aligned}$$

- \hat{X} è gaussiana con media μ .
- Calcoliamoci la varianza:

$$\begin{aligned}
 Var\{\hat{X}\} &= \frac{1}{4} Var\{X_1\} + \frac{1}{4(n-1)^2} \sum_{j=2}^n Var\{X_j\} \\
 &= \frac{\sigma^2}{4} + \frac{1}{4(n-1)^2} (n-1)\sigma^2 \\
 &= \frac{\sigma^2}{4} \left(1 + \frac{1}{n-1} \right) \\
 &\rightarrow_{\lim n \rightarrow \infty} \frac{\sigma^2}{4} \neq 0
 \end{aligned}$$

E' diverso da zero quindi non è consistente.

Esercizio 4.3:

Si tengono, indipendentemente tra loro, 100 osservazioni di una v.a. esponenziale di parametro incognito λ . E' risultato che 70 tra esse si trovano nell'intervallo $[0, 6]$ e le rimanenti 30 sono maggiori di 6. Determinare una stima di λ .

Soluzione: Ci sono 2 modi per affrontare il problema, il primo è intuitivamente:

$$\begin{aligned} F_X(6) &= \Pr\{X \leq 6\} \\ &\simeq \frac{70}{100} \\ &\rightarrow 1 - e^{-\lambda 6} = 0.7 \\ &= e^{-\lambda 6} = 0.3 \\ \lambda &= -\frac{1}{6} \ln(0.3) \\ &= 0.2 \end{aligned}$$

Si può invece seguire la strada più formale, dividiamo gli eventi in modo che i primi 70 diano risultato compreso tra 0 e 6, mentre i rimanenti 30 diano risultato > 6 . Su questo costruiamo una funzione di verosimiglianza della quale andremo a calcolare il massimo:

$$\Pr\{(X_1 \leq 6 \cap \dots \cap (X_{70} \leq 6)) \cap (X_{71} > 6) \cap \dots \cap (X_{100} > 6)\} = F_X(6)^{70} [1 - F_X(6)]^{30}$$

Cosa vuol dire costruire una Likelihood, vuol dire prendere le osservazioni a disposizione e creare una funzione che rappresenti la distribuzione di quelle osservazioni. Questa può esser usata all'interno della definizione di probabilità oppure può essere studiata per ricavarne le proprietà. In questo caso la verosimiglianza ha la seguente forma:

$$\begin{aligned} L(\lambda, X_n) &= (1 - e^{-\lambda 6})^{70} (e^{-\lambda 6})^{30} \\ &\text{Quindi la derivata della Likelihood:} \\ \frac{dL}{d\lambda} &= 70 (1 - e^{-\lambda 6})^{69} 6 e^{-\lambda 6} - 6 \cdot 30 (1 - e^{-\lambda 6})^{70} e^{-\lambda 6 \cdot 30} = 0 \\ &= 70 \cdot 6 \cdot e^{-\lambda 6} - 6 \cdot 30 \cdot (1 - e^{-\lambda 6}) \\ \frac{7}{3} &= e^{\lambda 6} - 1 \\ e^{\lambda 6} &= \frac{10}{3} \\ \lambda &= -\frac{1}{6} \ln 0.3 \\ &= 0.2 \end{aligned}$$

Esercizio 4.4:

sia X una variabile aleatoria gaussiana e sia (X_1, \dots, X_n) un campione casuale di X . Determinare con il metodo della massima verosimiglianza, uno stimatore della media e della varianza. Verificarne la consistenza.

Soluzione: La v.a. gaussiana ha densità di probabilità:

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Lo scopo di MLE è di trovare il modo ottimale in cui i dati si distribuiscono. Il primo passo è costruire la funzione di Likelihood:

$$\begin{aligned} L(\hat{\mu}, \hat{\sigma}, X_1, \dots, X_n) &= \prod_{j=1}^n \frac{1}{\hat{\sigma}\sqrt{2\pi}} e^{-\frac{(x_j - \hat{\mu})^2}{2\hat{\sigma}^2}} \\ &= (\hat{\sigma}\sqrt{2\pi})^{-n} \prod_{j=1}^n e^{-\frac{(x_j - \hat{\mu})^2}{2\hat{\sigma}^2}} \end{aligned}$$

A questo punto possiamo definire lo stimatore

$$\begin{aligned} \ell &= \ln L \\ &= -n \ln \hat{\sigma} - \frac{n}{2} \ln(2\pi) - \frac{1}{2\hat{\sigma}^2} \sum_{j=1}^n (x_j - \hat{\mu})^2 \end{aligned}$$

A questo punto dobbiamo massimizzare lo stimatore, quindi guardiamo quando la derivata è nulla:

$$\begin{aligned} \frac{\partial L}{\partial \hat{\mu}} &= \frac{1}{2\hat{\sigma}^2} 2 \sum_{j=1}^n (x_j - \hat{\mu}) = 0 \\ \sum_{j=1}^n x_j - n \hat{\mu} &= 0 \\ \hat{\mu} &= \frac{1}{n} \sum_{j=1}^n x_j \\ &= \bar{x} \end{aligned}$$

Valutiamo la correttezza e la consistenza:

$$\begin{aligned} E\{\hat{\mu}\} &= \frac{1}{n} \sum_{j=1}^n E\{X_j\} \\ &= \frac{1}{n} n\mu \\ &= \mu \end{aligned}$$

è corretta.

$$\begin{aligned} Var\{\hat{\mu}\} &= \frac{1}{n^2} \sum_{j=1}^n Var\{X_j\} \\ &= \frac{\sigma^2}{n} \\ \lim_{n \rightarrow \infty} &\rightarrow 0 \end{aligned}$$

b) Facciamo la stessa cosa con la varianza, usando la medesima formula ottenuta per la Likelihood cerchiamo la derivata rispetto alla varianza:

$$\begin{aligned} \frac{\partial L}{\partial \hat{\sigma}} &= -\frac{n}{\hat{\sigma}} + \frac{1}{\hat{\sigma}^3} \sum_{j=1}^n (x_j - \hat{\mu})^2 = 0 \\ \hat{\sigma}^2 &= \frac{1}{n} \sum_{j=1}^n (x_j - \bar{x})^2 \end{aligned}$$

Valutiamo ora la consistenza:

$$\begin{aligned}
 E\{\hat{\sigma}^2\} &= \frac{1}{n} \sum_{j=1}^n E\{(x_j - \bar{x})^2\} \\
 &= \frac{1}{n} \sum_{j=1}^n E\{[(x_j - \mu) - (\bar{x} - \mu)]^2\} \\
 &= \frac{1}{n} \sum_{j=1}^n E\{(x_j - \mu)^2 + (\bar{x} - \mu)^2 - 2(\bar{x} - \mu)(x_j - \mu)\} \\
 &= \frac{1}{n} \sum_{j=1}^n E\{(x_j - \mu)^2\} + \frac{1}{n} \sum_{j=1}^n E\{(\bar{x} - \mu)^2\} - 2E\left\{\frac{1}{n} \sum_{j=1}^n (\bar{x} - \mu)(x_j - \mu)\right\} \\
 &= \frac{1}{n} \sum_{j=1}^n E\{(x_j - \mu)^2\} + \frac{1}{n} \sum_{j=1}^n E\{(\bar{x} - \mu)^2\} - 2E\{(x_j - \mu)(x_j - \mu)\} \\
 &= \sigma^2 - E\{(x_j - \mu)^2\} \\
 &= \sigma^2 \frac{\sigma^2}{n} \\
 &= \frac{n-1}{n} \sigma^2
 \end{aligned}$$

é uno stimatore distorto, se lo dividessi per $n - 1$ invece che per n allora sarebbe corretto.

Esercizio 4.5:

Per riordinare un archivio si separano i documenti antecedenti all'anno 2000 da quelli successivi. I documenti, in ordine temporale casuale, sono in totale 1000 e vengono esaminati a lotti di 20 alla volta, per un totale di 50 lotti. Il numero dei documenti successivi al 2000 in ogni lotto é una v.a. con media 7 e varianza 4.

- Determinare uno stimatore della probabilità p che un documento sia successivo al 2000.
- Verificare correttezza e consistenza dello stimatore ottenuto.

Soluzione: Definiamo la v.a. in modo da poter scrivere esplicitamente le quantità riportate nel testo, nello specifico media e varianza.

$$Y_j = \# \text{ documenti datati post 2000 nel } j\text{-esimo lotto} \quad j = 1, \dots, 50$$

quindi la media e la varianza:

$$\bar{Y} = \frac{1}{n} \sum_{j=1}^n Y_j, \quad \sigma_Y^2 = \sum_{j=1}^n \frac{(Y_j - \bar{Y})^2}{n}$$

Ora ragioniamo sul tipo di distribuzione che la v.a. segue, abbiamo una successione di prove indipendenti che possono risultare in 2 possibilità equiprobabili e mutualmente esclusive, quindi usiamo una binomiale. La probabilità nel j -esimo lotto è:

$$\Pr\{Y_j\} = \binom{m}{y_j} p^{y_j} (1-p)^{m-y_j}$$

Creiamo una funzione di verosimiglianza:

$$L(\hat{p}, y_1, \dots, y_n) = \prod_{j=1}^n \binom{m}{y_j} \hat{p}^{y_j} (1-\hat{p})^{m-y_j}$$

come abbiamo imparato ricaviamo il logaritmo della Likelihood:

$$\begin{aligned}\ell &= \ln L \\ &= \sum_{j=1}^n \left[\ln \binom{m}{y_j} + y_j \ln(\hat{p}) + (m - y_j) \ln(1 - \hat{p}) \right]\end{aligned}$$

Ora cerchiamo il massimo:

$$\begin{aligned}\frac{d\ell}{d\hat{p}} &= \frac{1}{\hat{p}} \sum_{j=1}^n y_j - \frac{1}{1 - \hat{p}} \sum_{j=1}^n (m - y_j) = 0 \\ (1 - \hat{p}) \sum_{j=1}^n y_j &= \hat{p} \sum_{j=1}^n (m - y_j) \\ \sum_{j=1}^n y_j - \hat{p} \sum_{j=1}^n y_j &= \hat{p} m n - \hat{p} \sum_{j=1}^n y_j \\ \hat{p} &= \frac{1}{m n} \sum_{j=1}^n y_j \\ &= \frac{\bar{y}}{m} \\ &= \frac{7}{20} \\ &= 0.35\end{aligned}$$

b) Valutiamo ora correttezza e consistenza dello stimatore \hat{p} .

$$\begin{aligned}E\{\hat{p}\} &= \frac{1}{m n} \sum_{j=1}^n E\{y_j\} \\ &= \frac{1}{m n} \sum_{j=1}^n m p \\ &= \frac{m n}{m n} p \\ &= p\end{aligned}$$

l'estimatore è corretto

$$\begin{aligned}Var\{\hat{p}\} &= Var\left\{ \frac{1}{m n} \sum_{j=1}^n y_j \right\} \\ &= \frac{1}{(m n)^2} \sum_{j=1}^n Var\{y_j\} \\ &= \frac{1}{(m n)^2} n \sigma_y^2 \\ &= \frac{1}{(m^2 n)} \sigma_y^2 \\ \lim_{n \rightarrow \infty} &\rightarrow 0\end{aligned}$$

è consistente.

Esercizio 4.6:

Sia X una variabile aleatoria esponenziale di parametro incognito λ e sia (X_1, \dots, X_n) un campione casuale di X .

- Determinare uno stimatore del parametro λ .
- Verificare che sia uno stimatore distorto e proporre uno stimatore non distorto.

Soluzione: Il primo passo consiste nella creazione di una Likelihood function che conosca lo stimatore $\hat{\lambda}$ e il campione casuale.

$$L(\hat{\lambda}, X_1, \dots, X_n) = \prod_{j=1}^n \hat{\lambda} e^{-\hat{\lambda} X_j}$$

e passiamo a costruire l'osservabile:

$$\begin{aligned} \ell &= \ln L \\ &= n \ln \hat{\lambda} - \hat{\lambda} \sum_{j=1}^n X_j \end{aligned}$$

ora possiamo cercare il massimo:

$$\begin{aligned} \frac{d\ell}{d\hat{\lambda}} &= \frac{n}{\hat{\lambda}} - \sum_{j=1}^n X_j = 0 \\ \hat{\lambda} &= \frac{1}{\frac{1}{n} \sum_{j=1}^n X_j} \\ &= \frac{n}{\sum_{j=1}^n X_j} \\ &= \frac{1}{\bar{X}} \end{aligned}$$

è uno stimatore distorto ma consistente. Se facessimo le cose

b) Verifichiamo che sia uno stimatore distorto e guardiamo come ottenerne uno non distorto. Dobbiamo valutare il valore di aspettazione dello stimatore trovato nel punto a. Per farlo utilizziamo il fatto che una somma in n variabili aleatorie indipendenti si può rappresentare con una variabile Γ .

$$Y_n = \sum_{i=1}^n X_i = \Gamma(n, \lambda)$$

(lo abbiamo dimostrato in un esercizio con le funzioni generatrici del momento). Infatti la funzione generatrice del momento in questo caso è:

$$M_Y(t) = \prod_{i=1}^n M_{X_i}(t) = \left(\frac{1}{1 - \lambda t} \right)^n$$

Se andiamo quindi a fare il valore atteso di $\hat{\lambda}$ stiamo facendo:

$$\begin{aligned} E\{\hat{\lambda}\} &= E\left\{ \frac{n}{\sum_{j=1}^n X_j} \right\} \\ &= E\left\{ \frac{n}{Y_n} \right\} \\ &= \int \frac{1}{y} f_Y(y) dy \\ &= \int_0^{\infty} \frac{1}{s} \frac{\lambda^n s^{n-1} e^{-\lambda s}}{(n-1)!} ds \\ &= \frac{n \lambda^n}{(n-1)!} \int_0^{\infty} s^{n-2} e^{-\lambda s} ds \\ &= \frac{n \lambda^n}{(n-1)!} \frac{(n-2)!}{\lambda^n} \\ &= \lambda \frac{n}{n-1} \end{aligned}$$

Deduciamo che $\hat{\lambda}$ è uno stimatore distorto. Se la forma fosse:

$$\hat{\lambda}' = \frac{n-1}{\sum_{j=1}^n X_j}$$

sarebbe non distorto.

Esercizio 4.7:

Sia X una v.a. della quale si conosce un campione casuale di 3 determinazioni sperimentali X_1, X_2, X_3 e si considerino i seguenti stimatori di $\mu = E\{X\}$:

$$M_1 = \frac{X_1 + 3X_2 + X_3}{5}, \quad M_2 = \frac{2X_1 + X_2 + 2X_3}{5}, \quad M_3 = \frac{X_1 + 2X_2 + X_3}{4}$$

- Dire quali tra i tre stimatori sono corretti, dimostrando quanto asserito.
- Determinare lo stimatore più efficiente tra i tre e lo si confronti con la media campionaria di X_1, X_2, X_3 .

Soluzione: Iniziamo valutando la correttezza degli stimatori:

$$\begin{aligned} E\{M_1\} &= \frac{1}{5}E\{X_1\} + \frac{3}{5}E\{X_2\} + \frac{1}{5}E\{X_3\} = \mu \\ E\{M_2\} &= \frac{2}{5}E\{X_1\} + \frac{1}{5}E\{X_2\} + \frac{2}{5}E\{X_3\} = \mu \\ E\{M_3\} &= \frac{1}{4}E\{X_1\} + \frac{1}{2}E\{X_2\} + \frac{1}{4}E\{X_3\} = \mu \end{aligned}$$

Sono tutti e 3 corretti.

- Per determinare il più efficiente dobbiamo valutare le varianze degli estimatori:

$$\begin{aligned} Var\{M_1\} &= \frac{1}{25}\sigma^2 + \frac{9}{25}\sigma^2 + \frac{1}{25}\sigma^2 = 0.44\sigma^2 \\ Var\{M_2\} &= \frac{4}{25}\sigma^2 + \frac{1}{25}\sigma^2 + \frac{4}{25}\sigma^2 = 0.36\sigma^2 \\ Var\{M_3\} &= \frac{1}{16}\sigma^2 + \frac{4}{16}\sigma^2 + \frac{1}{16}\sigma^2 = 0.375\sigma^2 \end{aligned}$$

Il più efficiente risulta essere M_2 . Confrontiamolo con la media campionaria:

$$M_0 = \frac{X_1 + X_2 + X_3}{3}$$

facciamo la varianza

$$Var\{M_0\} = \frac{3}{9}\sigma^2 = 0.33\sigma^2$$

La media campionaria è lo stimatore più efficace.

Esercizio 4.8:

Le misurazioni del diametro di un campione casuale di 200 sferette da cuscinetto costruite da una macchina in una settimana mostrano una media di 0.824 cm ed una deviazione di 0.042 cm.

- Calcolare l'intervallo di confidenza al 95% del valore del diametro delle sferette.
- Come al punto precedente, con un livello di confidenza del 99%.

Soluzione: Chiamo X la v.a. "diametro delle sferette" e sappiamo le definizioni di media e varianza campionaria:

$$\bar{X} = \frac{1}{n} \sum_{j=1}^n X_j, \quad S^2 = \frac{1}{n-1} \sum_{j=1}^n (X_j - \bar{X})^2$$

Non conosco X ma so che le statistiche \bar{X} e S^2 sono distribuite gaussianamente come se X fosse anch'essa gaussiana. Sappiamo questo perchè $n = 200$ è un numero grande e il processo non ha

motivo per avere una distribuzione in particolare. Quando siamo in queste situazioni, ossia conosciamo le statistiche campionarie ma non quelle della distribuzione si usa la T-student:

$$T_n = \frac{\bar{X}_n - \mu}{S_n/\sqrt{n}}$$

questa T-student ha $\nu = 200 - 1 = 199$ gradi di libertà.

a) Calcoliamo il livello di confidenza al 95% considerando che la distribuzione che stiamo studiando è una distribuzione a 2 code:

$$\Pr \left\{ -t_{\frac{\alpha}{2}} < \frac{\bar{X} - \mu}{S/\sqrt{n}} < t_{\frac{\alpha}{2}} \right\} = 1 - \alpha$$

$$-t_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}} < \bar{X} - \mu < t_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}}$$

$$t_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}} > \mu - \bar{X} > -t_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}}$$

$$\Pr \left\{ \bar{X} - t_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}} < \mu < \bar{X} + t_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}} \right\} = 1 - \alpha$$

Il grado di fiducia se sto facendo il 95% è tabulato in:

$$t_{\frac{\alpha}{2}} = t_{0.05} = 1.645$$

$$t_{\frac{\alpha}{2}} = t_{0.025} = 1.96$$

$$t_{\frac{\alpha}{2}} = t_{0.005} = 2.576$$

Quindi:

$$\Pr \left\{ 0.824 - 1.96 \frac{0.042}{\sqrt{200}} < \mu < 0.824 + 1.96 \frac{0.042}{\sqrt{200}} \right\} = 0.95$$

$$\Pr \{ 0.818 < \mu < 0.830 \} = 0.95$$

b) al 99% si cambia il valore del grado di fiducia e si ottiene:

$$\Pr \{ 0.816 < \mu < 0.832 \} = 0.99$$

Esercizio 4.9:

Una equipe di medici vuole misurare il tempo di reazione medio di automobilisti con età compresa tra i 18 e i 50 anni e stima che la deviazione standard di questa sia $\sigma = 0.20$ s.

- Quanto deve essere grande il campione di misurazioni affinché si sia confidenti al 95% che l'errore nello stimare il tempo di reazione non sia superiore a 0.04 s?
- Ripetere il punto a considerando un livello di confidenza del 99%.
- Il metodo utilizzato nei punti precedenti per stimare la grandezza del campione da estrarre presuppone che la popolazione da cui questo viene estratto sia di tipo gaussiano. Non sapendo a priori di che tipo di variabile aleatoria si tratta, giustificare il metodo usato.

Soluzione: Impostiamo il problema considerando la v.a. X "tempo di reazione, a questo punto dobbiamo trovare il valore di n "grandezza del campione" tale da ottenere un valore di $\delta \leq 0.04$. Quindi quello che dobbiamo ottenere è:

$$\mu = \bar{X} \pm \delta, \quad 95\%$$

In questo caso dobbiamo trovare la dimensione del c.c. per ottenere un livello di confidenza del 95%. Quindi usiamo la definizione di intervallo di confidenza per una distribuzione gaussiana standard.

Questa quantità si trova usando la T-statistics:

$$t = \frac{\hat{\mu} - \mu}{\sigma/\sqrt{n}} \\ = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$$

a livello di formalismo una volta che usiamo la gaussiana standard possiamo scrivere che

$$T_n \rightarrow U$$

quindi

$$t_{\frac{\alpha}{2}} \simeq U_{\frac{\alpha}{2}}$$

a) Cerchiamo a questo punto il valore al 95%. Notiamo che abbiamo a che fare con una distribuzione a due code, quindi in questo caso:

$$\Pr \left\{ -U_{\frac{\alpha}{2}} < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < U_{\frac{\alpha}{2}} \right\} = 0.95 \\ \Pr \left\{ \bar{X} - U_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + U_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right\} = 0.95 \\ \text{possiamo a questo punto considerare che } \delta = U_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \\ n \geq \left(U_{\frac{\alpha}{2}} \frac{\sigma}{\delta} \right)^2 \\ \geq 96$$

b) Se cerchiamo il 99% basta considerare un valore diverso del grado di fiducia:

$$U_{\frac{\alpha}{2}} = 2.57$$

Se rifacciamo lo stesso conto del punto precedente allora otteniamo un valore di

$$n \geq 166$$

c) Il numero di sample del c.c. è grande quindi si può approssimare con la statistica usata. Inoltre se pensiamo alle variabili che si sommano per ottenere un tempo di reazione medio notiamo che sono una somma di vari tempi fisiologici indipendenti, il che rende plausibile l'uso della distribuzione gaussiana.

Esercizio 4.10:

Un lotto di sfere presenta un diametro medio μ incognito; la varianza del diametro delle sfere è nota e vale 0.01 cm. Si estrae un campione casuale di $n = 1000$ sfere sulle quali si osserva un diametro medio pari a 1.2 cm.

- Si determini l'intervallo di confidenza per μ avendo fissato un livello di confidenza del 99%.
- SI determini l'ampiezza di tale intervallo.

Soluzione: Per determinare l'intervallo di confidenza bisogna in primo luogo ricavare il quantile, questo può essere utile per generalizzare il problema, ad esempio nel caso il livello di confidenza richiesto non si trovi nelle tabelle. Il quantile di ordine $1 - \alpha/2$ della distribuzione normale standard:

$$1 - \alpha = 0.99 \\ \alpha = 1 - 0.99 \\ \text{dato che studiamo un caso a due code} \\ \frac{\alpha}{2} = 0.005$$

quindi:

$$U_{\frac{\alpha}{2}} = U_{0.005} = 2.57$$

a) A questo punto possiamo scrivere:

$$\Pr \left\{ -U_{\frac{\alpha}{2}} < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < U_{\frac{\alpha}{2}} \right\} = 0.99$$

$$\Pr \left\{ \bar{X} - U_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + U_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right\} = 0.99$$

$$\Pr \left\{ 1.2 - 2.576 \sqrt{\frac{0.01}{1000}} < \mu < 1.2 + 2.576 \sqrt{\frac{0.01}{1000}} \right\} = 0.99$$

In questo caso possiamo scrivere l'intervallo come:

$$[1.1918, 1.2081]$$

b) Per calcolare l'ampiezza dell'intervallo si può agire in due modi, se si conosce l'intervallo, come in questo caso, allora possiamo scrivere:

$$\begin{aligned} Amp &= Estremo superiore - Estremo inferiore \\ &= 1.2081 - 1.1918 \\ &= 0.01629 \text{ cm} \end{aligned}$$

oppure se viene chiesto direttamente di calcolare l'ampiezza possiamo usare la formula seguente:

$$\begin{aligned} Amp &= 2 \cdot U_{\frac{\alpha}{2}} \sqrt{\frac{\sigma^2}{n}} \\ &= 2 \cdot 2.576 \sqrt{\frac{0.01}{1000}} \\ &= 0.01629 \end{aligned}$$

Esercizio 4.11:

Da un insieme di bomboloni se ne estraggono 100 e si stima che il peso medio è pari a 82 g. Sapendo che $\sigma^2 = 25$,

a) si determini l'intervallo di confidenza per il peso medio μ dei bomboloni al livello di confidenza del 97%.

b) Si determini la probabilità che la differenza in valore assoluto fra la media campionaria e il peso medio μ dei bomboloni sia inferiore ai 3 g.

Soluzione: Iniziamo col determinare il quantile per la distribuzione a due code:

a) calcoliamo il fattore $1 - \alpha = 0.97$

$$1 - \alpha = 0.97$$

$$\alpha = 0.03$$

$$\frac{\alpha}{2} = 0.015$$

usando le tavole della distribuzione normale standard

$$U_{\frac{\alpha}{2}} = 2.17$$

Calcoliamoci l'intervallo di confidenza:

$$\Pr \left\{ -U_{\frac{\alpha}{2}} < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < U_{\frac{\alpha}{2}} \right\} = 0.97$$

$$\Pr \left\{ \bar{X} - U_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + U_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right\} = 0.97$$

$$\Pr \left\{ 82 - 2.17\sqrt{\frac{25}{100}} < \mu < 82 + 2.17\sqrt{\frac{25}{100}} \right\} = 0.97$$

Possiamo scrivere quindi l'intervallo:

$$[80.915, 83.085]$$

b) In questo caso vogliamo trovare il valore di probabilità che la differenza in valore assoluto fra la media campionaria e il peso medi sia minore di 3 g.

$$\begin{aligned} \Pr\{|\bar{X} - \mu| < 3\} &= \Pr\{-3 < (\bar{X} - \mu) < 3\} \\ &= \Pr \left\{ -\frac{3}{\sigma/\sqrt{n}} < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < \frac{3}{\sigma/\sqrt{n}} \right\} \\ &\text{ricordiamoci che } \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = Z \equiv N(0, 1) \\ &= \Pr \left\{ -\frac{3}{\sigma/\sqrt{n}} < Z < \frac{3}{\sigma/\sqrt{n}} \right\} \\ &= \Pr \left\{ -\frac{3}{\sqrt{25}/\sqrt{100}} < Z < \frac{3}{\sqrt{25}/\sqrt{100}} \right\} \\ &= \Pr\{-6 < Z < 6\} \\ &= \Pr\{Z > -6\} + \Pr\{Z < 6\} \\ &= \Phi(6) - [1 - \Phi(6)] \\ &= 2\Phi(6) - 1 \\ &= 1 \end{aligned}$$

Esercizio 4.12:

Durante il monitoraggio climatico della regione A sono state effettuate nel mese di luglio 30 misurazioni della temperatura media mensile in 30 diversi punti rappresentativi dell'intero territorio.

22.3	24.1	22.9	24.0	23.8	23.2
24.9	22.6	23.3	20.9	24.0	20.3
25.7	24.0	25.1	25.1	25.7	24.9
24.6	24.2	22.7	22.0	23.2	22.0
20.9	21.6	19.9	23.2	22.3	22.7

Dai dati in tabella di ricavano la media campionaria $\bar{x} = 23.2$ e la varianza campionaria $S^2 = 2.39$. Si vuole appurare se il campione di misure smentisce il dato storico che assegna il valore $\sigma^2 = 2$ alla varianza della temperatura. E' possibile affermare che il valore della varianza della temperatura è significativamente maggiore del valore storico?

Soluzione: Riassumiamo i dati che abbiamo in possesso:

$$\sigma_0^2 = 2, \quad \bar{X} = \frac{1}{30} \sum_{j=1}^3 0X_j = 23.20, \quad S^2 = \frac{1}{29} \sum_{j=1}^3 0(X_j - \bar{X})^2$$

La domanda che ci facciamo è la seguente: è vero che la varianza è $\sigma > 2$ oppure è vero che $\sigma = 2$ e che il valore della varianza campionaria è una fluttuazione statistica? La varianza campionaria $S^2 = \frac{1}{n-1} \sum_{j=1}^n (X_j - \bar{X})^2$ è distribuita come un χ^2 con 29 gradi di libertà:

$$(n-1) \frac{S^2}{\sigma^2}$$

Come la distribuzione T-student, la distribuzione χ^2 non è una unica distribuzione, ma una famiglia di distribuzioni che dipendono dal numero di gradi di libertà ν . La distribuzione χ^2 ha quindi media $\mu = \nu$ e varianza $\sigma^2 = 2\nu$.

Iniziamo il test declinando le ipotesi:

$$H_0 : \sigma^2 = 2$$

$H_0 : \sigma^2 > 2$ notiamo che è una ipotesi direzionale o a una coda.

Prendiamo la tabella della distribuzione χ^2 e cerchiamo il valore che individua la coda della distribuzione:

$$\begin{aligned} \Pr\{K_{29} > x_0\} &= 0.05 \\ \rightarrow x_0 &= 42.46 \end{aligned}$$

ora valutiamo il valore di χ^2 che otteniamo dalla nostra campionatura:

$$n-1 \frac{S^2}{\sigma^2} = 34.66$$

Questo già ci dà una indicazione di come il test risulterà, ma facciamo il test nella maniera opportuna.

$$\begin{aligned} \Pr\left\{n-1 \frac{S^2}{\sigma^2} > x_0\right\} &= \Pr\left\{S^2 > \frac{\sigma^2 x_0}{n-1}\right\} \\ &= \Pr\{S^2 > 2.94\} = 0.05 \end{aligned}$$

Il valore di varianza campionaria misurato è $S^2 = 2.34$ quindi possiamo dire che:

$$\Pr\{S^2 \leq 2.94\} = 0.95$$

e quindi il valore di varianza campionaria è all'interno del 95% di confidenza.

Definendo l'errore di prima e seconda specie come:

$$\begin{aligned} \alpha &= \Pr\{D_1|H_0\} \\ \beta &= \Pr\{D_0|H_1\} \end{aligned}$$

In questo caso possiamo dire che D_0 è accettata.

Esercizio 4.13:

Un campione di 8 pazienti è stato sottoposto ad una nuova terapia. Misurando un certo indice ematico sono stati registrati differenti valori (prima e dopo la cura) riportati in tabella.

Paziente	Prima	Dopo	Differenza
	x_1	x_2	$y = x_2 - x_1$
1	1.00	0.34	-0.66
2	0.69	0.62	-0.07
3	0.76	1.20	0.44
4	0.99	0.82	-0.17
5	0.77	1.03	0.26
6	1.24	1.48	0.24
7	1.06	1.24	0.18
8	1.19	1.11	-0.08

Dai valori si ricavano media campionaria delle differenze $\bar{Y} = 0.018$ e varianza campionaria delle differenze $S^2 = 0.118$.

E' possibile affermare che la terapia utilizzata ha prodotto una differenza significativa sul valore prima e dopo il trattamento?

Soluzione: Abbiamo chiamato $Y = X_2 - X_1$ e possiamo contare $n = 8$ pazienti. Formuliamo le ipotesi:

$$\begin{aligned} H_0 : \mu = 0 &\rightarrow D_0 : |\bar{Y}| < C \\ H_1 : \mu \neq 0 &\rightarrow D_1 : |\bar{Y}| \geq C \end{aligned}$$

E' un test a due code o test bilaterale.

Andiamo a valutare l'errore di primo tipo:

$$\begin{aligned} \alpha &= \Pr\{D_1|H_0\} = \Pr\{|\bar{Y}| \geq c \mid \mu = 0\} \\ &= \Pr\left\{\frac{|\bar{Y}|}{S/\sqrt{n}} \geq \frac{c}{S/\sqrt{n}} \mid \mu = 0\right\} = 0.05 \end{aligned}$$

Abbiamo ottenuto una T-student con $n - 1 = 7$ gradi di libert , dobbiamo stare attenti perch  i gradi di libert  sono pochi e quindi dobbiamo consultare la tabella della distribuzione T-student.

$$\Pr\{|T_7| > t_{\frac{\alpha}{2}}\} = 0.05$$

Leggendo la tabella otteniamo:

$$t_{\frac{\alpha}{2}} = 2.365$$

Quindi possiamo riscrivere il tutto come:

$$\begin{aligned} \Pr\left\{|\bar{Y}| \geq t_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}}\right\} &= \Pr\{|\bar{Y}| \geq 0.29\} \\ &= 0.05 \end{aligned}$$

Quindi possiamo affermare che D_0 sia accettata.

Esercizio 4.14:

All'interno di una azienda si vuole verificare se un incentivo salariale aumenta la produttivit . A questo scopo vengono messi a confronto due reparti: il reparto 1 con $n = 8$ dipendenti e il reparto 2 con $m = 6$ dipendenti. Solo ai dipendenti del reparto 2 viene dato un incentivo salariale. Le produttivit  dei due reparti dopo l'esperienza sono riportate in tabella.

x_{1i}	1.17	0.94	1.10	0.81	1.05	0.89	1.00	0.56
x_{2i}	1.44	1.54	0.90	1.14	0.91	1.13		

E' possibile affermare che la produttivit  del reparto 2   significativamente migliore di quella del reparto 1?

Soluzione: Sappiamo scrivere le medie campionarie e le varianze campionarie:

$$\begin{aligned} \bar{X}_1 &= 0.94, & S_1^2 &= 0.037 \\ \bar{X}_2 &= 1.18, & S_2^2 &= 0.05 \end{aligned}$$

Dato che abbiamo due campioni diversi ci serve una quantità che possa essere utilizzabile come confronto:

$$Y = X_2 - X_1$$

la cui media campionaria è:

$$\bar{Y} = 0.24$$

Fondamentale è anche calcolare la varianza campionaria della nuova variabile che abbiamo definito:

$$\begin{aligned} S^2 &= \frac{(n-1)S_1^2 + (m-1)S_2^2}{(n-1) + (m-1)} \\ &= \frac{7S_1^2 + 5S_2^2}{n+m-2} \\ &= \frac{7S_1^2 + 5S_2^2}{12} \\ &= 0.042 \end{aligned}$$

Dove ancora una volta il numero che compare al denominatore è il numero di gradi di libertà della distribuzione associata. Decidiamo quindi di fare un test unilaterale ad una coda le cui ipotesi sono:

$$\begin{aligned} H_0 : \mu_2 = \mu_1 &\rightarrow D_0 : \bar{Y} \leq C \\ H_1 : \mu_2 > \mu_1 &\rightarrow D_1 : \bar{Y} > C \end{aligned}$$

Valutiamo l'errore di primo tipo:

$$\alpha = \Pr\{D_1|H_0\} = \Pr\{\bar{Y} > c|\mu_2 = \mu_1\} = 0.05$$

Abbiamo ancora a che fare con una T-student con $m+n-2=12$ gradi di libertà. A questo punto possiamo scrivere quindi:

$$\Pr\{T_{12} > t_\alpha\} = 0.05$$

e ricavare dalle tabelle il valore di $t_\alpha = 1.782$. Ricordiamoci che questa volta il valore è da trovare per una distribuzione ad una coda.

$$\begin{aligned} \Pr\left\{\frac{\bar{Y}}{S/\sqrt{\frac{1}{n} + \frac{1}{m}}}\right\} > t_\alpha &= 0.05 \\ \Pr\{\bar{Y} > 0.197\} &= 0.05 \end{aligned}$$

Controlliamo che il valore di \bar{Y} cada o no all'interno dell'area del sottografico che conta il 95% della probabilità. In questo caso no dato che $0.24 > 0.197$ quindi è verificata D_1 .

Esercizio 4.15:

I cavalli vincitori delle ultime 144 edizioni del Palio di Siena avevano occupato le 8 possibili posizioni in griglia con le seguenti frequenze.

posizione	1	2	3	4	5	6	7	8	totale
frequenze	29	19	18	25	17	10	15	11	144

- Si formuli una opportuna ipotesi nulla e si esegua il test del χ^2 verificandone la significatività.
- Si esprima un parere in relazione alla possibile dipendenza dalla posizione del cavallo vincitore. Che certezza abbiamo di questa affermazione?

Soluzione: L'ipotesi nulla che dobbiamo formulare è:

$$H_0 = \text{nessuna dipendenza dalla posizione.} \rightarrow P_j = \frac{1}{8}$$

Le frequenze attese se la probabilità fosse uniforme sarebbero:

$$E_j = n P_j = 144 \frac{1}{8} = 18$$

A questo punto possiamo effettuare un test del χ^2 :

$$\sum_{j=1}^8 \frac{(E_j - O_j)^2}{E_j} = 16.3$$

Ora che abbiamo il valore del χ^2 dobbiamo capire quale significatività ha questo valore. Consideriamo i gradi di libertà della distribuzione che sono $\nu = n - 1 = 7$.

$$\Pr\{K_7 \geq 16.0\} = 0.025$$

$$\Pr\{K_7 \geq 18.5\} = 0.01$$

Abbiamo questi due valori dalle tabelle e dobbiamo capire a che livello sta il valore che abbiamo trovato noi. Sostanzialmente abbiamo 2 punti nel piano e dobbiamo ricavarci la retta passante per i 2 punti:

$$\begin{aligned} \frac{x - x_1}{x_2 - x_1} &= \frac{y - y_1}{y_2 - Y - 1} \\ \frac{x - 16}{18.5 - 16} &= \frac{y - 0.01}{0.025 - 0.01} \\ a &= -0.006 \\ b &= 0.121 \\ y &= -0.006 \cdot x + 0.121 \end{aligned}$$

Quindi il valore per la x che abbiamo ricavato dal nostro χ^2 è:

$$y = 0.0232$$

scegliamo D_1 con un livello di confidenza quindi è $1 - \alpha = 97.7\%$.

Esercizio 4.16:

Considerare il seguente insieme di misure sperimentali di una grandezza fisica.

731	772	771	681	722	688	653	757	733	742
739	780	709	676	760	748	672	687	766	645
678	748	689	810	805	778	764	753	709	675
698	770	754	830	725	710	738	638	787	712

a) Raggruppare l'insieme dei dati in 4 classi: $D < \bar{X} - S$, $\bar{X} - S < D < \bar{X}$, $\bar{X} < D < \bar{X} + S$, $D > \bar{X} + S$.

b) Si formuli una ipotesi ragionevole in relazione alla variabile aleatoria che governa la distribuzione dei dati e la si discuta.

Soluzione: Calcoliamoci la media campionaria, la deviazione campionaria e le rispettive differenze.

$$\begin{aligned}\bar{X} &= 730.1, & S &= 46.8, & n &= 40 \\ \bar{X} - S &= 683.3, & \bar{X} + S &= 776.9\end{aligned}$$

Sono quindi 8 nel primo intervallo, 10 nel secondo, 16 nel terzo e 6 nel quarto. Ricordiamoci che nell'intervallo $[\bar{X} - S, \bar{X} + S]$ cadono il 68% dei dati.

b) Possiamo immaginare che la variabile da usare sia Gaussiana e quindi poniamo come ipotesi che:

$$H_0 = (X_1, \dots, X_{40}) \text{ è un campione casuale di } X.$$

Per effettuare il test del χ^2 dobbiamo avere sia i dati osservati sia i dati stimati, quindi procediamo ad una stima delle frequenze considerando la distribuzione che abbiamo scelto.

intervallo	P_j	E_j	O_j
1	0.16	6.4	8
2	0.34	13.6	10
3	0.34	13.6	16
4	0.16	6.4	6

Quindi il valore del test del χ^2 è:

$$\begin{aligned}\chi_0 &= \sum_{j=1}^4 \frac{(O_j - E_j)^2}{E_j} \\ &= 1.80\end{aligned}$$

Per il discorso dei gradi di libertà devo stare molto attento, infatti in questo caso ho 4 classi e $\nu = 1$ gradi di libertà. Il numero di gradi di libertà ν delle distribuzione chi-quadro risulta essere:

$\nu = n - 1$ se la distribuzione delle frequenze teoriche può essere determinata senza usare dati campionari.

$\nu = n - m - 1$ se la distribuzione delle frequenze teoriche può essere determinata stimando m parametri della popolazione tramite i valori campionari.

$$\Pr\{K_1 \geq 1.32\} = 0.25$$

$$\Pr\{K_1 \geq 2.71\} = 0.1$$

facciamo una interpolazione lineare:

$$\begin{aligned}\frac{x - x_1}{x_2 - x_1} &= \frac{y - y_1}{y_2 - Y - 1} \\ \frac{x - 1.32}{2.71 - 1.32} &= \frac{y - 0.1}{0.25 - 0.1} \\ y &= -0.108 \cdot x + 0.392 \\ &= 0.198\end{aligned}$$

Se accettassi D_0 il livello di rischio sarebbe del 20%.

Esercizio 4.17:

Si lanciano insieme 3 dadi 400 volte e si registra il numero di "6" in ciascun tiro con i seguenti conteggi.

numero di "6"	frequenze
0	217
1	148
2	33
3	2

a) Formulare una adeguata ipotesi statistica nulla ed una alternativa finalizzate all'esecuzione del metodo del χ^2 . Scegliere un opportuno raggruppamento per i dati riportati in tabella e calcolare il valore del χ^2 campionario.

b) Calcolare, sulla base della suddivisione scelta nel punto precedente, il livello di significatività del test discutendone il risultato.

Soluzione: Iniziamo pensando alla distribuzione che questi lanci di dati seguono, possiamo pensare che sia la binomiale in cui:

$$0 : \binom{3}{0} \left(\frac{1}{6}\right)^0 \left(\frac{5}{6}\right)^3 = 0.5787$$

$$1 : \binom{3}{1} \left(\frac{1}{6}\right)^1 \left(\frac{5}{6}\right)^2 = 0.3472$$

$$2 : \binom{3}{2} \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^1 = 0.0694$$

$$3 : \binom{3}{3} \left(\frac{1}{6}\right)^3 \left(\frac{5}{6}\right)^0 = 0.0046$$

A questo punto dobbiamo formulare una ipotesi

$$H_0 = \text{i dadi non sono truccati}$$

e scriviamoci l'opportuna tabella in cui effettuiamo un raggruppamento dal momento che un ingresso ha conteggi inferiori a 5:

intervallo	E_j	O_j
0	231.48	217
1	138.88	148
2	29.63	35

e calcoliamoci il χ^2 :

$$\chi_0 = \sum_{j=0}^2 \frac{(O_j - E_j)^2}{E_j} = 2.478$$

Calcoliamoci la probabilità della distribuzione associata che ha $\nu = n - 1 = 2$ gradi di libertà:

$$\Pr\{K_2 \geq 3.22\} = 0.2$$

$$\Pr\{K_2 \geq 2.41\} = 0.3$$

Possiamo dire che c'è una probabilità compresa tra il 20 e il 30% che i dadi siano truccati.

Esercizio 4.18:

Dal modello teorico è noto che le letture di una certa v.a. X dovrebbero seguire una v.a. gaussiana con deviazione standard $\sigma = 2$ e media $\mu = 10$. Eseguendo 20 determinazioni sperimentali si ottengono i risultati contenuti in tabella.

3.84	7.01	7.96	8.09	8.11
8.89	8.95	9.00	9.18	9.95
10.07	10.40	10.91	11.01	11.24
11.37	12.33	12.69	12.71	13.97

- Si scelga un opportuno numero di raggruppamenti per il calcolo del χ^2 .
- Si indichino chiaramente in una tabella gli intervalli scelti, le popolazioni effettive, sulla base dei dati e quelle attese sulla base della distribuzione data.
- Si calcoli infine sulla base dei raggruppamenti precedentemente ottenuti, il valore del χ^2 per i dati forniti. Si traggano le conseguenze sulla validità delle assunzioni statistiche iniziali.

Soluzione: Sappiamo che la variabile aleatoria X dovrebbe seguire una gaussiana con $\mu = 10$ e $\sigma = 2$, tuttavia la media e la deviazione standard campionarie dicono che $\bar{X} = 9.88$ e $S = 2.33$. Usando il valore di media e varianza della distribuzione gaussiana otteniamo 4 intervalli $D < \bar{X} - S$, $\bar{X} - S < D < \bar{X}$, $\bar{X} < D < \bar{X} + S$, $D > \bar{X} + S$. Riassumiamo in una tabella:

intervallo	E_j	O_j
$-\infty - 8$	3.2	3
$8 - 10$	6.4	7
$10 - 12$	6.4	6
$12 - \infty$	3.2	4

Se facciamo il χ^2 con questa tabella otteniamo un valore di:

$$\chi_0^2 = \sum_{j=1}^4 \frac{(O_j - E_j)^2}{E_j} = 0.294$$

Questo non è esatto però, dal momento che la distribuzione è gaussiana dobbiamo riportarci ad una gaussiana standard, la cui formula è:

$$\frac{x - \mu}{\sigma} = \frac{x - 10}{2}$$

E raggruppiamo in modo da avere i valori osservati equipartiti in 4 classi. Calcoliamo per ogni ripartizione i valori e otteniamo:

intervallo	E_j	O_j
$-\infty - 8.5$?	5
$8.5 - 10$?	5
$10 - 11.3$?	5
$11.3 - \infty$?	5

Per completarlo bisogna valutare le probabilità delle aree nella gaussiana normalizzata e per farlo basta controllare la tabella della gaussiana normalizzata che riporta le seguenti probabilità:

$$0.2265, \quad 0.2735, \quad 0.242, \quad 0.258$$

e quindi:

intervallo	E_j	O_j
$-\infty - 8.5$	4.53	5
$8.5 - 10$	5.47	5
$10 - 11.3$	4.84	5
$11.3 - \infty$	5.16	5

Facendo il test del χ^2 con 3 gradi di libertà troviamo che:

$$\chi_0^2 = \sum_{j=1}^4 \frac{(O_j - E_j)^2}{E_j} = 0.099$$

Controllando nella tabella del χ^2 vediamo che cade nell'intervallo:

$$\Pr\{K_3 \geq 0.072\} = 0.995$$

$$\Pr\{K_3 \geq 0.115\} = 0.99$$

Esercizio 4.19:

Sia X una v.a. discreta ricavata a partire da una v.a. di Poisson, ma con la possibilità che X possa assumere anche valori negativi, la cui distribuzione di probabilità vale:

$$P_j = c \frac{\mu^{|j|}}{|j|!} e^{-\mu}$$

- a) Determinare c in modo che sia verificata la condizione di normalizzazione e disegnarne qualitativamente l'andamento della distribuzione di probabilità.
 b) Calcolare la media, varianza, e la funzione generatrice dei momenti X . c) Sia data la successione $\{X\}_n$ ottenuta ponendo $\mu = \frac{1}{n}$; calcolare il limite di tale successione indicando se si tratta di convergenza in distribuzione e/o in probabilità.

Soluzione: Troviamo c e disegniamo P_j .

$$\begin{aligned} \sum_{j=-\infty}^{\infty} c \frac{\mu^{|j|}}{|j|!} e^{-\mu} &= 1 \\ &= c e^{-\mu} \left(2 \sum_{j=0}^{\infty} \frac{\mu^j}{j!} - 1 \right) \\ &= c e^{-\mu} (2e^{\mu} - 1) \\ c &= \frac{e^{\mu}}{2e^{\mu} - 1} \end{aligned}$$

Quindi possiamo scrivere:

$$P_j = \frac{1}{2e^{\mu} - 1} \frac{\mu^{|j|}}{|j|!}$$

- b) Troviamo la media e la varianza e funzione generatrice:

$$\begin{aligned} E\{X\} &= \sum_{j=-\infty}^{\infty} j \frac{1}{2e^{\mu} - 1} \frac{\mu^{|j|}}{|j|!} \\ &= 0 \end{aligned}$$

dal momento che è una funzione dispari.

$$E\{X^2\} = \sum_{j=-\infty}^{\infty} j^2 \frac{1}{2e^\mu - 1} \frac{\mu^{|j|}}{|j|!}$$

considerando che:

$$\begin{aligned} \sum_{j=1}^{\infty} j^2 \frac{\mu^j}{j!} &= \sum_{j=1}^{\infty} j \frac{\mu^j}{(j-1)!} \\ &= \sum_{k=0}^{\infty} (k+1) \frac{\mu^{k+1}}{(k)!} \\ &= \sum_{k=0}^{\infty} \mu \frac{\mu^k}{(k-1)!} + \mu e^\mu \\ &= \sum_{l=0}^{\infty} \mu \frac{\mu^{l+1}}{(l)!} + \mu e^\mu \\ &= \mu^2 e^\mu + \mu e^\mu \end{aligned}$$

Otteniamo che

$$E\{X^2\} = \frac{2}{2e^\mu - 1} e^\mu (\mu + 1)$$

Infine la funzione generatrice dei momenti è:

$$\begin{aligned} M_X(t) &= E\{e^{tX}\} \\ &= \sum_{j=-\infty}^{\infty} \frac{1}{2e^\mu - 1} e^{tj} \frac{\mu^{|j|}}{|j|!} \\ &= \sum_{j=-\infty}^0 \frac{1}{e^\mu - 1} e^{tj} \frac{\mu^{-j}}{-j!} + \sum_{j=0}^{\infty} \frac{1}{e^\mu - 1} e^{tj} \frac{\mu^j}{j!} - 1 \\ &= \sum_{j=0}^{\infty} \frac{1}{e^\mu - 1} e^{-tj} \frac{\mu^j}{j!} + \sum_{j=0}^{\infty} \frac{1}{e^\mu - 1} e^{tj} \frac{\mu^j}{j!} - 1 \\ &= \frac{1}{2e^\mu - 1} e^{\mu e^{-t}} + e^{\mu e^t} - 1 \end{aligned}$$

c) Per valutare la convergenza in distribuzione o probabilità della nuova successione ripassiamo le definizioni. Una successione di v.a. con funzione di ripartizione $F_X(x)$ converge in distribuzione se esiste:

$$\lim_{n \rightarrow \infty} F_X(x) = F_X(x)$$

si dice che converge in probabilità se:

$$\lim_{n \rightarrow \infty} \Pr\{|X_n - X| < \epsilon\} = 1$$

Abbiamo la funzione generatrice dei momenti possiamo usare la funzione caratteristica che si ricava in modo analogo:

$$\begin{aligned}
 H_X(t) &= E\{e^{itX}\} \\
 &= \frac{1}{2e^\mu - 1} e^{\mu e^{-it}} + e^{\mu e^{it}} - 1 \\
 &\text{dove sappiamo che } \mu = \frac{1}{n} \\
 &= \frac{1}{2e^{\frac{1}{n}} - 1} e^{\frac{e^{-it}}{n}} + e^{\frac{e^{-it}}{n}} - 1 \\
 &\lim_{n \rightarrow \infty} \frac{1}{2e^{\frac{1}{n}} - 1} e^{\frac{e^{-it}}{n}} + e^{\frac{e^{-it}}{n}} - 1 = e^{it0} = 1
 \end{aligned}$$

Il limite per $n \rightarrow \infty$ ci dice che la successione $X_n \rightarrow X$ in distribuzione, dato che la v.a. X è degenera allora la convergenza è anche in probabilità. Degenera significa che la distribuzione di probabilità è concentrata in un solo valore, nel nostro caso abbiamo che la funzione caratteristica ottenuta ci sta descrivendo la funzione caratteristica di una delta di Dirac che ha come risultato $e^{it0} = 1$.

$$\int_{-\infty}^{\infty} \delta(X=0) e^{itX} dX = e^{it0} = 1$$

Esercizio 4.20:

Sia X_n con $n \in \mathbb{N}$ un insieme di variabili aleatorie indipendenti tali che X_n siano uniformi sull'intervallo $(0, 3)$. Sia $Z_n = \min(X_1, \dots, X_n)$. Studiare la convergenza in distribuzione di Z_n .

Soluzione: Per prima cosa dobbiamo calcolarci F_{Z_n} quindi:

$$F_{X_n} = \begin{cases} 0 & a \leq 0 \\ \frac{a}{3} & 0 \leq a \leq 3 \\ 1 & a \geq 3 \end{cases}$$

Calcoliamoci quindi :

$$\begin{aligned}
 F_{Z_n}(t) &= \Pr\{Z_n \leq t\} \\
 &= \Pr\{\min(X_1, \dots, X_n) \leq t\} \\
 &= 1 - \Pr\{\min(X_1, \dots, X_n) > t\} \\
 &= 1 - (1 - F_{X_1}(t)) \cdot (1 - F_{X_2}(t)) \cdot \dots \cdot (1 - F_{X_n}(t)) \\
 &= 1 - (1 - F_{X_i}(t))^n
 \end{aligned}$$

Quindi per $t \in (0, 3)$

$$F_{Z_n} = \begin{cases} 0 & t \leq 0 \\ 1 - (1 - \frac{t}{3})^n & 0 \leq t \leq 3 \\ 1 & t \geq 3 \end{cases}$$

Ora per studiare la convergenza di distribuzione basta fare il limite per ogni intervallo:

$$\begin{aligned}
 t < 0 & \quad \lim_{n \rightarrow \infty} 0 = 0 \\
 t \in (0, 3) & \quad \lim_{n \rightarrow \infty} 1 - (1 - \frac{t}{3})^n = 1 \\
 t > 3 & \quad \lim_{n \rightarrow \infty} 1 = 1
 \end{aligned}$$

sostanzialmente converge a 0 se $t < 0$ e a 1 se $t \geq 0$. Si dice quindi che in distribuzione converge ad una costante nello specifico a 0. Dato che converge ad una costante in distribuzione allora converge anche in probabilità.

Esercizio 4.21:

Sia X_n con $n \in \mathbb{N}$ una successione di variabili aleatorie indipendenti. Supponiamo che ciascuna X_n abbia distribuzione bernulliana di parametro $\frac{1}{n^\alpha}$ con $\alpha > 0$. Per quali valori di $\alpha > 0$ la successione converge in distribuzione?

Soluzione: Scriviamo in primo luogo la funzione di ripartizione:

$$F_{X_n}(x) = \begin{cases} 0 & x \leq 0 \\ 1 - \frac{1}{n^\alpha} & 0 \leq x < 1 \\ 1 & x \geq 1 \end{cases}$$

facciamo il limite di questa funzione di ripartizione:

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = \begin{cases} 0 & x \leq 0 \\ 1 & x \geq 1 \end{cases}$$

La successione converge a 0 in distribuzione per ogni $\alpha > 0$.

$$X_n \xrightarrow[\lim_{n \rightarrow \infty}]{d} 0 \quad \forall \alpha > 0$$