Esame di *Calcolo delle probabilità e statistica* (per studenti di Informatica)
corso A

Università degli studi di Bari Aldo Moro

Docente: Stefano Rossi

01-02-2021

Esercizio 1. Una moneta equa viene lanciata N volte, dove N è una variabile aleatoria geometrica di parametro p, cioè $P[N = k] = (1-p)^{k-1}p$, $k \geq 1$, con $0 < p < 1$. Indichiamo con X il numero di teste ottenute.

(1) Calcolare $P[X = 0]$, verificando che $P[X = 0] = \frac{p}{1+p}$.
(2) Calcolare $P[N = 2|X = 0]$, la probabilità che siano stati fatti 2 lanci sapendo che non si sono avute teste.
(3) Determinare p affinché la probabilità determinata sopra sia pari a $\frac{1}{16}$.

(Ricordare che $\sum_{k=0}^{\infty} q^k = \frac{1}{1-q}$ per ogni $q \in \mathbb{R}$ tale che $|q| < 1$.)

Esercizio 2. Dato il parametro $\theta > 0$, si considera la funzione $f(x;\theta) := cx$ per $0 \leq x \leq \theta$ e $f(x;\theta) = 0$ altrove. Determinare c affinché f sia la densità di probabilità di una variabile aleatoria X, verificando che $c = \frac{2}{\theta^2}$.

(1) Determinare il valore atteso della variabile aleatoria X
(2) Usare il valore determinato sopra per ricavare uno stimatore corretto del parametro θ in corrispondenza a un campione $(X_1, X_2, \ldots, X_n)$ di rango n.
(3) Determinare lo stimatore di massima verosimiglianza di θ in corrispondenza a un campione $(X_1, X_2, \ldots, X_n)$ di rango n.

Esercizio 3. Un'azienda vinicola imbottiglia vini in due stabilimenti diversi. Si vuole controllare se il contenuto medio di una bottiglia di vino sia lo stesso nei due stabilimenti. A tal scopo si prendono a caso $n = 30$ bottiglie per ciascuno dei due stabilimenti e si trovano le medie campionarie $\overline{x_1} = 748\text{cl}$ e $\overline{x_2} = 751\text{cl}$, rispettivamente. Assumendo che i contenuti medi siano variabili aleatorie gaussiane con la stessa varianza $\sigma^2 = 40\text{cl}^2$:

(1) Dire quale test occorre eseguire se si vuole verificare che il contenuto medio delle bottiglie sia lo stesso per i due stabilimenti, specificando se si tratta di un test unilaterale o bilaterale.
(2) Condurre il test a un livello di significatività del 10% e del 5%.
(3) Calcolare il p-value del test.

Esercizio 1

1. $P[X=0] = \sum_{k=1}^{\infty} P[X=0 \mid N=k] \cdot P[N=k]$ (formula della probabilità totale)

Inserendo i dati, troviamo

$$\sum_{k=1}^{\infty} \frac{1}{2^k} \cdot (1-p)^{k-1} \cdot p = \frac{p}{2} \sum_{k=1}^{\infty} \left(\frac{1-p}{2}\right)^{k-1} = \frac{p}{2} \sum_{\ell=0}^{\infty} \left(\frac{1-p}{2}\right)^{\ell}$$

$$= \frac{p}{2} \cdot \frac{1}{1-\frac{1-p}{2}} = \frac{p}{\cancel{2}} \frac{\cancel{2}}{2-1+p} = \frac{p}{1+p}$$

avendo usato che $P[X=0 \mid N=k] = \frac{1}{2^k}$ (la probabilità di avere k croci in k lanci)

2. $P[N=2 \mid X=0] = \frac{P[N=2 \text{ e } X=0]}{P[X=0]}$

Ora

$P[X=0 \mid N=2] = \frac{P[X=0 \text{ e } N=2]}{P[N=2]}$, da cui troviamo

$P[X=0 \text{ e } N=2] = P[X=0 \mid N=2] \cdot P[N=2]$, per cui

$$P[N=2 \mid X=0] = \frac{P[X=0 \mid N=2] \cdot P[N=2]}{P[X=0]} =$$

$$\frac{1}{2^2} \cdot \cancel{p}(1-p) \cdot \frac{1+p}{\cancel{p}} = \frac{1}{4}(1-p^2)$$

3. $\frac{1}{4}(1-p^2) = \frac{1}{16}$, $\quad 1-p^2 = \frac{1}{4}$, $\quad p^2 = \frac{3}{4}$, $\quad p = \frac{\sqrt{3}}{2} \simeq 0.86$

Esercizio 2

Bisogna imporre la condizione $\int_0^\theta cx\,dx \overset{!}{=} 1$.

Ora $\int_0^\theta x\,dx = \left.\frac{x^2}{2}\right|_0^\theta = \frac{\theta^2}{2}$; perciò $c \cdot \frac{\theta^2}{2} = 1 \Rightarrow c = \frac{2}{\theta^2}$

1. $E[X] = \int_0^\theta \frac{2}{\theta^2} x \cdot x\,dx = \frac{2}{\theta^2}\int_0^\theta x^2 dx = \frac{2}{\theta^2} \left.\frac{x^3}{3}\right|_0^\theta = \frac{2}{\theta^2}\frac{\theta^3}{3} =$

$= \frac{2}{3}\theta$

2. Dato che $\theta = \frac{3}{2}E[X] = E\left[\frac{3}{2}X\right]$, si deduce che

$\frac{3}{2}\left(\frac{X_1 + X_2 + \dots + X_n}{n}\right)$ è uno stimatore corretto di θ, cioè

$$E\left[\frac{3}{2}\left(\frac{X_1+X_2+\dots+X_n}{n}\right)\right] = \theta$$

3. $h(x_1, \dots, x_n; \theta) = f(x_1; \theta) \cdots f(x_n, \theta)$

$$= \begin{cases} \frac{2^n}{\theta^{2n}} x_1 x_2 \dots x_n & \text{se } 0 \le x_i \le \theta \text{ per ogni } i \\ 0 & \text{altrimenti} \end{cases}$$

$\log h = n \log 2 - 2n \log\theta + \log(x_1 \cdots x_n)$

$\frac{\partial}{\partial\theta} \log h = -\frac{2n}{\theta} < 0$

il valore massimo è dunque assunto in corrispondenza del più piccolo valore possibile di θ, dato che $\theta \ge x_i \quad \forall i = 1, \dots, n$

troviamo $\theta = \max\{x_1, x_2, \dots, x_n\}$

Il corrispondente stimatore è allora $\hat{\Theta} = \max\{X_1, X_2, \dots, X_n\}$.

Esercizio 3

$\overline{X_1} = 748 \, cl$

$\overline{X_2} = 751 \, cl$

$\sigma^2 = 40 \, cl^2$ (note)

$n = 30$

1. test gaussiano bilaterale sfruttando la statistica

$$\frac{\overline{X_1} - \overline{X_2}}{\sigma \sqrt{\frac{1}{n} + \frac{1}{n}}}$$

Si rifiuta H_0 : "$\mu_1 = \mu_2$" al livello α se

$$\left| \frac{\overline{X_1} - \overline{X_2}}{\sigma \sqrt{\frac{1}{n} + \frac{1}{n}}} \right| > z_{1-\frac{\alpha}{2}}$$

2. $$\frac{|\overline{X_1} - \overline{X_2}|}{\sigma \sqrt{\frac{1}{n} + \frac{1}{n}}} = \frac{3}{\sqrt{40} \sqrt{\frac{1}{30} + \frac{1}{30}}} = 1.837$$

- con $\alpha = 0.1$ $\quad z_{1-\frac{\alpha}{2}} = z_{0.95} = 1.65$

$\Rightarrow$ si rifiuta H_0

- con $\alpha = 0.05$ $\quad z_{1-\frac{\alpha}{2}} = z_{0.975} = 1.96$

$\Rightarrow$ si accetta H_0

3. il p-value è determinato dalla relazione

$$z_{1-\frac{p}{2}} = 1.837$$

da cui $1 - \frac{p}{2} = \mathrm{Erf}(1.837)$, $p = 2(1 - \mathrm{Erf}(1.837)) = 2(1 - 0.96711)$

$\simeq 6.6\%$

Esame di *Calcolo delle Probabilità e Statistica* (per studenti di Informatica) corso A

Università degli studi di Bari Aldo Moro

12-04-2021

Esercizio 1. Una scatola contiene 100 monete con due facce, testa e croce. Tra queste, 10 monete sono truccate, e fanno uscire testa con probabilità $\frac{2}{3}$. Le restanti 90 monete sono equilibrate. Si prende una moneta a caso e la si lancia 1000 volte.

(1) Sia N la variabile aleatoria che dà il numero di teste ottenute nell'esperimento di sopra. Quanto vale l'attesa $E[N]$?
(2) Calcolare la probabilità di ottenere almeno 500 teste.
(3) Sapendo di aver ottenuto più di 600 teste, qual è la probabilità che la moneta sia truccata?

Esercizio 2. Sia X una variabile aleatoria uniformemente distribuita tra 0 e a, dove $a > 0$ è un parametro. Scrivere la densità di probabilità di X e calcolarne la media.
Sia $(X_1, X_2, \ldots, X_n)$ un campione di rango n distribuito come X.

(1) Sfruttando il calcolo della media di X, esibire uno stimatore corretto del parametro a.
(2) Determinare lo stimatore di massima verosimiglianza per il parametro a corrispondente al campione $(X_1, X_2, \ldots, X_n)$ mostrando che è dato da $T_n = \max\{X_1, X_2, \ldots, X_n\}$.
(3) Verificare che la funzione di distribuzione di T_n è data da $F_n(t) = (\frac{t}{a})^n$ per $0 \leq t \leq a$, $F_n(t) = 0$ per $t < 0$, e $F_n(t) = 1$ per $t \geq a$.
(4) Calcolare $E[T_n]$ e verificare che T_n è asintoticamente corretto.

Esercizio 3. Una compagnia petrolifera dichiara che il contenuto di zolfo del suo carburante diesel non supera lo 0.15%. Si vuole verificare questa ipotesi analzzando 40 campioni. Si trova un contenuto medio di 0.162% con deviazione standard *campionaria* di 0.40%.

(1) Dire quale test statistico occorre usare, specificando se si tratta di un test unilaterale o bilaterale.
(2) Verificare l'ipotesi al 10% di significatività.
(3) Verificare l'ipotesi al 5% di significatività.

H_1 : la moneta è equilibrata ; $p(H_1) = \frac{9}{10}$

H_2 : la moneta è truccata ; $p(H_2) = \frac{1}{10}$

T = "La moneta dà testa" $p(T|H_1) = \frac{1}{2} = p_1$

$p(T|H_2) = \frac{2}{3} = p_2$

$N = 1000$ (lanci della moneta)

1. $E[N] = \sum_{i=0}^{1000} i\, P[N=i] = \sum_{i=0}^{1000} i\, P[N=i\,|H_1] \cdot p(H_1) +$

$\sum_{i=0}^{1000} i\, P[N=i\,|H_2] \cdot p(H_2) =$

$= p(H_1)\, N \cdot p_1 + p(H_2)\, N\, p_2 =$

$\frac{9}{10} \cdot 1000 \cdot \frac{1}{2} + \frac{1}{10} \cdot 1000 \cdot \frac{2}{3}$

$\frac{9}{10} \cdot 500 + \frac{1}{10}\, 666.6 \simeq 516.6$

2. $P[N \geq 500] = P[N \geq 500\,|H_1] \cdot p(H_1) + p[N \geq 500\,|H_2] \cdot p(H_2)$

$= \frac{9}{10}\, P[N \geq 500\,|H_1] + \frac{1}{10}\, P[N \geq 500\,|H_2]$

$P[N \geq 500\,|H_1] = \frac{1}{2}$

$P[N \geq 500\,|H_2] =$

Nell'ipotesi H_2 (moneta truccata) N ha media 666.6

e varianza $1000 \cdot p_2\, q_2 = 1000\, \frac{2}{3} \cdot \frac{1}{3} = 222.22$

La deviazione standard di N è pertanto $\sigma = \sqrt{222.22} = 14.9$

Sfruttando l'approssimazione del limite centrale

$P[N \geq 500 \mid H_2]$

$= P\left[\frac{N - 666.6}{14.907} \geq \frac{500 - 666.6}{14.907}\right] = P[\underset{\substack{\uparrow \\ \text{normale standard}}}{Z} \geq -11.18] \simeq 1$

Quindi $P[N \geq 500] = \frac{9}{10} \cdot \frac{1}{2} + \frac{1}{10} \cdot 1 =$

$= \frac{9}{20} + \frac{1}{10} = \frac{9+2}{20} = \frac{11}{20} = 0.55$

3. $P[H_2 \mid N \geq 600] = \frac{P[N \geq 600 \mid H_2] \cdot p(H_2)}{p(N \geq 600)} = \frac{p[N \geq 600 \mid H_2] \cdot p(H_2)}{\sum_{i=1}^{2} p[N \geq 600 \mid H_i] p(H_i) +}$

$p[N \geq 600 \mid H_2] = P\left(Z \geq \frac{600 - 666.6}{14.907}\right) = P(Z \geq -4.46) = 1$

$p[N \geq 600 \mid H_1] = ?$

Sotto H_1 N ha media 500 e varianza $1000 \cdot \frac{1}{2} \cdot \frac{1}{2} = 250$.
La deviazione standard è allora $\sqrt{250} = 15.81$

$P[N \geq 600 \mid H_1] = P\left(\underset{\substack{\uparrow \\ \text{normale standard}}}{Z} \geq \frac{600 - 500}{15.81}\right) = P(Z \geq 6.32) = 0$

ma allora $P[H_2 \mid N \geq 600] = \frac{1 \cdot \frac{1}{10}}{0 \cdot \frac{9}{10} + \frac{1}{10} \cdot 1} \simeq 1$

Esame di *Calcolo delle probabilità e statistica* (per studenti di Informatica)
corso A

Università degli studi di Bari Aldo Moro

08-06-2021

Esercizio 1. Si lancia N volte un dado equo a sei facce, numerate da 1 a 6, dove N è una variabile aleatoria di Poisson di parametro $\lambda = \frac{6}{5}$.

- Determinare la probabilità di ottenere 6 almeno una volta.
- Determinare la probabilità di ottenere 6 esattamente una volta.
- Sapendo di non avere mai ottenuto 6, calcolare la probabilità che N sia pari a n, per ogni $n \in \mathbb{N}$, e calcolarne il limite per $n \to \infty$.

(Si ricorda che $\sum_{n=0}^{\infty} \frac{1}{n!} = e$.)

Esercizio 2. Verificare che per ogni $a > 0$ la funzione $f(x) = \frac{3}{a^3} x^2 \chi_{[0,a]}(x)$ è la densità di probabilità di una variabile aleatoria X.

- Determinare l'attesa di X.
- Sfruttare il calcolo precedente per esibire uno stimatore corretto di a con un campione di rango n distribuito come X.
- Dire se lo stimatore ottenuto al punto precedente è consistente.
- Determinare lo stimatore di massima verosimiglianza di a relativo a un campione di rango n.

Esercizio 3.

Una casa vinicola ha due impianti di imbottigliamento. Si vuole controllare che il contenuto medio di vino per le bottiglie dei due impianti sia lo stesso. A tal scopo si analizzano $n_1 = 25$ bottiglie provenienti dal primo impianto e $n_2 = 25$ provenienti dal secondo. Per il primo campione di bottiglie si trova un contenuto medio di vino pari a $\overline{x_1} = 749.5\,\mathrm{cl}$ con varianza campionaria $S_1^2 = 3.5\,\mathrm{cl}^2$; per il secondo si trova un contenuto medio pari a $\overline{x_2} = 750.6\,\mathrm{cl}$ con varianza campionaria $S_2^2 = 4\,\mathrm{cl}^2$. Assumendo che la varianza delle due popolazioni di bottiglie sia la stessa (ma non nota):

- Verificare l'ipotesi nulla "le bottiglie dei due impianti hanno lo stesso contenuto medio" con un livello di significatività del 5% e del 10%.
- Determinare il *p-value* del test.

Scritto A

(1) E = "Esce sei almeno una volta"

- $p(E) = 1 - p(E^c)$

$$p(E^c) = \sum_{n=0}^{\infty} p(E^c \mid N=n) \cdot p(N=n) = \sum_{n=0}^{\infty} \left(\frac{5}{6}\right)^n e^{-6/5} \frac{6^n}{5^n} \frac{1}{n!} =$$

$$= e^{-6/5} \sum_{n=0}^{\infty} \frac{1}{n!} = e^{-6/5} \cdot e = e^{-1/5}$$

$$p(E) = 1 - \frac{1}{\sqrt[5]{e}}.$$

- F = "Esce 6 esattamente una volta"

$$p(F) = \sum_{n=1}^{\infty} p(F \mid N=n) \cdot p(N=n) = \sum_{n=1}^{\infty} \binom{n}{1} \frac{1}{6} \left(\frac{5}{6}\right)^{n-1} \frac{6^n}{5^n} \frac{1}{n!}$$

$$= \frac{1}{6} \sum_{n=1}^{\infty} n \frac{\cancel{5^{n-1}}}{\cancel{6^{n-1}}} \frac{6^{\cancel{n}}}{5^{\cancel{n}}} \frac{1}{n!} = \frac{1}{\cancel{6}} \cdot \frac{1}{5} \cdot \cancel{6} \sum_{n=1}^{\infty} \frac{n}{n!}$$

$$= \frac{1}{5} \sum_{n=1}^{\infty} \frac{1}{(n-1)!} = \frac{1}{5} \sum_{n=0}^{\infty} \frac{1}{n!} = \frac{1}{5} e = \frac{e}{5}$$

- $$p(N=n \mid E^c) = \frac{p(E^c \mid N=n) \cdot p(N=n)}{p(E^c)} = \frac{\cancel{\left(\frac{5}{6}\right)^n} \frac{\cancel{6^n}}{\cancel{5^n}} \frac{1}{n!}}{e^{-1/5}} =$$

$$\frac{e^{1/5}}{n!} \xrightarrow[n \to \infty]{} 0$$

② $f(x) = \begin{cases} \frac{3}{a^3} x^2 & 0 \le x \le a \\ 0 & \text{altrimenti} \end{cases}$

- $f(x) \ge 0 \quad \forall x \in \mathbb{R}$ e $\int_{\mathbb{R}} f(x)dx = \int_0^a \frac{3}{a^3} x^2 dx = \frac{3}{a^3} \frac{x^3}{3}\Big|_0^a = \frac{3}{a^3}\frac{a^3}{3} = 1$

- $E[X] = \int_{\mathbb{R}} x f(x) dx = \int_0^a \frac{3}{a^3} x^3 dx = \frac{3}{a^3} \frac{x^4}{4}\Big|_0^a = \frac{3}{a^3}\frac{a^4}{4} = \frac{3}{4} a$

- Dato che $a = \frac{4}{3} E[X]$, si ha che $T_n = \frac{4}{3} \frac{X_1 + X_2 + \dots + X_n}{n}$

è uno stimatore corretto di a

- Per la legge debole dei grandi numeri $\frac{X_1 + X_2 + \dots + X_n}{n}$ tende

in probabilità a $E[X] = \frac{3}{4} a$; ne segue che T_n converge

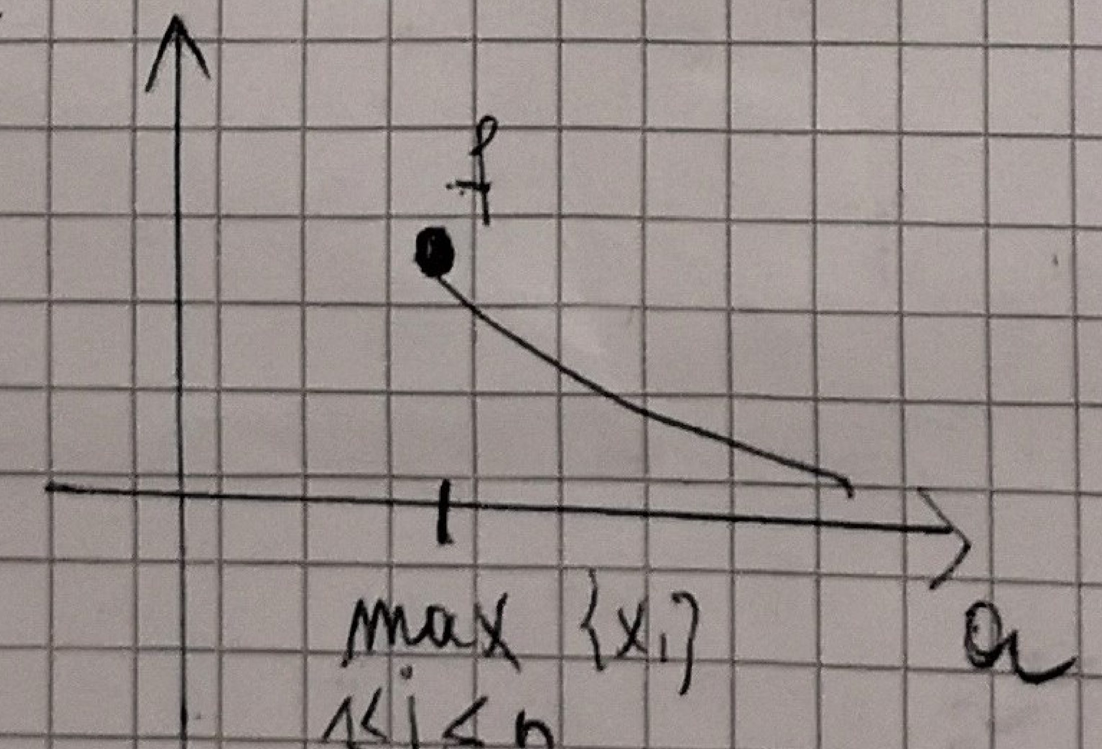
in probabilità ad a, cioè lo stimatore è consistente

- La densità congiunta di $(X_1, X_2, \dots, X_n)$ è

$$f(x_1, x_2, \dots, x_n \mid a) = \begin{cases} \frac{3^n}{a^{3n}} x_1^2 x_2^2 \dots x_n^2 & \text{se } 0 \le x_i \le a \ \forall i \\ 0 & \text{altrimenti} \end{cases}$$

Se $a \ge x_i \quad \forall i = 1, 2, \dots, n$, cioè se $a \ge \max_{1 \le i \le n} x_i$

la f.n.e vale $\frac{3^n}{a^{3n}} x_1^2 x_2^2 \dots x_n^2$; altrimenti è nulla



La verosimiglianza è, dunque, massima per $a = \max_{1 \le i \le n} x_i$

Lo stimatore di massima verosimiglianza è allora

$\max_{1 \le i \le n} \{X_i\}$

③ $n_1 = n_2 = 25$

$\overline{X_1} = 749.5\ cl \qquad S_1^2 = 3.5\ cl^2$

$\overline{X_2} = 750.6\ cl \qquad S_2^2 = 4\ cl^2$

$H_0 : \mu_1 = \mu_2$

- Si usa la statistica $T = \dfrac{\overline{X_1} - \overline{X_2}}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$

con $S_p = \sqrt{\dfrac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-1}}$

che è una t di Student con $n_1+n_2-2 = 48$ gradi di libertà

Si ha $S_p = \sqrt{\dfrac{24 \cdot 3{,}5 + 24 \cdot 4}{49}} = 1.916629615$

per cui $T = \dfrac{749.5 - 750.6}{1.916629615} \cdot \dfrac{5}{\sqrt{2}} = -2.029 \simeq -2.03$

Si rifiuta H_0 se $|T| > t_{1-\frac{\alpha}{2}}(n_1+n_2-2)$

Con $\alpha = 5\%$ $\quad 1 - \frac{\alpha}{2} = 0.975$

Ora $t_{0.975}(48) \simeq t_{0.975}(50) = 2.00856$

Quindi si rifiuta H_0 con $\alpha = 5\%$. A maggior ragione si rifiuta H_0 anche al 10%.

- Dato che il numero di gradi di libertà è abbastanza alto, approssimiamo i quantili della t di Student con quelli gaussiani; e il p-value è allora dato da

$z_{1-\frac{p}{2}} = 2.029; \quad 1 - \frac{p}{2} = \mathrm{Erf}(2.029); \quad p = 2(1 - \mathrm{Erf}(2.029)) \simeq 4.2\%$

Esame di *Calcolo delle probabilità e statistica* (per studenti di Informatica) corso B

Università degli studi di Bari Aldo Moro

08-06-2021

Esercizio 1. Si lancia N volte un dado equo a sei facce, numerate da 1 a 6, dove N è una variabile aleatoria di Poisson di parametro $\lambda = \frac{6}{5}$.

- Determinare la probabilità di ottenere 6 almeno una volta.
- Determinare la probabilità di ottenere 6 esattamente una volta.
- Sapendo di aver ottenuto 6 esattamente una volta, calcolare la probabilità che N sia pari a n, per ogni $n \in \mathbb{N}$, e dire per quali valori di n tale probabilità è massima.

(Si ricorda che $\sum_{n=0}^{\infty} \frac{1}{n!} = e$.)

Esercizio 2. Verificare che per ogni $a > 0$ la funzione $f(x) = \frac{4}{a^4} x^3 \chi_{[0,a]}(x)$ è la densità di probabilità di una variabile aleatoria X.

- Determinare l'attesa di X.
- Sfruttare il calcolo precedente per esibire uno stimatore corretto di a con un campione di rango n distribuito come X.
- Dire se lo stimatore ottenuto al punto precedente è consistente.
- Determinare lo stimatore di massima verosimiglianza di a relativo a un campione di rango n.

Esercizio 3.

Una casa vinicola ha due impianti di imbottigliamento. Si vuole controllare che il contenuto medio di vino per le bottiglie dei due impianti sia lo stesso. A tal scopo si analizzano $n_1 = 30$ bottiglie provenienti dal primo impianto e $n_2 = 30$ provenienti dal secondo. Per il primo campione di bottiglie si trova un contenuto medio di vino pari a $\overline{x_1} = 749.3\,\mathrm{cl}$ con varianza campionaria $S_1^2 = 3.8\,\mathrm{cl}^2$; per il secondo si trova un contenuto medio pari a $\overline{x_2} = 750.1\,\mathrm{cl}$ con varianza campionaria $S_2^2 = 4\,\mathrm{cl}^2$. Assumendo che la varianza delle due popolazioni di bottiglie sia la stessa (ma non nota):

- Verificare l'ipotesi nulla "le bottiglie dei due impianti hanno lo stesso contenuto medio" con un livello di significatività del 5% e del 10%.
- Determinare il *p-value* del test.

Scritto B

1) I primi due punti sono come nello scritto A

- $p(N=n \mid F) = \dfrac{p(F \mid N=n) \cdot p(N=n)}{p(F)}$

con $n=0$ la probabilità di sopra è evidentemente nulla

con $n \geq 1$ si ha

$$p(N=n \mid F) = \left(n \; \tfrac{1}{6} \left(\tfrac{5}{6}\right)^{n-1} e^{-6/5} \frac{6^n}{5^n} \frac{1}{n!}\right) \cdot \frac{5}{e}$$

$$= \frac{n}{\cancel{6}} \frac{\cancel{5^{n-1}}}{\cancel{6^{n-1}}} \frac{\cancel{6^n}}{5^{\cancel{n}}} e^{-6/5} \frac{1}{n!} \frac{5}{e} =$$

$$= \frac{n}{\cancel{5}} e^{-6/5} \frac{\cancel{5}}{e} \frac{1}{n!} = \frac{1}{(n-1)!} e^{-\frac{6}{5}-1}$$

che è massima per $n=1$ e $n=2$ (cioè quando $(n-1)!$ è minimo)

2) $f(x) = \begin{cases} \frac{4}{a^4} x^3 & 0 \leq x \leq a \\ 0 & \text{altrove} \end{cases}$

$f(x) \geq 0 \;\; \forall x \in \mathbb{R}$; $\displaystyle\int_{\mathbb{R}} f(x)\,dx = \int_0^a \frac{4}{a^4} x^3\,dx = \frac{4}{a^4} \frac{x^4}{4}\Big|_0^a =$

$$= \frac{4}{a^4} \frac{a^4}{4} = 1$$

- $\displaystyle E[X] = \int_{\mathbb{R}} x f(x)\,dx = \int_0^a \frac{4}{a^4} x^4\,dx$

$$= \frac{4}{a^4} \frac{x^5}{5}\Big|_0^a = \frac{4}{a^4} \frac{a^5}{5} = \frac{4}{5} a$$

- Dato che $a = \frac{5}{4} E[X]$, $T_n := \frac{5}{4} \dfrac{X_1 + X_2 + \ldots + X_n}{n}$

è uno stimatore corretto di a

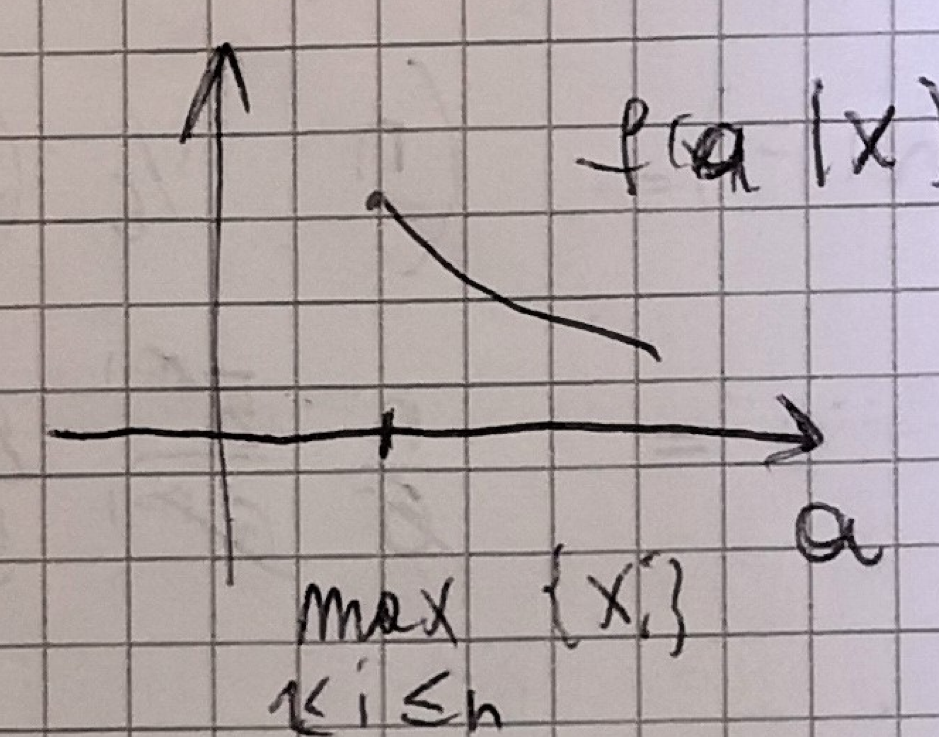
- Lo stimatore T_n è consistente per la legge debole dei grandi numeri (vedi compito A)

- La densità congiunta del campione $(X_1, X_2, \ldots, X_n)$ è

$$f(x_1, x_2, \ldots, x_n | a) = \begin{cases} \dfrac{4^n}{a^{4n}} x_1^3 x_2^3 \ldots x_n^3 & \text{per } a \geq x_i \quad \forall_i \\ 0 & \text{altrimenti} \end{cases}$$

Ora $a \geq x_i \quad \forall i = 1, 2, \ldots, n$ è lo stesso che

$a \geq \max\limits_{1 \leq i \leq n} \{x_i\}$



La f.ne f raggiunge il suo massimo per $a = \max\limits_{1 \leq i \leq n} \{X_i\}$

Lo stimatore di massima verosimiglianza è dunque

$\max\limits_{1 \leq i \leq n} \{X_i\}$.

(3) $n_1 = n_2 = 30$

$\overline{X_1} = 749.3 \, cl \qquad S_1^2 = 3.8 \, cl^2$

$\overline{X_2} = 750.1 \, cl \qquad S_2^2 = 4 \, cl^2$

$H_0: \mu_1 = \mu_2$

Si usa la statistica $T = \dfrac{\overline{X}_1 - \overline{X}_2}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$

con $S_p = \sqrt{\dfrac{(n_1 - 1) S_1^2 + (n_2 - 1) S_2^2}{n_1 + n_2 - 1}}$

che è una t di Student con $n_1 + n_2 - 2 = 58$ gradi di libertà

Si ha $S_p = \sqrt{\frac{29 \cdot 3,8 + 29 \cdot 4}{59}} = 1.958034$

$T = \frac{749.3 - 750.1}{1.958034} \cdot \sqrt{\frac{29}{2}} \simeq -1.556$

Si rifiuta H_0 se $|t| > t_{1-\frac{\alpha}{2}}(n_1+n_2-2)$

Con $\alpha = 5\%$ $1-\frac{\alpha}{2} = 0.975$ e $t_{0.975}(58)$

$\sim t_{0.975}(60) = 2.003$

Si accetta H_0

Con $\alpha = 10\%$ $1-\frac{\alpha}{2} = 0.95$ e $t_{0.95}(58) \sim t_{0.95}(60)$

$= 1.670$

Si accetta ancora H_0

- Dato che il numero di gradi di libertà è abbastanza alto, approssimiamo i quantili di Student con quelli gaussiani; il p-value del test è allora dato da

$$z_{1-\frac{p}{2}} = 1.56$$

$$1-\frac{p}{2} = \mathrm{Erf}(1.56), \quad p = 2(1-\mathrm{Erf}(1.56))$$

$\sim 11\%$

Esame di *Calcolo delle probabilità e statistica* (per studenti di Informatica)
corso A
Università degli studi di Bari Aldo Moro

24–06–2021

Esercizio 1. Si lancia N volte ua moneta equa, dove N è una variabile aleatoria geometrica di parametro p, vale a dire $P[N = n] = p(1-p)^{n-1}$ per $n \geq 1$.

- Determinare la probabilità di ottenere almeno una testa.
- Determinare la probabilità che N sia n sapendo di non aver ottenuto mai testa.
- Calcolare il limite per $n \to \infty$ della probabilità calcolata al punto precedente.

(Si ricorda che $\sum_{n=0}^{\infty} \frac{1}{n!} = e$.)

Esercizio 2. Verificare che per ogni $\theta \geq 1$ la funzione $f(x) = \frac{\theta}{2^\theta} x^{\theta-1} \chi_{[0,2]}(x)$, dove $\chi_{[0,1]}(x)$ vale 1 per $0 \leq x \leq 2$ e 0 altrimenti, è la densità di probabilità di una variabile aleatoria X.

- Calcolare l'attesa di X.
- Calcolare l'attesa di $\ln(X)$, verificando che è pari a $\ln(2) - \frac{1}{\theta}$.
- Dato un campione $(X_1, X_2, \ldots, X_n)$ di rango n distribuito come X, determinare il corrispondente stimatore di massima verosimiglianza di θ.
- Dire se lo stimatore di massima verosimiglianza ottenuto al punto precedente è consistente.

Esercizio 3. Una casa produttrice sostiene di produrre transistor bipolari con un valore medio del guadagno almeno di 210. Si prova un campione di $n = 71$ transistor trovando una deviazione standard campionaria $\overline{x} = 202$ con una varianza *campionaria* $S = 43$.

- Dire quale test statistico occorre condurre per verificare l'affermazione della casa produttrice, ed effettuarlo con un livello di significatività del 5% e del 10%.
- Calcolare il p-value del test.

(A)

(1)

- Con le stesse notazioni del compito B

$$p(N=n \mid E^c) = \frac{p(E^c \mid N=n)}{p(E^c)} = \left(\frac{1}{2}\right)^n \cdot \frac{1+p}{p}$$

- $$\lim_{n\to\infty} p(N=n \mid E^c) = \frac{1+p}{p} \lim_{n\to\infty} \left(\frac{1}{2}\right)^n = 0$$

(2) $$f(x) = \begin{cases} \frac{\theta}{2^\theta} x^{\theta-1} & \text{per } 0 \le x \le 2 \\ 0 & \text{altrove} \end{cases}$$

- $f(x) \geq 0 \quad \forall x \in \mathbb{R}$ e $\int_{\mathbb{R}} f(x)\,dx = \frac{\theta}{2^\theta} \int_0^2 x^{\theta-1}\,dx =$

$$= \frac{\theta}{2^\theta} \frac{x^\theta}{\theta}\Big|_0^2 = \frac{\theta}{2^\theta} \cdot \frac{2^\theta}{\theta} = 1$$

- $$E[X] = \int_{\mathbb{R}} x f(x)\,dx = \frac{\theta}{2^\theta} \int_0^2 x \cdot x^{\theta-1}\,dx = \frac{\theta}{2^\theta} \int_0^2 x^\theta\,dx =$$

$$= \frac{\theta}{2^\theta} \frac{x^{\theta+1}}{\theta+1}\Big|_0^2 = \frac{\theta}{\cancel{2^\theta}} \cdot \frac{2^{\cancel{\theta}+1}}{\theta+1} = 2\frac{\theta}{\theta+1}$$

- $$E[\ln X] = \int_{\mathbb{R}} \ln x\, f(x)\,dx = \int_0^2 \underbrace{\ln x}_{g} \frac{\theta}{2^\theta} \underbrace{x^{\theta-1}}_{f'}\,dx$$

$$= \frac{\theta}{2^\theta}\left[\frac{x^\theta}{\theta} \ln x \Big|_0^2 - \int_0^2 \frac{x^\theta}{\theta} \frac{1}{x}\,dx\right]$$

$$= \frac{\theta}{2^\theta}\left[\frac{2^\theta}{\theta}\ln 2 - \frac{1}{\theta}\int_0^2 x^{\theta-1}\,dx\right] =$$

$$\frac{\theta}{2^\theta}\left[\frac{2^\theta}{\theta}\ln 2 - \frac{1}{\theta}\,\frac{x^\theta}{\theta}\Big|_0^2\right] = \frac{\cancel{\theta}}{\cancel{2^\theta}}\,\frac{\cancel{2^\theta}}{\cancel{\theta}}\ln 2 - \frac{\theta}{\cancel{2^\theta}}\cdot\frac{1}{\theta^2}\,2^\theta$$

$$= \ln 2 - \frac{1}{\theta}$$

- $$f(x_1, x_2, \dots, x_n|\theta) = \begin{cases} \dfrac{\theta^n}{2^{n\theta}}\, x_1^{\theta-1} \cdots x_n^{\theta-1} & 0 \le x_i \le 2 \\ 0 & \text{altrimenti} \end{cases}$$

$$\ln f(x_1, x_2, \dots, x_n|\theta) = n\ln\theta - n\theta\ln 2 + (\theta-1)\sum_{i=1}^n \ln x_i$$

$$\frac{\partial}{\partial\theta}\ln f(x_1, x_2, \dots, x_n|\theta) = \frac{n}{\theta} - n\ln 2 + \sum_{i=1}^n \ln x_i = 0$$

$$\frac{n}{\theta} = n\ln 2 - \sum_{i=1}^n \ln x_i$$

$$\theta = \frac{n}{n\ln 2 - \sum_{i=1}^n \ln x_i} = \frac{1}{\ln 2 - \frac{1}{n}\sum_{i=1}^n \ln x_i}$$

$$\hat{\theta}_n = \frac{1}{\ln 2 - \frac{1}{n}\sum_{i=1}^n \ln X_i}$$

- Per la legge debole dei grandi numeri

$$\frac{1}{n}\sum_{i=1}^n \ln X_i \xrightarrow[n\to\infty]{P} \ln 2 - \frac{1}{\theta}$$

per cui $\ln 2 - \frac{1}{n}\sum_{i=1}^{n} \ln X_i \xrightarrow{P} \frac{1}{\theta}$

e allora $\hat{\theta}_n \xrightarrow{P} \theta$.

(3) Come nel corrispondente esercizio del compito B

$$T = \frac{\overline{X} - \mu}{\frac{S}{\sqrt{n}}} = \frac{202 - 210}{43}\sqrt{71} = -1.57$$

Zona di rifiuto $T < -t_{1-\alpha}(n-1)$

- $\alpha = 0.05$ $\quad t_{0.95}(70) = 1.67$

quindi si accetta H_0

- $\alpha = 0.1$ $\quad t_{0.90}(70) = 1.29$

quindi si rifiuta H_0

Calcolo (approssimato) del p-value

$-1.57 = -z_{1-p}$ $\qquad 1 - p = Erf(1.57)$

$p = 1 - Erf(1.57) = 1 - 0.9418 = 6\%$

Esame di *Calcolo delle probabilità e statistica* (per studenti di Informatica)
corso B

Università degli studi di Bari Aldo Moro

24–06–2021

Esercizio 1. Si lancia N volte ua moneta equa, dove N è una variabile aleatoria geometrica di parametro p, vale a dire $P[N = n] = p(1-p)^{n-1}$ per $n \geq 1$.

- Determinare la probabilità di ottenere almeno una testa.
- Determinare la probabilità che N sia n sapendo di aver ottenuto almeno una testa.
- Calcolare il limite per $n \to \infty$ della probabilità calcolata al punto precedente.

(Si ricorda che, per ogni q con $|q| < 1$, si ha $\sum_{n=0}^{\infty} q^n = \frac{1}{1-q}$.)

Esercizio 2. Verificare che per ogni $\theta \geq 1$ la funzione $f(x) = \theta x^{\theta-1}\chi_{[0,1]}(x)$, dove $\chi_{[0,1]}(x)$ vale 1 per $0 \leq x \leq 1$ e 0 altrimenti, è la densità di probabilità di una variabile aleatoria X.

- Calcolare l'attesa di X.
- Calcolare l'attesa di $\ln(X)$, verificando che è pari a $-\frac{1}{\theta}$.
- Dato un campione $(X_1, X_2, \ldots, X_n)$ di rango n distribuito come X, determinare il corrispondente stimatore di massima verosimiglianza di θ.
- Dire se lo stimatore di massima verosimiglianza ottenuto al punto precedente è consistente.

Esercizio 3.

Una casa produttrice sostiene di produrre transistor bipolari con un valore medio del guadagno almeno di 210. Si prova un campione di $n = 51$ transistor trovando una deviazione standard campionaria $\overline{x} = 200$ con una varianza *campionaria* $S = 35$.

- Dire quale test statistico occorre condurre per verificare l'affermazione della casa produttrice, ed effettuarlo con un livello di significatività del 5% e del 10%.
- Calcolare il p-value del test.

Esercizio 1.

- $E =$ " Esce almeno una testa"

$p(E) = 1 - p(E^c)$

$E^c =$ " Non esce nemmeno una testa"

$p(E^c) = \sum_{n=1}^{\infty} p(E^c|N=n) \cdot p(N=n)$ (probabilità totale)

Ora $p(E^c|N=n) = \left(\frac{1}{2}\right)^n$ per cui

$$\sum_{n=1}^{\infty} p(E^c|N=n) \cdot p(N=n) = \sum_{n=1}^{\infty} \left(\frac{1}{2}\right)^n p(1-p)^{n-1} = \frac{p}{2} \sum_{n=1}^{\infty} \left(\frac{1-p}{2}\right)^{n-1} =$$

$$= \frac{p}{2} \sum_{n=0}^{\infty} \left(\frac{1-p}{2}\right)^n = \frac{p}{2} \frac{1}{1 - \frac{1-p}{2}} = \frac{\cancel{p}}{\cancel{2}} \frac{\cancel{2}}{2-1+p} = \frac{p}{1+p}$$

$$p(E) = 1 - \frac{p}{1+p} = \frac{1+p-p}{1+p} = \frac{1}{1+p}$$

- $p(N=n|E) = \dfrac{p(E|N=n)}{p(E)} = \dfrac{1-\left(\frac{1}{2}\right)^n}{1+p} =$

$p(E|N=n) = 1 - p(E^c|N=n) = 1 - \left(\frac{1}{2}\right)^n$

- $\lim_{n\to\infty} p(N=n|E) = \lim_{n\to\infty} \dfrac{1-\left(\frac{1}{2}\right)^n}{1+p} = \dfrac{1}{1+p}$

2) $f(x) = \begin{cases} \theta x^{\theta-1} & 0 \leq x \leq 1 \\ 0 & \text{altrimenti} \end{cases}$

$f(x) \geq 0 \quad \forall x \in \mathbb{R}$ e $\int_{\mathbb{R}} f(x)dx = \int_0^1 f(x)dx = \theta \int_0^1 x^{\theta-1} dx =$

$$\theta \left. \frac{x^\theta}{\theta} \right|_0^1 = 1$$

• $E[X] = \int_{\mathbb{R}} x f(x) dx = \int_0^1 \theta \, x \cdot x^{\theta-1} dx = \theta \int_0^1 x^\theta dx = \theta \left. \frac{x^{\theta+1}}{\theta+1} \right|_0^1 = \frac{\theta}{\theta+1}$

• $E[\ln X] = \int_{\mathbb{R}} \ln x \, f(x) dx = \int_0^1 \ln x \; \theta x^{\theta-1} =$

$\theta \int_0^1 \underbrace{\ln x}_{g} \; \underbrace{x^{\theta-1}}_{f'} dx = \theta \left[\cancel{\left. \frac{x^\theta}{\theta} \ln x \right|_0^1} - \int_0^1 \frac{x^\theta}{\theta} \cdot \frac{1}{x} \right]$

$= - \int_0^1 x^{\theta-1} dx = - \left. \frac{x^\theta}{\theta} \right|_0^1 = - \frac{1}{\theta}$

• $f(x_1, x_2, \dots, x_n | \theta) = \begin{cases} \theta^n \, x_1^{\theta-1} \, x_2^{\theta-1} \dots x_n^{\theta-1} & 0 \leq x_i \leq 1 \\ 0 & \text{altrimenti} \end{cases}$

$\ln f(x_1, \dots, x_n | \theta) = n \ln \theta + (\theta - 1) \sum_{i=1}^n \ln x_i$

$\frac{\partial}{\partial \theta} \ln f(x_1, \dots, x_n | \theta) = \frac{n}{\theta} + \sum_{i=1}^n \ln x_i \overset{!}{=} 0$

$\frac{n}{\theta} = - \sum_{i=1}^n \ln x_i \, , \qquad \frac{\theta}{n} = - \frac{1}{\sum_{i=1}^n \ln x_i} \, , \qquad \theta = - \frac{n}{\sum_{i=1}^n \ln x_i}$

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

• Per la legge debole dei grandi numeri

$$\frac{1}{n}\sum_{i=1}^{n} \ln X_i \xrightarrow{P} -\frac{1}{\theta}$$

Ne segue che $\dfrac{1}{-\frac{1}{n}\sum_{i=1}^{n} \ln X_i} \to \theta$ in probabilità

$= \frac{\theta}{\theta+1}$

3). $\mu_0 = 210$

$H_0: \overline{X} \geq \mu_0$

$n = 51$ $\quad \overline{X} = 200$ $\quad S = 35$

- Occorre eseguire un test di Student unilaterale.

La statistica da usare è

$$T = \frac{\overline{X} - \mu_0}{\frac{S}{\sqrt{n}}}$$

$i \leq 1$

La zona di rifiuto è $\quad T < -t_{1-\alpha}(n-1)$

ti.

Con i nostri dati $\quad T = \dfrac{200-210}{35}\sqrt{51} = -2.04$

Con $\alpha = 5\%$ $\quad t_{1-\alpha}(n-1) = t_{0.95}(50) = 1.675$

Quindi con $\alpha = 5\%$ H_0 viene rifiutata.

A maggior ragione si rifiuta H_0 anche al livello $\alpha = 10\%$

• Calcolo (approssimato) del p-value

$$-z_{1-p} = -2.04 \quad ; \quad 1-p = Erf(2.04)$$

$$p = 1 - Erf(2.04) = 1 - 0.9793 = 0.0261 \simeq 3\%.$$

Esame di *Calcolo delle probabilità e statistica* (per studenti di Informatica)
corso A e B
Università degli studi di Bari Aldo Moro

12-07-2021

Esercizio 1. Una scatola contiene 10 dadi con le facce numerate da 1 a 6. Uno di questi dadi è truccato e non fa mai uscire i numeri pari, mentre i numeri $1, 3, 5$ escono tutti con probabilità $\frac{1}{3}$. I restanti 9 dadi sono equilibrati. Si prende a caso un dado e lo si lancia 1000 volte.

- Calcolare la probabilità che il 2 esca meno di 50 volte.
- Sapendo che il 2 è uscito meno di 50 volte, determinare la probabilità che il dado sia truccato.
- Calolare la probabilità che il numero 1 esca più di 150 volte.

Esercizio 2. Verificare che per ogni valore del parametro $\theta > 0$ la funzione

$$f(x;\theta) := \frac{4x^2}{\sqrt{\pi}\theta^3} e^{-\frac{x^2}{\theta^2}} \chi_{[0,\infty]}(x)$$

è la densità di probabilità di una certa variabile aleatoria X.

- Determinare l'attesa di X, verificando che si ha $E[X] = \frac{2\theta}{\sqrt{\pi}}$.
- Sfruttare il calcolo precedente per esibire uno stimatore corretto di θ con un campione di rango n distribuito come X.
- Determinare lo stimatore di massima verosimiglianza di θ relativo a un campione di rango n distribuito come X.
- Esibire una statistica sufficiente.

(Per il calcolo degli integrali si ricorda che $\int_0^{+\infty} e^{-x^2} \mathrm{d}x = \frac{\sqrt{\pi}}{2}$.)

Esercizio 3. Data una popolazione gaussiana X, si effettuano $n = 10$ osservazioni, che danno i seguenti valori:

1.2	1.3	1.2	1.4	1.2
1.3	1.1	1.1	1.3	1.4

- Calcolare media e varianza campionaria.
- Verificare l'ipotesi "$\sigma_X = 0.1$" al livello di significatività dell' 1%.
- Verificare l'ipotesi "$\mu_X \geq 1.3$" con un livello di significatività del 5%.

In entrambi i casi dire quale test occorre condurre, specificando se è un test unilaterale o bilaterale.

Esercizio 1.

T = " Il dado è truccato " $p(T) = \frac{1}{10}$

T^c = " Il dado è equilibrato " $p(T^c) = \frac{9}{10}$

$n = 1000$ (numero di lanci del dado)

E = " Il due esce meno di 50 volte "

$p(E) = p(E|T) \cdot p(T) + p(E|T^c) \cdot p(T^c)$

Ora $p(E|T) = 1$ (se il dado è truccato, il 2 non uscirà mai)

Occorre calcolare

$p(E|T^c)$

In un singolo lancio la probabilità che esca 2 è $\frac{1}{6}$ (quando il dado è equilibrato)

Sia X la v.a. che conta il numero di volte che esce 2.

$X = \sum_{i=1}^{n} X_i$ con X_i la v.a. che vale 1 se all'i-esimo lancio esce 2 e 0 altrimenti

X_i è una v.a. binomiale di parametro $p = 1/6$

per cui $E[X_i] = 1/6$ $Var[X_i] = 1/6 \cdot 5/6 = 5/36$

$$\left(\sum_{i=1}^{1000} X_i - \frac{1000}{6} \right) \Big/ \sqrt{1000 \cdot 5/36}$$ è approssimativamente normale

Ora

$$P(E \mid T^c) = P(X \leq 50) = P\Big(\underset{\substack{\uparrow \\ \text{normale} \\ \text{standard}}}{Z} \leq \frac{50 - \frac{1000}{6}}{\sqrt{1000 \cdot 5/36}}\Big)$$

$$P(Z \leq -26.5) = 0$$

Quindi

$$p(E) = p(E \mid T) \cdot p(T) = 1 \cdot \frac{1}{10} = \frac{1}{10}$$

- $$p(T \mid E) = \frac{p(E \mid T) \cdot p(T)}{p(E)} = \underline{1}$$

- F = "Il numero 1 esca più di 150 volte"

$$p(F) = p(F \mid T) \cdot p(T) + p(F \mid T^c) \cdot p(T^c)$$

$p(F \mid T)$

Se il dado è truccato la probabilità di ottenere 1 ad ogni lancio è $1/3$

La media degli 1 ottenuti è dunque $\frac{1000}{3} = 333.3$

Con varianza $\sigma^2 = 1000 \cdot 1/3 \cdot \frac{2}{3} = 222.22$ e

deviazione standard $\sigma = 14.907$

Allora standardizzando

$$P(F|T) = P\left(Z \geq \frac{150 - 333.3}{14.907}\right) =$$

$$P(Z \geq -12.29) = 1$$

$P(F|T^c)$

se il dado è equilibrato la probabilità di ottenere 1 ad ogni lancio è $1/6$

La media degli 1 ottenuti è $1000/6 = 166.67$

con varianza

$$\sigma^2 \; 1000 \cdot \frac{1}{6} \cdot \frac{5}{6} = 138.88 \quad e \quad \sigma = 11.78$$

Quindi

$$P(F|T^c) = P\left(Z \geq \frac{150 - 166.67}{11.78}\right) =$$

$$P(Z \geq -1.42)$$

$$\cancel{1 -}\, \mathrm{Erf}(-1.42) = \mathrm{Erf}(1.42) = 0.922$$

$$P(F) = 1 \cdot \frac{1}{10} + 0.922 \cdot \frac{9}{10} = 0.93 \quad (93\%)$$

ESERCIZIO 2

$$f(x;\theta) = \begin{cases} \dfrac{4x^2}{\sqrt{\pi}\,\theta^3}\, e^{-\frac{x^2}{\theta^2}} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

- $f(x,\theta) \geq 0 \quad \forall x \in \mathbb{R}$

Inoltre

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_0^{+\infty} \frac{4x^2}{\sqrt{\pi}\,\theta^3}\, e^{-\frac{x^2}{\theta^2}}\,dx =$$

$$\frac{2}{\sqrt{\pi}\,\theta} \int_0^{\infty} \underbrace{x}_{g}\ \underbrace{\frac{2x}{\theta^2}\, e^{-\frac{x^2}{\theta^2}}}_{f'}\,dx = \frac{2}{\sqrt{\pi}\,\theta}\left[\cancel{-e^{-\frac{x^2}{\theta^2}}\, x}\Big|_0^{\infty} - \int_0^{\infty} -e^{-\frac{x^2}{\theta^2}}\,dx \right]$$

$$= \frac{2}{\sqrt{\pi}\,\theta} \int_0^{\infty} e^{-\frac{x^2}{\theta^2}}\,dx = \frac{2}{\sqrt{\pi}\,\theta} \int_0^{\infty} e^{-t^2}\,\theta\,dt = \frac{2}{\cancel{\sqrt{\pi}\,\theta}}\,\cancel{\theta}\,\frac{\cancel{\sqrt{\pi}}}{\cancel{2}} = 1$$

- $$E[X] = \int_{-\infty}^{\infty} x\, f(x,\theta)\,dx = \int_0^{\infty} \frac{4x^3}{\sqrt{\pi}\,\theta^3}\, e^{-\frac{x^2}{\theta^2}}\,dx$$

$$= \frac{2}{\sqrt{\pi}}\,\frac{1}{\theta} \int_0^{\infty} \underbrace{x^2}_{g} \cdot \underbrace{\frac{2x}{\theta^2}\, e^{-\frac{x^2}{\theta^2}}}_{f'} = \frac{2}{\sqrt{\pi}\,\theta}\left[e^{-\frac{x^2}{\theta^2}}\, x^2 \Big|_0^{\infty} - \int_0^{\infty} -e^{-\frac{x^2}{\theta^2}}\, 2x \right]$$

$$= \frac{2}{\sqrt{\pi}\,\theta}\,\theta^2 \int_0^{\infty} \frac{2x}{\theta^2}\, e^{-\frac{x^2}{\theta^2}}\,dx = \frac{2\theta}{\sqrt{\pi}}\left(-e^{-\frac{x^2}{\theta^2}}\right)\Big|_0^{\infty} = \frac{2\theta}{\sqrt{\pi}}$$

- Dato che $\theta = \frac{\sqrt{\pi}}{2}\, E[X]$, si ha che

$\frac{X_1 + X_2 + \dots + X_n}{n} \cdot \frac{\sqrt{\pi}}{2}$ è uno stimatore corretto di θ

$$f(x_1, x_2, \dots, x_n \mid \theta) = \begin{cases} \dfrac{4^n \, x_1^2 \, x_2^2 \dots x_n^2}{\pi^{n/2} \quad \theta^{3n}} \quad e^{-\sum_{i=1}^{n} \frac{x_i^2}{\theta^2}} & x_i \geq 0 \\ 0 & \text{altrimenti} \end{cases}$$

$$\ln f(x_1, x_2, \dots, x_n \mid \theta) = -3n \ln \theta - \frac{1}{\theta^2} \sum_{i=1}^{n} x_i^2$$

(a meno di costanti additive)

$$\frac{\partial}{\partial \theta} \ln f(x_1, \dots, x_n \mid \theta) = -\frac{3n}{\theta} + \frac{2}{\theta^3} \sum_{i=1}^{n} x_i^2 \overset{!}{=} 0$$

$$\frac{2}{\cancel{\theta^3}\,\theta^2} \sum_{i=1}^{n} x_i^2 = \frac{3n}{\cancel{\theta}}$$

$$\frac{\theta^2}{2 \sum_{i=1}^{n} x_i^2} = \frac{1}{3n}$$

$$\theta^2 = \frac{2}{3n} \sum_{i=1}^{n} x_i^2 \quad ; \quad \theta = \left(\frac{2}{3n} \sum_{i=1}^{n} x_i^2 \right)^{1/2}$$

$$\hat{\theta}_n = \left(\frac{2}{3n} \sum_{i=1}^{n} X_i^2 \right)^{1/2}$$

- Dal teorema di fattorizzazione di Fisher si vede che

$T = \sum_{i=1}^{n} X_i^2$ è una statistica sufficiente

Esercizio 3

Si ha

$\overline{x} = \frac{1}{10} \sum_{i=1}^{10} x_i = 1.25$

$\overline{x^2} = \frac{1}{10} \sum_{i=1}^{10} x_i^2 = 1.573$

$S^2 = \frac{n}{n-1} (\overline{x^2} - \overline{x}^2) = \frac{10}{9} (1.573 - 1.25^2)$

$= 0.01166$

$S = 0.108$

– " $\sigma_0 = 0.1$ " ipotesi nulla con $\alpha = 0.01$

Occorre eseguire un test bilaterale

la zona di accettazione è

$\chi^2_{\alpha/2}(n-1) \leq \frac{S^2}{\sigma_0^2}(n-1) \leq \chi^2_{1-\frac{\alpha}{2}}(n-1)$

Nel nostro caso $\frac{S^2}{\sigma_0^2}(n-1) = \frac{0.01166}{0.1^2} \cdot 9 = 10.494$

Dato che $\chi^2_{0.005}(9) = 1.735$ e $\chi^2_{0.995}(9) = 23.589$

si accetta H_0

– $\mu_X \geq \mu_0 = 1.3$ $\alpha = 0.05$

test unilaterale di student (la varianza non è nota)

Si usa la statistica

$$T = \frac{\overline{X} - \mu_0}{\frac{S}{\sqrt{n}}}$$

che assume il valore $T = \frac{1.25 - 1.3}{0.108} \sqrt{10} = -1.464$

e si rifiuta se $T < - t_{1-\alpha}(n-1)$

↑ quantile Student

Ora $t_{0.95}(9) = 3.24$

quindi H_0 è accettata.

Esame di *Calcolo delle probabilità e statistica* (per studenti di Informatica)
corso A e B
Università degli studi di Bari Aldo Moro

06-09-2021

Esercizio 1. Si hanno due monete a due facce, testa e croce. Una moneta è equa, mentre l'altra è truccata e dà testa con probabilità $\frac{2}{3}$. Si sceglie una delle due monete e la si lancia $n = 100$ volte.

- Calcolare la probabilità di ottenere più di 65 teste.
- Calcolare la probabilità che la moneta sia truccata se si sono ottenute più di 65 teste.
- Calcolare la probabilità che la moneta sia equa se si sono ottenute meno di 65 teste.

Esercizio 2. Verificare che per ogni valore del parametro $a > 0$ la funzione

$$f(x; a) := \frac{5a^5}{x^6} \chi_{[a,\infty]}(x)$$

è la densità di probabilità di una certa variabile aleatoria X.

- Determinare l'attesa di X, verificando che si ha $E[X] = \frac{5}{4}a$.
- Sfruttare il calcolo precedente per esibire uno stimatore corretto di a a partire da un campione di rango n distribuito come X.
- Determinare lo stimatore di massima verosimiglianza di a relativo a un campione di rango n distribuito come X.
- Esibire una statistica sufficiente.

Esercizio 3.

- Fornire la definizione di intervallo di confidenza a un fissato livello di confidenza $1 - \alpha$, con $0 < \alpha < 1$, per un parametro incognito di una data distribuzione.
- Si sospetta che una certa terapia farmacologica possa aumentare la probabilità di concepire un bambino di sesso femminile. Su 574 gravidanze in cui la madre era sotto questo trattamento, 525 hanno dato una bambina. Costruire un intervallo di fiducia al 95% per la probabilità di concepire una bambina quando la madre è sotto trattamento al momento del concepimento.

Esercizio 1

E = "La moneta è equa" $\quad p(E) = \frac{1}{2}$

$T = E^c$ = "La moneta è truccata" $\quad p(T) = \frac{1}{2}$

F = " In $n = 100$ lanci si ottengono più di 65 teste"

- $p(F) = p(F|E) \cdot p(E) + p(F|T) \cdot p(T) = \frac{1}{2} [p(F|E) + p(F|T)]$

Calcolo di $p(F|E)$.

Se la moneta è equa, la v.a. X che conta il numero di teste è binomiale di parametri $p = \frac{1}{2}$ e $n = 100$, pertanto ha media

$\mu = np = 100 \cdot \frac{1}{2} = 50$ e varianza $\sigma^2 = n \cdot p \cdot (1-p) = 100 \cdot \frac{1}{2} \cdot \frac{1}{2} = 25$

$\Rightarrow \sigma = 5$

Allora $P[X \geq 65] = P\left[\frac{X-50}{5} \geq \frac{65-50}{5}\right] = P(Z \geq 3)$

$= 1 - Erf(3)$

$= 1 - 0,99861$

$= 0,00139$

Calcolo di $p(F|T)$

Ora X ha parametri $p = \frac{2}{3}$ e $n = 100$ per cui $\mu = 66,66$

e $\sigma^2 = 100 \frac{2}{3} \cdot \frac{1}{3} = 22.22$ e $\sigma = 4,714$

Quindi abbiamo $p(F|T) = p(X \geq 65) = p\left(\frac{X-66}{4.714} \geq \frac{65-66.66}{4.714}\right) =$

$= p(Z \geq -0.35) = p(Z \leq 0.35) = Erf(0.35) = 0.63683$

$p(F) = \frac{1}{2} (0.00139 + 0.63683) = 0.319$

- $p(T|F) = \frac{p(F|T) \cdot p(T)}{p(F)} = \frac{0.63683 \cdot \frac{1}{2}}{0.319} = 0.998$

$\cdot\ p(E \mid F^c) = \dfrac{p(F^c \mid E) \cdot p(E)}{p(F^c)}$

$p(F^c) = 1 - p(F) = 1 - 0.319 = 0.681$

$p(F^c \mid E) = P(X \leq 65)$

$\downarrow$

Gh. $\mu = 50$ e $\sigma = 5$

$= P\left(\dfrac{X-50}{5} \leq \dfrac{65-50}{5} \right) = P(Z \leq 3) = 0{,}99861$

Ma oltre

$p(E \mid F^c) = \dfrac{0{,}99861 \cdot \frac{1}{2}}{0{,}681} = 0{,}733$

Esercizio n°2

$$f(x;a) = \begin{cases} \frac{5a^5}{x^6} & \text{se } x \geq a \\ 0 & \text{se } x < a \end{cases}$$

- $f(x;a) \geq 0 \quad \forall x \in \mathbb{R}$ e inoltre $\int_{-\infty}^{\infty} f(x)\,dx = \int_{a}^{+\infty} f(x)\,dx = 5a^5 \int_{a}^{+\infty} x^{-6}\,dx$

$$= 5a^5 \left.\frac{x^{-5}}{-5}\right|_a^{\infty} = 5a^5\left(0-\left(-\frac{a^{-5}}{5}\right)\right) = 5a^5 \cdot \frac{a^{-5}}{5} = 1$$

- $E[X] = \int_{-\infty}^{\infty} x f(x)\,dx = 5a^5 \int_{a}^{+\infty} x^{-5}\,dx = 5a^5 \left.\frac{x^{-4}}{-4}\right|_a^{\infty} = \frac{5}{4} a^5 a^{-4} = \frac{5}{4} a$

- Dato che $a = \frac{4}{5} E[X]$, segue che $\frac{4}{5}\bar{X} = \frac{4}{5} \frac{X_1 + X_2 + \dots + X_n}{n}$ è uno stimatore corretto di a

- La densità congiunta è data da $f(x_1, \dots, x_n; a) = \frac{5^n a^{5n}}{x_1^6 x_2^6 \dots x_n^6} \chi_{[a,\infty)}(x_1) \dots \chi_{[a,\infty)}(x_n)$

$$= \begin{cases} \frac{5^n a^{5n}}{x_1^6 \dots x_n^6} & \text{se } x_1 \geq a \text{ e } x_2 \geq a \dots \text{ e } x_n \geq a \\ 0 & \text{altrimenti} \end{cases}$$

$$= \begin{cases} \frac{5^n a^{5n}}{x_1^6 \dots x_n^6} & a \leq x_1 \text{ e } a \leq x_2 \text{ e } \dots a \leq x_n \\ 0 & \text{altrimenti} \end{cases}$$

$$= \begin{cases} \frac{5^n a^{5n}}{x_1^6 \dots x_n^6} & \text{se } a \leq \min\{x_1, x_2, \dots x_n\} \\ 0 & \text{altrimenti} \end{cases}$$

Dato che $a \to f(x_1, \dots, x_n; a)$ è monotone crescente in a ad $x = (x_1 \dots x_n)$ fissato, il valore massimo di tale funzione è assunto in corrispondenza di $\min\{x_1, x_2, \dots, x_n\}$, per cui $\hat{a} = \min\{X_1, X_2, \dots, X_n\}$ è lo stimatore di massima verosimiglianza

- Dal teorema di fattorizzazione di Fisher si vede facilmente che $T = \max\{X_1, X_2, \dots X_n\}$ è una statistica sufficiente.

3). È un intervallo aleatorio $[X_1, X_2]$ (X_1 e X_2 sono v.a.) tale che

$$P[X_1 \leq a \text{ e } X_2 \geq a] \geq 1-\alpha$$

- L'intervallo di confidenza di livello $1-\alpha$ per il parametro p di una distribuzione di Bernoulli è dato da

$$\left[\overline{X} - \sqrt{\frac{\overline{X}(1-\overline{X})}{n}}\, z_{1-\alpha/2}, \overline{X} + \sqrt{\frac{\overline{X}(1-\overline{X})}{n}}\, z_{1-\alpha/2}\right]$$

Nel nostro caso $\alpha = 0.05$, per cui $1-\frac{\alpha}{2} = 0.975$

e $z_{0.975} = 1.96$

$\overline{X} = \frac{525}{574} = 0.9146$ $\qquad n = 574$

$$\sqrt{\frac{\overline{X}(1-\overline{X})}{574}} \cdot z_{1-\alpha/2} = 0{,}023$$

Quindi l'intervallo richiesto è

$[0{,}892, 0{,}938]$

Esame di *Calcolo delle probabilità e statistica* (per studenti di Informatica)
corso A e B
Università degli studi di Bari Aldo Moro

22-09-2021

Esercizio 1. Si dispone di una moneta equa e di un dado equilibrato a sei facce numerate da 1 a 6. Si conduce il seguente esperimento. Si lancia prima la moneta: se esce testa, si lancia $n \geq 1$ volte il dado; se esce croce lo si lancia $2n$ volte.

- Calcolare la probabilità che il 6 esca almeno una volta.
- Calcolare la probabilità che il 6 esca esattamente n volte.
- Calcolare la probabilità che sia uscita testa sapendo che il 6 è uscito n volte.

Esercizio 2. Dato il parametro reale $\theta > 0$, verificare che la funzione $f(x;\theta) := \theta x^{\theta-1}\chi_{[0,1]}(x)$ è la densità di probabilità di una certa variabile aleatoria X.

- Determinare l'attesa di X, verificando che si ha $E[X] = \frac{\theta}{\theta+1}$.
- Sfruttare il calcolo precedente per determinare lo stimatore dei momenti di θ corrispondente a un campione $(X_1, X_2, \ldots, X_n)$ di rango n distribuito come X.
- Determinare lo stimatore di massima verosimiglianza di θ corrispondente a un campione $(X_1, X_2, \ldots, X_n)$.
- Trovare la legge (funzione di ripartizione e densità) della variabile aleatoria $Y := -\ln X$.

Esercizio 3.

- Dare la definizione di errore di prima e seconda specie nell'ambito dei test statistici di verifica delle ipotesi, ed illustrarla attraverso qualche esempio o analogia.
- Per studiare le abitudini di caccia dei pipistrelli, 22 esemplari sono stati muniti di un segnalatore per essere monitorati via radio. Di questi 22 pipistrelli, $n = 12$ erano femmine e $m = 10$ erano maschi. Nell'esperimento sono state misurate le distanze percorse tra un pasto e il successivo, ottenendo i risulati seguenti. Per il gruppo dei pipistrelli femmine la media e la varianza campionaria sono rispettivamente $\overline{X} = 180m$ e $S_x^2 = 92m^2$. Per il gruppo dei pipistrelli maschi la media e la varianza campionaria sono rispettivamente $\overline{Y} = 136m$ e $S_y^2 = 86m^2$.
  Verificare con un livello di significatività del 5% e del 10% che la distanza media percorsa sia la stessa per maschi e femmine (assumendo che la varianza della popolazione maschile coincida con quella della popolazione femminile).

Esercizio 1

$H_1 =$ "La moneta dà testa" $\quad p(H_1) = \frac{1}{2}$

$H_2 = H_1^c =$ "La moneta dà croce" $\quad p(H_2) = \frac{1}{2}$

- $E =$ "Esce 6 almeno una volta"

$p(E) = 1 - p(E^c)$ con $E^c =$ "Il 6 non esce nemmeno una volta"

$p(E^c) = p(E^c|H_1) \cdot p(H_1) + p(E^c|H_2) \cdot p(H_2)$ (probabilità totale)

$p(E^c|H_1) = \left(\frac{5}{6}\right)^n$

$p(E^c|H_2) = \left(\frac{5}{6}\right)^{2n}$

per cui $p(E^c) = \frac{1}{2}\left[\left(\frac{5}{6}\right)^n + \left(\frac{5}{6}\right)^{2n}\right]$

$p(E) = 1 - \frac{1}{2}\left[\left(\frac{5}{6}\right)^n + \left(\frac{5}{6}\right)^{2n}\right]$

- $F =$ "Il 6 esce n volte"

$p(F) = p(F|H_1) \cdot p(H_1) + p(F|H_2) \cdot p(H_2)$

$p(F|H_1) = \left(\frac{1}{6}\right)^n$

$p(F|H_2) = \binom{2n}{n}\left(\frac{1}{6}\right)^n\left(\frac{5}{6}\right)^n$ (dalla binomiale)

$p(F) = \frac{1}{2}\left[\left(\frac{1}{6}\right)^n + \binom{2n}{n}\left(\frac{1}{6}\right)^n\left(\frac{5}{6}\right)^n\right] = \frac{1}{2}\left(\frac{1}{6}\right)^n\left[1 + \binom{2n}{n}\left(\frac{5}{6}\right)^n\right]$

- $p(H_1|F) = \dfrac{p(F|H_1) \cdot p(H_1)}{p(F)} = \dfrac{\cancel{\frac{1}{2}}\cancel{\left(\frac{1}{6}\right)^n}}{\cancel{\frac{1}{2}}\cancel{\left(\frac{1}{6}\right)^n}\left[1 + \binom{2n}{n}\left(\frac{5}{6}\right)^n\right]} =$

$$= \frac{1}{1 + \binom{2n}{n}\left(\frac{5}{6}\right)^n}$$

Esercizio 2

- la funzione $f(x;\theta) = \begin{cases} \theta x^{\theta-1} & 0 \le x \le 1 \\ 0 & \text{altrimenti} \end{cases}$

prende valori maggiori o uguali a 0. Inoltre

$$\int_{\mathbb{R}} f(x;\theta)dx = \int_0^1 \theta x^{\theta-1} dx = \theta \int_0^1 x^{\theta-1} dx = \theta \frac{x^\theta}{\theta}\Bigg|_0^1 = \frac{\theta}{\theta} = 1$$

- $$E[X] = \int_{\mathbb{R}} x f(x;\theta)dx = \int_0^1 x\,\theta x^{\theta-1} dx = \theta \int_0^1 x^\theta dx = \theta \frac{x^{\theta+1}}{\theta+1}\Bigg|_0^1 = \frac{\theta}{\theta+1}$$

- Da $E[X] = \frac{\theta}{\theta+1} \Rightarrow E[X]\cdot\theta + E[X] = \theta \Rightarrow$

$$(E[X]-1)\theta = -E[X] \Rightarrow \theta = \frac{E[X]}{1-E[X]}$$

Lo stimatore dei momenti è allora $\hat{\theta} = \frac{\overline{X}_n}{1-\overline{X}_n}$

dove $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ (media campionaria)

- La densità congiunta è

$$g(x_1, x_2, \dots, x_n; \theta) = \begin{cases} \theta^n (x_1 x_2 \cdots x_n)^{\theta-1} & 0 \le x_i \le 1 \ \forall i \\ 0 & \text{altrimenti} \end{cases}$$

$$\ln g(x_1, x_2, \dots, x_n; \theta) = n \ln\theta + (\theta-1)\ln(x_1 \cdots x_n)$$

$$\frac{\partial}{\partial\theta} \ln g(x_1, x_2, \dots, x_n; \theta) = \frac{n}{\theta} + \ln(x_1 x_2 \cdots x_n) \overset{!}{=} 0$$

$$\frac{n}{\theta} = -\sum_{i=1}^n \ln x_i \quad ; \quad \frac{\theta}{n} = \frac{1}{-\sum_{i=1}^n \ln x_i}$$

$$\theta = \frac{1}{-\frac{1}{n}\sum_i \ln x_i} \quad ; \quad \hat{\theta}_n = \frac{1}{-\frac{1}{n}\sum_{i=1}^n \ln X_i}$$

• $Y = -\ln X$

dato che X prende valori in $[0,1]$, Y

prende valori ~~tra~~ in $[0, \infty)$

Detta $G(t) := P(Y \leq t)$, abbiamo $F(t) = 0 \ \forall t < 0$

Se $t \geq 0$ si ha invece

$$P(Y \leq t) = P(-\ln X \leq t) = P(\ln X \geq -t) =$$
$$P(X \geq e^{-t}) = 1 - P(X < e^{-t})$$
$$= 1 - \int_0^{e^{-t}} \theta x^{\theta - 1} \, dx$$

La densità di Y è $g(t) = G'(t)$
ed è nulla per $t < 0$

Per $t \geq 0$ è data da $g(t) = -\theta (e^{-t})^{\theta - 1} (-e^{-t})$

$$= \theta e^{t} e^{-\theta t - t} = \theta e^{-\theta t}$$

(variabile esponenziale)

Esercizio 3

- L'errore di prima specie consiste nel rifiutare l'ipotesi nulla quando è vera.
  L'errore di seconda specie consiste nell'accettare l'ipotesi nulla quando è falsa.

Nell'analogia "giudiziaria" ~~[illegible]~~, l'errore di prima specie porta alla condanna di un innocente; quello di seconda specie all'assoluzione di un colpevole.

- $\overline{X} = 180$ m $\quad S_x^2 = 92$ m^2 $\quad n = 12$

  $\overline{Y} = 136$ m $\quad S_y^2 = 86$ m^2 $\quad m = 10$

$n+m-2 = 22-2 = 20$

$$S_p^2 = \frac{(n-1)S_x^2 + (m-1)S_y^2}{n+m-2} = \frac{11 \cdot 92 + 9 \cdot 86}{20} \; m^2 = 89.3 \; m^2$$

$S_p = 9.45$

$$T = \frac{\overline{X} - \overline{Y}}{S_p\sqrt{\frac{1}{n}+\frac{1}{m}}} = \frac{180 - 136}{9.45\sqrt{\frac{1}{12}+\frac{1}{10}}} = 10.87$$

~~H0:~~

L'ipotesi nulla è $H_0 : \mu_x = \mu_y$

Si rifiuta H_0 se $|T| > t_{1-\alpha/2, \, n+m-2}$

Nel nostro caso: $\alpha = 5\%$ $\quad t_{0.975, \, 20} = 2.0860$

$\Rightarrow$ si rifiuta H_0

$\alpha = 10\%$ $\quad t_{0.95, \, 20} = 1.724 \Rightarrow$ si rifiuta H_0

Esame di *Calcolo delle probabilità e statistica* (per studenti di Informatica) corso A e B

Università degli studi di Bari Aldo Moro

16-11-2021

Esercizio 1. Si lancia una moneta equa N volte, dove N è una variabile aleatoria geometrica di parametro $0 < p < 1$, ossia $P[N = k] = p(1-p)^{k-1}$ per ogni $k \geq 1$.

- Calcolare la probabilità di ottenere almeno una testa.
- Calcolare la probabilità che $N = 1$ sapendo che non si è ottenuta alcuna testa.
- Calcolare il limite della probabilità determinata al punto precedente quando $p \to 1^-$, interpretando il risultato ottenuto.

(Si ricorda che $\sum_{l=0}^{\infty} q^l = \frac{1}{1-q}$ per ogni q con $|q| < 1$.)

Esercizio 2. Verificare che per ogni valore del parametro $a > 0$ la funzione

$$f(x; a) := \frac{7a^7}{x^8} \chi_{[a,\infty]}(x)$$

è la densità di probabilità di una certa variabile aleatoria X.

- Calcolare l'attesa di X, verificando che si ha $E[X] = \frac{7}{6}a$.
- Sfruttare il calcolo precedente per esibire uno stimatore corretto di a a partire da un campione di rango n distribuito come X.
- Determinare lo stimatore di massima verosimiglianza di a relativo a un campione di rango n distribuito come X.
- Calcolare la varianza di X.

Esercizio 3.

- Fornire la definizioni di livello di significatività α e di *p-value* di un dato test statistico.
- I salmoni cresciuti ogni anno in un allevamento commerciale hanno dei pesi con distribuzione normale di deviazione standard $\sigma = 1,2$ libbre. La ditta dichiara che il peso medio dei suoi pesci quest'anno è superiore alle 7.6 libbre. Un campione casuale di $n = 16$ pesci ha dato una media campionaria pari a 7.2 libbre. Si può dire che questo dato sia abbastanza forte da rigettare l'affermazione dell'azienda al 5% di significatività?
- Calcolare il *p-value* del test statistico utilizzato al punto precedente .

(1) Indichiamo con X la variabile aleatoria che conta il numero di teste ottenute. Abbiamo

- $P[X \geq 1] = 1 - P[X=0]$

e

$$P[X=0] = \sum_{k=1}^{\infty} P[X=0 \mid N=k] \cdot P[N=k] = \sum_{k=1}^{\infty} \left(\frac{1}{2}\right)^k p(1-p)^{k-1}$$

$$= \frac{1}{2} p \sum_{k=1}^{\infty} \left(\frac{1-p}{2}\right)^{k-1} = \frac{p}{2} \sum_{\ell=0}^{\infty} \left(\frac{1-p}{2}\right)^{\ell} = \frac{p}{2} \frac{1}{1-\left(\frac{1-p}{2}\right)} =$$

$$= \frac{p}{2} \frac{2}{2-1+p} = \frac{p}{1+p}$$

da cui

$$P[X \geq 1] = 1 - \frac{p}{1+p} = \frac{1}{1+p}$$

- $$P[N=1 \mid X=0] = \frac{P[X=0 \mid N=1] \cdot P[N=1]}{P[X=0]} = \frac{1}{2} p \cdot \frac{(1+p)}{p}$$

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

- $$\lim_{p \to 1^-} \frac{1+p}{2} = 1$$

Interpretazione: se $p=1$ $N=1$ con certezza e quindi $P[N=1 \mid X=0]$ deve tendere a 1 quando $p \to 1^-$.

2

$$f(x;a) = \begin{cases} \dfrac{7a^7}{x^8} & x \geq a \\ 0 & x < a \end{cases}$$

Si ha $f(x;a) \geq 0 \quad \forall x$ in $\mathbb{R}$ e

$$\int_{\mathbb{R}} f(x;a)\,dx = \int_a^{+\infty} \frac{7a^7}{x^8}\,dx = 7a^7 \int_a^{+\infty} x^{-8}\,dx = 7a^7 \left.\frac{x^{-7}}{-7}\right|_a^{\infty} = \frac{7a^7}{7a^7} = 1$$

- $$E[X] = \int_{\mathbb{R}} x f(x;a)\,dx = \int_a^{+\infty} \frac{7a^7}{x^8}\,x\,dx = 7a^7 \int_a^{+\infty} x^{-7}\,dx = 7a^7 \left.\frac{x^{-6}}{-6}\right|_a^{\infty} = \frac{7}{6}a$$

- Da $E[X] = \frac{7}{6}a$ si trova $a = E\left[\frac{6}{7}X\right]$, per cui

$$\frac{6}{7}\,\frac{X_1 + \dots + X_n}{n}$$ è uno stimatore corretto

- La densità congiunta del campione è

$$f(x_1, x_2, \dots, x_n; a) = \begin{cases} \dfrac{7^n a^{7n}}{x_1^8 x_2^8 \cdots x_n^8} & x_i \geq a \quad \forall i = 1, 2, \dots, n \\ 0 & \text{altrimenti} \end{cases}$$

La condizione $x_i \geq a \quad \forall i = 1, \dots, n$ implica

$a \leq x_i \quad \forall i = 1, \dots, n$, cioè $a \leq \min\limits_{1 \leq i \leq n} \{x_i\}$

La densità congiunta è una funzione monotona crescente in a. Segue che il suo massimo è raggiunto in corrispondenza del valore $\min\limits_{1 \leq i \leq n} \{x_i\}$.

Lo stimatore di massima verosimiglianza è allora $\hat{\Theta}_n = \min\limits_{1 \leq i \leq n} \{X_i\}$

- $Var[X] = E[X^2] - E[X]^2$

$$E[X^2] = \int_a^{+\infty} x^2 \frac{7a^7}{x^8}\,dx = 7a^7 \int_a^{+\infty} x^{-6}\,dx = 7a^7 \left.\frac{x^{-5}}{-5}\right|_a^{\infty} = \frac{7}{5}a^2$$

$$Var[X] = \frac{7}{5}a^2 - \left(\frac{7}{6}a\right)^2 = \frac{7}{5}a^2 - \frac{49}{36}a^2$$

③ Il livello di significatività α è la probabilità di commettere errore di prima specie (rifiutare l'ipotesi nulla quando è vera).

Il p-value è il più piccolo valore di α al di sopra del quale i dati portano a rifiutare l'ipotesi nulla.

$\sigma = 1.2$ (libbre) è la deviazione standard della popolazione

$H_0 : \mu \geq \mu_0 = 7.6$ (libbre)

$H_1 : \mu < \mu_0 = 7.6$ (libbre)

$n = 16 \quad \overline{X} = 7.2$

$\alpha = 5\% = 0.05$

Occorre eseguire un test di Gauss unilaterale.

La statistica da usare è $T = \dfrac{\overline{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$

Si rifiuta H_0 se $\dfrac{\overline{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} < -Z_{1-\alpha}$

Ora $\dfrac{\overline{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \dfrac{7.2 - 7.6}{\frac{1.2}{4}} = -1.33$

mentre $Z_{1-\alpha} = Z_{0.95} = 1.65$

Quindi non si può rifiutare H_0

- Il p-value è dato dalla relazione $Z_{1-p} = 1.33$

$\Rightarrow 1 - p = Erf(1.33)$; $p = 1 - Erf(1.33) = 1 - 0.90824 \simeq 9\%$.