

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

Università degli studi di Bari Aldo Moro

Docente: Stefano Rossi

04-02-2022

Esercizio 1. Un dado equilibrato a sei facce numerate da 1 a 6 viene lanciato N volte, dove N è una variabile di Poisson di parametro $\lambda > 0$.

- (1) Calcolare la probabilità che il numero 6 esca almeno una volta, verificando che vale $1 - e^{-\frac{\lambda}{6}}$.
- (2) Calcolare la probabilità che $N = 1$ sapendo che il numero 6 non è mai uscito.
- (3) Calcolare il limite per $\lambda \rightarrow \infty$ della probabilità al punto precedente e fornire un'interpretazione probabilistica del risultato trovato.

(Ricordare che $\sum_{k=0}^{\infty} \frac{a^k}{k!} = e^a$, per ogni numero reale a .)

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

$$f(x) = e^{-(x-a)} \chi_{[a, \infty)}(x), \quad x \in \mathbb{R}$$

dove $\chi_{[a, \infty)}$ è la funzione caratteristica della semiretta $[a, \infty)$, è la densità di probabilità di una certa variabile aleatoria X .

- (1) Calcolare l'aspettazione della variabile aleatoria X , verificando che $E[X] = a + 1$.
- (2) Sfruttare il valore determinato sopra per ricavare uno stimatore corretto del parametro a in corrispondenza a un campione (X_1, X_2, \dots, X_n) di rango n distribuito come X .
- (3) Determinare lo stimatore di massima verosimiglianza di a in corrispondenza a un campione (X_1, X_2, \dots, X_n) di rango n distribuito come X .
- (4) Determinare una statistica sufficiente per un campione (X_1, X_2, \dots, X_n) di rango n distribuito come X .

Esercizio 3.

- Dare le definizioni di errore di prima e seconda specie nell'ambito dei test statistici di verifica delle ipotesi.
- Un'azienda vinicola sostiene che il contenuto medio di vino nelle sue bottiglie sia superiore a 750cl. Un'associazione di consumatori decide di verificare quest'affermazione, analizzando un campione casuale di $n = 40$ bottiglie. Si trova una media campionaria $\bar{x} = 748$ cl con una varianza campionaria $\sigma_x^2 = 65$ cl². Quali conclusioni si possono trarre con un livello di significatività del 5% e del 10%?
- Calcolare il p -value (approssimato) del test al punto precedente.

Esercizio n° 1

$E = \text{"Il numero 6 esce almeno una volta"}$

$E^C = \text{"Il numero 6 non esce mai"}$

$$P(E) = 1 - P(E^C)$$

$$\begin{aligned} P(E^C) &= \sum_{k=0}^{\infty} P(E^C | N=k) \cdot P(N=k) = \sum_{k=0}^{\infty} \left(\frac{5}{6}\right)^k e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \left(\frac{5}{6}\right)^k \frac{\lambda^k}{k!} = \\ &= e^{-\lambda} \cdot e^{5/6 \lambda} = e^{-\lambda/6} \end{aligned}$$

$$\begin{aligned} P(N=1 | E^C) &= \frac{P(E^C | N=1) \cdot P(N=1)}{P(E^C)} = \frac{\left(\frac{5}{6}\right) e^{-\lambda} \lambda}{e^{-\lambda/6}} = \frac{5}{6} \lambda e^{-\lambda/6} = \\ &= \frac{5}{6} \lambda e^{-5/6 \lambda} \end{aligned}$$

$$\lim_{\lambda \rightarrow \infty} \frac{5}{6} \lambda e^{-5/6 \lambda} = 0$$

Interpretazione: $E[N] = \lambda$, per cui quando $\lambda \rightarrow \infty$

$P(N=1)$ tende a zero anche se però che

che non si è mai ottenuto elen 6.

Esercizio 2

$$f(x) = \begin{cases} e^{-(x-a)} & x \geq a \\ 0 & x < a \end{cases}$$

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

$$- e^{(x-a)} \Big|_a^{+\infty} = 0 - (-1) = 1$$

$$- E[X] = \int_{\mathbb{R}} x f(x) dx = \int_{-\infty}^{+\infty} x e^{-(x-a)} dx = -e^{-(x-a)} \Big|_{-\infty}^{\infty} - \int_{-\infty}^{+\infty} e^{-(x-a)} dx$$

$$a + \int_{-\infty}^{+\infty} e^{-(x-a)} dx = a + 1$$

$$- a = E[X] - 1, \text{ per cui } \bar{X} - 1 = \frac{x_1 + \dots + x_n}{n} - 1$$

perché

è uno stimatore corretto
del parametro a .

- La densità condizionata del campione è la funzione

$$h(x_1, x_2, \dots, x_n; a) = e^{-(x_1-a)} e^{-(x_2-a)} \dots e^{-(x_n-a)} \\ = e^{-\sum_{i=1}^n x_i / n} \quad \text{se } x_i \geq a \quad \forall a$$

altrimenti vale 0

La condizione $x_i \geq a \quad \forall i$ si rilega in termine del parametro

ignoto come $a \leq x_i \quad \forall i$, cioè $a \leq \min_{1 \leq i \leq n} \{x_i\}$

$a \rightarrow h(x_1, x_2, \dots, x_n, a)$ è una f.m. monotone crescente in a
per cui per lui il suo massimo in corrispondenza del valore più grande
possibile per a , che è $\min_{1 \leq i \leq n} \{x_i\}$.

- stimatore di massima verosimiglianza è allora

$$\hat{a} = \min_{1 \leq i \leq n} \{x_i\}$$

- ~~Nella statistica deterministica la statistica sufficiente
utilizza tutte le informazioni di fatto~~ ~~di Fisher~~, perché

con $\alpha=0.1$ si ha $t_{0.90, 39} = 1.304$ e di conseguenza
si rifiuta H_0 (perché $-1.569 < -1.304$)

Poiché il numero di gradi di libertà è abbastanza alto, si possono approssimare i quantili di Student con quelli di Gauss - Le condizione che determina il p-value è allora

$$Z_{x-p} = 1.57$$

$$p = 1 - \text{Erf}(1.57) = 1 - 0.9418 \sim 5.8\%.$$

Applicando il teorema di sufficienza di Fisher

si conclude che $\min_{1 \leq i \leq n} \{X_i\}$ è una statistica sufficiente, dato che

$$h(x_1, x_2, \dots, x_n) = u(x_1, \dots, x_n) v(t(x), \alpha)$$

$$\text{con } u(x_1, \dots, x_n) = \bar{x} \sum_{i=1}^n x_i$$

$$v(t(x); \alpha) = \chi_{[t(x), +\infty)} \quad \text{con } t(x_1, \dots, x_n) = \min_{1 \leq i \leq n} \{x_i\}$$

3)

Errore di prima specie = rifiutare l'ipotesi nulla, quando invece è vera

Errore di seconda specie = accettare l'ipotesi nulla, quando è falsa

- Occorre condurre un test di Student unilaterale

La statistica da usare è la V.d.

$$T = \frac{\bar{X} - \mu_0}{\frac{s}{\sqrt{n}}}$$

$$\text{con } \mu_0 = 750 \text{ c.l.} \quad n = 40 \quad s = \sqrt{s_x^2} = \sqrt{65} \text{ c.l.} = 8.0623 \text{ c.l.}$$

$$\text{e } \bar{X} = 748 \text{ c.l.}$$

$$\text{Si ha } T = \frac{748 - 750}{\frac{8.0623}{\sqrt{40}}} = -1.569$$

Si rifiuta $H_0: \mu > \mu_0$ se

$$T < -t_{1-\alpha, n-1}$$

$$\text{con } \alpha = 0.05 \quad t_{0.95, 34} = 1.685 \Rightarrow \text{si accetta } H_0$$

Esame di *Calcolo delle probabilità e statistica* (per studenti di Informatica)
corso A e B
Università degli studi di Bari Aldo Moro
13-04-2022

Esercizio 1. Una scatola contiene 100 monete, di cui 50 sono truccate e danno testa con probabilità $\frac{2}{3}$, mentre le restanti 50 sono monete equilibrate. Si prende a caso una moneta e la si lancia.

- Calcolare la probabilità che dia testa.
- Calcolare la probabilità che la moneta sia truccata sapendo che è uscita croce.
- Supponendo che la moneta presa a caso venga lanciata due volte, calcolare la probabilità che al secondo lancio dia testa sapendo che al primo lancio ha dato testa. I due lanci sono indipendenti?

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

$$f(x) = \alpha 2^\alpha x^{-(\alpha+1)} \chi_{[2,\infty)}(x)$$

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

- Determinare i valori di α per i quali X ha aspettazione finita e determinare il valore dell'aspettazione in corrispondenza di tali valori.
- Determinare lo stimatore di massima verosimiglianza di α relativo a un campione di rango n distribuito come X .
- Sapendo che $E[\ln(X)] = \ln 2 + \frac{1}{\alpha}$, dire se lo stimatore ottenuto al punto precedente è consistente.
- Esibire una statistica sufficiente per un campione di rango n distribuito come X .

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.
- Costruire un intervallo di confidenza (approssimato) al livello $1 - \alpha$ per il parametro p di una distribuzione di Bernoulli.
- Su un campione di 2500 individui 725 sono disoccupati. Costruire l'intervallo di confidenza della proporzione di disoccupati nella popolazione al livello di confidenza $1 - \alpha = 0.99$.

Esercizio n° 1

H_1 = "La moneta è truccata"

$$p(H_1) = \frac{1}{2}$$

H_2 = "La moneta è equa"

$$p(H_2) = \frac{1}{2}$$

T = "Esce testa"

C = "Esce croce"

$$p(T|H_1) = \frac{2}{3} \quad - \quad p(C|H_1) = \frac{1}{3}$$

$$p(T|H_2) = p(C|H_2) = \frac{1}{2}$$

$$P(T) = P(T|H_1) \cdot p(H_1) + P(T|H_2) \cdot p(H_2)$$

$$= \frac{1}{2} \left[\frac{2}{3} + \frac{1}{2} \right] = \frac{1}{2} \cdot \frac{4+3}{6} = \frac{7}{12} \approx 0.583$$

formula
delle probabilità
totali

$$p(H_1|C) = \frac{p(C|H_1) \cdot p(H_1)}{p(C|H_1) \cdot p(H_1) + p(C|H_2) \cdot p(H_2)} = \text{(formula di Bayes)}$$

$$= \frac{\frac{1}{3}}{\frac{1}{3} + \frac{1}{2}} = \frac{\frac{1}{3}}{\frac{5}{6}} = \frac{\frac{1}{3}}{\frac{5}{6}} = \frac{1}{5} \cdot \frac{6}{5} = \frac{3}{5} = 0.4$$

E_1 = "Al primo lancio esce teste"

E_2 = "Al secondo lancio esce teste"

$$P(E_2|E_1) = \frac{P(E_1 \cap E_2)}{P(E_1)} \quad \text{(definizione di probabilità condizionata)}$$

$$\text{Ora } P(E_1) = P(T)$$

$$P(E_1 \cap E_2) = P(E_1 \cap E_2 | H_1) P(H_1) + P(E_1 \cap E_2 | H_2) P(H_2)$$

$$= \frac{1}{2} \left[\frac{2}{3} \cdot \frac{2}{3} + \frac{1}{2} \cdot \frac{1}{2} \right] = \frac{1}{2} \left[\frac{4}{9} + \frac{1}{4} \right]$$

$$= \frac{1}{2} \cdot \frac{16+9}{36} = \frac{1}{2} \cdot \frac{25}{36} =$$

$$P(E_2 | E_1) = \frac{\frac{1}{2} \cdot \frac{25}{36}}{\frac{7}{12}} = \frac{1}{2} \cdot \frac{25}{36} \cdot \frac{12}{7} = \frac{25}{142} = 0.5952$$

mentre

$$P(E_2) = P(T) = 0.583$$

$E_1 \text{ e } E_2$ non sono indipendenti.

Esercizio 2

$$f(x) = \begin{cases} 0 & x < 2 \\ \lambda 2^{\lambda} x^{-(\lambda+1)} & x \geq 2 \end{cases}$$

$f(x) \geq 0 \quad \forall x \in \mathbb{R}$ e inoltre

$$\int_{-\infty}^{\infty} f(x) dx = \lambda 2^{\lambda} \int_2^{+\infty} x^{-(\lambda+1)} dx =$$

$$\lambda 2^{\lambda} \frac{x^{-\lambda}}{-\lambda} \Big|_2^{+\infty} = \lambda 2^{\lambda} \left[0 - \frac{2^{-\lambda}}{-\lambda} \right] = \lambda 2^{\lambda} \frac{2^{-\lambda}}{\lambda} = 1$$

$$E[X] = \lambda 2^{\lambda} \int_2^{+\infty} x^{-(\lambda+1)} \cdot x dx = \lambda 2^{\lambda} \int_2^{+\infty} x^{-\lambda} dx$$

l'integrale è convergente per $\lambda > 1$ e infatti così vale

$$\lambda 2^{\lambda} \frac{x^{-(\lambda+1)}}{-\lambda+1} \Big|_2^{+\infty} = \lambda 2^{\lambda} \left[0 - \frac{2^{-\lambda+1}}{1-\lambda} \right] = \lambda 2^{\lambda} \frac{2^{-\lambda+1}}{\lambda-1}$$

• La densità congiunta è

$$f(x_1, x_2, \dots, x_n; \lambda) = \begin{cases} 2^n 2^{2n} (x_1 - x_n)^{-(\lambda+1)} & \text{per } x_i \geq 2 \\ 0 & \text{se } \exists i \text{ con } x_i < 2 \end{cases}$$

$$\log f(x_1, x_2, \dots, x_n; \lambda) = n \log 2 + 2n \log 2 - (\lambda+1) \log(x_1 - x_n)$$

$$\frac{\partial}{\partial \lambda} \log f(x_1, x_2, \dots, x_n; \lambda) = \frac{n}{2} + n \log 2 - \log(x_1 - x_n) \stackrel{!}{=} 0$$

$$\frac{n}{2} = \log(x_1 - x_n) - n \log 2$$

$$\frac{1}{2} = \frac{1}{n} \sum_{i=1}^n \log(x_i) - \log 2$$

$$\lambda = \frac{1}{\frac{1}{n} \sum_{i=1}^n \log(x_i) - \log 2}$$

$$\lambda_n = \frac{1}{\frac{1}{n} \sum_{i=1}^n \log(X_i) - \log 2}$$

Per la legge dei grandi numeri (debole)

$$\frac{1}{n} \sum_{i=1}^n \log X_i \xrightarrow{P} E[\log X] = \log 2 + \frac{1}{2}$$

Ne segue che λ_n converge in probabilità a

$$\frac{1}{\log 2 + \frac{1}{2} - \log 2} = \lambda$$

$$\text{ci\`e} \quad P[\lvert \hat{\lambda}_n - \lambda \rvert \geq S] \xrightarrow[n \rightarrow \infty]{} 0$$

dunque $\hat{\lambda}_n$ i consist..

Dal teorema di fattorizzazione di Fisher si vede che

$T = X_1 X_2 \dots X_n$ i una statistica sufficiente.

Esercizio 3

E' un intervallo aleatorio $I = [X_1, X_2]$ tale che

$$P_{\theta} [X_1 < \psi(\theta) < X_2] \geq 1-\alpha$$

per ogni $\theta \in \Theta$ se $\theta \mapsto \psi(\theta)$ i la f.m che si intende stimare

L'intervallo richiesto i dato dalla formula

$$[\bar{X} - \frac{Z_{1-\alpha/2}}{\sqrt{n}} \sqrt{\bar{X}(1-\bar{X})}, \bar{X} + \frac{Z_{1-\alpha/2}}{\sqrt{n}} \sqrt{\bar{X}(1-\bar{X})}]$$

$$\text{con } \bar{X} = \frac{1}{n} \sum_i x_i \quad (\text{medio campionario})$$

e $Z_{1-\alpha/2}$ quantile gaussiano

$$\bar{X} = \frac{725}{2500} = 0.29 \quad \alpha = 0.01 \quad 1 - \frac{\alpha}{2} = 0.995$$

$$\frac{1}{\sqrt{n}} Z_{0.995} \sqrt{\bar{X}(1-\bar{X})} = \frac{1}{50} 2.50 \sqrt{0.29(0.71)} = 0.023$$

L'intervallo i allora

$$[0.267; 0.313]$$

Esame di *Calcolo delle probabilità e statistica* (per studenti di Informatica)
corso A e B
Università degli studi di Bari Aldo Moro
Docente: Stefano Rossi
19-05-2022

Esercizio 1. Una moneta equa viene lanciata N volte, dove N è una variabile di Poisson di parametro $\lambda > 0$. Detta X la variabile aleatoria che conta il numero di teste ottenute,

- (1) calcolare $P[X \geq 1]$;
- (2) calcolare $P[N = n|X = 0]$;
- (3) calcolare il limite della probabilità ottenuta al punto precedente quando n tende a infinito.

Esercizio 2. Per ogni valore del parametro reale θ , si considera la funzione

$$f(x; \theta) := \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\theta)^2}{2}}, \quad x \in \mathbb{R}.$$

Dopo aver specificato di quale densità di probabilità notevole si tratta,

- (1) determinare lo stimatore di massima verosimiglianza di θ relativo a un campione di rango n ;
- (2) calcolare la varianza dello stimatore ottenuto al punto precedente;
- (3) esibire una statistica sufficiente per θ ;
- (4) calcolare l'informazione di Fisher di un campione di rango n ;
- (5) dire se lo stimatore di massima verosimiglianza ottenuto sopra è efficiente;

Esercizio 3. Un oleificio imbottiglia olio extravergine di oliva in due stabilimenti diversi. Si vuole controllare se il contenuto medio di una bottiglia di olio sia lo stesso nei due stabilimenti. A tal scopo si prendono a caso $n = 30$ bottiglie per il primo stabilimento e $m = 40$ per il secondo, e si trovano le medie campionarie $\bar{x}_1 = 998$ cl e $\bar{x}_2 = 1001$ cl, rispettivamente. Assumendo che i contenuti siano variabili aleatorie gaussiane indipendenti con la stessa varianza nota $\sigma^2 = 40$ cl²:

- (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

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

$$P[X=0] = \sum_{n=0}^{\infty} P[X=0 | N=n] \cdot P[N=n]$$

(formule
delle probabilità
totali)

$$= \sum_{n=0}^{\infty} \left(\frac{1}{2}\right)^n e^{-\lambda} \frac{\lambda^n}{n!} =$$

$$e^{-\lambda} \sum_{n=0}^{\infty} \left(\frac{\lambda}{2}\right)^n \frac{1}{n!} = e^{-\lambda} \cdot e^{\lambda/2} = e^{-\lambda/2}$$

$$P[X \geq 1] = 1 - e^{-\lambda/2}$$

$$\cdot P[N=n | X=0] = \frac{P[X=0 | N=n] \cdot P[N=n]}{P[X=0]}$$

(formula
di Bayes)

$$= \left(\frac{1}{2}\right)^n e^{-\lambda} \frac{\lambda^n}{n!} \cdot \frac{1}{e^{-\lambda/2}}$$

$$= e^{-\lambda/2} \left(\frac{1}{2}\right)^n \frac{\lambda^n}{n!}$$

$$\lim_{n \rightarrow \infty} P[N=n | X=0] = 0 \quad \text{perché} \quad \left(\frac{1}{2}\right)^n \xrightarrow[n \rightarrow \infty]{\rightarrow 0} e$$

$$\frac{\lambda^n}{n!} \xrightarrow[n \rightarrow \infty]{\rightarrow 0}$$

Esercizio 2

La funzione $f(x; \theta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\theta)^2}{2}}$ è la densità normale di media θ e varianza 1.

$$1. h(x_1, x_2, \dots, x_n; \theta) = \frac{1}{(2\pi)^{n/2}} e^{-\sum_i \frac{(x_i - \theta)^2}{2}} \quad \text{densità congruente}$$

$$\ln h(x_1, x_2, \dots, x_n; \theta) = -\sum_i \frac{(x_i - \theta)^2}{2} \quad (\text{a meno di costanti additive})$$

$$\frac{\partial}{\partial \theta} \ln h(x_1, x_2, \dots, x_n; \theta) = \sum_i (x_i - \theta) \stackrel{!}{=} 0 \Rightarrow \sum_i x_i - n\theta = 0 \Rightarrow$$

$$\theta = \frac{1}{n} \sum_i x_i \quad \text{Lo stimatore di massima verosimiglianza } \hat{\theta}$$

$$\text{dunque } \hat{\theta}_n = \frac{1}{n} \sum_i x_i$$

$$2. \text{Var}[\hat{\theta}_n] = \frac{1}{n} \text{Var}[X] = \frac{1}{n} \quad \text{(perché } (X \sim N(\theta, 1)))$$

$$3. h(x_1, x_2, \dots, x_n; \theta) = \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2} \left(\sum_i x_i^2 + \theta^2 - 2\theta x_i \right)} \\ = \underbrace{\frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2} \sum_i x_i^2}}_{u(x)} \underbrace{e^{-\frac{1}{2} \sum_i \theta^2 - 2\theta x_i}}_{T(t(x); \theta)}$$

$$\text{con } t(x) = \sum_i x_i$$

$t(X) = \sum_i X_i$ è una statistica sufficiente grazie al teorema di Fisher

$$4. \frac{I}{n}(\theta) = \text{Var} \left[\sum_i (X_i - \theta) \right] = \sum_i \text{Var}[X_i - \theta] = n \text{Var}[X] = n$$

$$5. \text{Il limite di Cramér-Rao è } \frac{1}{I_n(\theta)} = \frac{1}{n}, \text{ per cui} \\ \hat{\theta}_n \text{ è efficiente}$$

Esercizio 3

$$n=30 \quad m=40$$

$$H_0: \mu_1 = \mu_2$$

$$\bar{X}_1 = 998 \text{ cl}$$

$$\bar{X}_2 = 1001 \text{ cl}$$

$$\sigma^2 = 40 \text{ cl}^2 \quad \Rightarrow \quad \sigma = 6.325 \text{ cl}$$

1) occorre eseguire un test di Gauss bilaterale.

La regione di rifiuto per H_0 è

$$T = \left| \frac{\bar{X}_1 - \bar{X}_2}{\sigma \sqrt{\frac{1}{n} + \frac{1}{m}}} \right| \geq Z_{1-\frac{\alpha}{2}}$$

$(Z_{1-\frac{\alpha}{2}} \text{ è il quantile superiore})$

$$2) \text{ Calcoliamo } T = \frac{|998 - 1001|}{6.325 \sqrt{\frac{1}{30} + \frac{1}{40}}} = 1.9638$$

$$\text{con } \alpha = 10\% (= 0.1) \quad \frac{\alpha}{2} = 0.05 \quad 1 - \frac{\alpha}{2} = 0.95 \quad \text{e } Z_{0.95} = 1.65$$

il quantile dunque si rifiuta H_0

$$\text{con } \alpha = 5\% (0.05) \quad Z_{\frac{\alpha}{2}} = 0.025 \quad 1 - Z_{\frac{\alpha}{2}} = 0.975 \quad \text{e } Z_{0.975} = 1.96$$

e dunque si rifiuta H_0

3) Il p-value è determinato dalla relazione

$$T = Z_{1-\frac{\alpha}{2}} \Rightarrow 1 - \frac{\alpha}{2} = \text{Erf}(T) \Rightarrow p = 2(1 - \text{Erf}(T))$$

e sarà poco meno del 5%.

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

Università degli studi di Bari Aldo Moro
Docente: Stefano Rossi, Simone Del Vecchio

08-06-2022

Esercizio 1. Una scatola contiene 10 monete, di cui una è truccata e fa uscire testa con probabilità $\frac{2}{3}$. Si prende una moneta a caso e la si lancia 200 volte. Detta X la variabile aleatoria che conta il numero di teste ottenute, calcolare:

- (1) $E[X]$
- (2) $P[X \geq 100]$
- (3) la probabilità che la moneta sia truccata sapendo che $X \geq 125$.

Esercizio 2. Per ogni valore del parametro $\lambda > 0$, si considera la funzione

$$f(x) := \frac{\lambda e^{-\lambda\sqrt{x}}}{2\sqrt{x}} \chi_{[0,+\infty)}(x).$$

- (1) Verificare che f è la densità di una certa variabile aleatoria X .
- (2) Determinare lo stimatore di massima verosimiglianza di λ relativo a un campione di rango n distribuito come X .
- (3) Esibire una statistica sufficiente per λ .
- (4) Determinare la legge di \sqrt{X} e dire se si tratta di una legge notevole.
- (5) Calcolare $E[\sqrt{X}]$.

Esercizio 3. Dare la definizione di livello di significatività di un test statistico.

Sulla base dei seguenti $n = 10$ valori ottenuti su un campione casuale proveniente da una popolazione normale

$$1.1, 3.1, 4.2, 4.6, 5.0, 5.2, 5.3, 6.5, 8.4, 9.6$$

verificare l'ipotesi che la media della popolazione sia pari a $\mu_0 = 4$ al livello di significatività 5% e 1%.

Esercizio 1

H_1 = "La moneta è rucaste"

$$P(H_1) = \frac{1}{10}$$

$H_2 = H_1^C$ = "La moneta è equa"

$$P(H_2) = \frac{9}{10}$$

E = "Esca testa con un singolo lancio"

$$P(E|H_1) = \frac{2}{3}$$

$$P(E|H_2) = \frac{1}{2}$$

$n=200$ (numero di lanci) - Per $0 \leq k \leq 200$ abbiamo

1.

$$P[X=k] = P[X=k|H_1] \cdot P[H_1] + P[X=k|H_2] \cdot P[H_2]$$

$$= \frac{1}{10} \binom{200}{k} \left(\frac{2}{3}\right)^k \left(\frac{1}{3}\right)^{200-k} + \frac{9}{10} \binom{200}{k} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{200-k}$$

Segue che

$$E[X] = \frac{1}{10} [200 \cdot \frac{2}{3}] + \frac{9}{10} \frac{200}{2} = 109.33$$

$$2. P[X \geq 100] = P[X \geq 100 | H_1] \cdot P[H_1] + P[X \geq 100 | H_2] \cdot P[H_2]$$

$$P[X \geq 100 | H_2] = \frac{1}{2} \quad (\text{per simmetria, perché la media di } X \text{ sotto } H_2 \text{ è } 100)$$

$$P[X \geq 100 | H_1] = P[X \geq 100] \quad \text{con } X \sim N(133, \sqrt{\frac{1}{3} \cdot \frac{2}{3} \cdot 200})$$

$$= P\left[\frac{X - 133}{6.66} \geq \frac{100 - 133}{6.66} = -4.85\right] = 1 = N(133, 0.66)$$

$$P[X \geq 100] = 1 - 1 + \frac{9}{10} \cdot 1 = 11 - 0.55$$

$$P[H_1 | X \geq 125] =$$

$$\frac{P[X \geq 125 | H_1] \cdot p[H_1]}{P[X \geq 125 | H_1] \cdot p[H_1] + P[X \geq 125 | H_2] \cdot p[H_2]}$$

formule
di Bayes

$$P[X \geq 125 | H_1] \cdot p[H_1] + P[X \geq 125 | H_2] \cdot p[H_2]$$

$$P[X \geq 125 | H_1] = P\left[\frac{X - 133}{6.67} \geq \frac{125 - 133}{6.67}\right] =$$

$$P[N \geq -1.20]$$

$$P[N \leq 1.20] = 0.8849$$

$$P[X \geq 125 | H_2] = P\left[\frac{X - 100}{\sqrt{200 \cdot \frac{1}{2} \cdot \frac{1}{2}}} \geq \frac{125 - 100}{\sqrt{\frac{200}{4}}} = \frac{+25}{\sqrt{50}} = 3.53\right] \sim 0$$

Nel seguito che

$$P[H_1 | X \geq 125] \sim 1 \quad \square$$

Esercizio n° 2

$$1) f(x) = \begin{cases} 0 & \text{se } x < 0 \\ \frac{\lambda e^{-\lambda \sqrt{x}}}{2\sqrt{x}} & x \geq 0 \end{cases}$$

è una densità, perché $f(x) \geq 0 \quad \forall x \in \mathbb{R}$ e inoltre

$$\int_{-\infty}^{\infty} f(x) dx = \int_0^{\infty} \frac{\lambda e^{-\lambda \sqrt{x}}}{2\sqrt{x}} = \int_0^{\infty} \lambda e^{-\lambda t} dt = 1 \quad \text{se } t = \sqrt{x}$$

$(dt = \frac{1}{2\sqrt{x}} dx)$

2) La densità condizionata è

$$h(x_1, x_2, \dots, x_n; \lambda) = \begin{cases} \lambda^n e^{-\lambda \sum_{i=1}^n \sqrt{x_i}} & x_i \geq 0 \quad \forall i = 1, 2, \dots \\ 0 & \text{altrimenti} \end{cases}$$

$$\ln h(x_1, x_2, \dots, x_n; \lambda) = n \ln \lambda - \lambda \sum_{i=1}^n \sqrt{x_i} \quad (\text{e meno di costanti additivi})$$

$$\frac{\partial}{\partial \lambda} \ln h(x_1, x_2, \dots, x_n; \lambda) = \frac{n}{\lambda} - \sum_{i=1}^n \frac{1}{\sqrt{x_i}} \geq 0 \quad (\lambda)$$

$$\frac{n}{\lambda} = \sum_{i=1}^n \frac{1}{\sqrt{x_i}} \quad \frac{\lambda}{n} = \frac{1}{\sum_{i=1}^n \sqrt{x_i}} \quad \lambda = \frac{1}{\frac{1}{n} \sum_{i=1}^n \sqrt{x_i}}$$

$$\hat{\lambda} = \frac{1}{\frac{1}{n} \sum_{i=1}^n \sqrt{x_i}} \quad \hat{\lambda} \text{ la stima di massima verosimiglianza}$$

3) Dal teorema di fattorizzazione di Fisher si vuole che

$$\bar{T} = \sum_{i=1}^n \sqrt{x_i} \quad \bar{T} \text{ una statistica sufficiente}$$

$$4) F(t) = P[\sqrt{x} \leq t] = P[X \leq t^2] = \int_0^{t^2} \frac{\lambda e^{-\lambda \sqrt{x}}}{2\sqrt{x}} dx$$

quindi $F(t) = \begin{cases} \frac{\lambda e^{-\lambda t}}{2t} & 2t \neq 0 \\ 0 & t < 0 \end{cases}$ per $t \geq 0$

la densità di \sqrt{x} è di tipo esponenziale di parametro λ .

5) $E[\sqrt{x}] = \frac{1}{\lambda}$ perché $\sqrt{x} \sim \text{exp}(\lambda)$

Esercizio n° 3

Il livello di significatività di un test è la probabilità
di commettere l'errore di prima specie, cioè di rifiutare l'ipotesi
nulle quando è vera.

$$H_0 : \mu_0 = 4$$

Con i dati a disposizione

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = 5.3$$

$$\bar{x}^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 = 33.472$$

$$S_n^2 = \frac{n}{n-1} (\bar{x}^2 - (\bar{x})^2) = \frac{10}{9} (33.472 - 5.3^2) = 5.98$$

Occorre utilizzare un test di Student (bilaterale)

Si rifiuta H_0 se

$$\left| \frac{\bar{x} - \mu_0}{S_n / \sqrt{n}} \right| > t_{1-\alpha/2}(n-1)$$

Nel nostro caso

$$T = \frac{\bar{X} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{5.3 - 4}{\sqrt{5.98}} \cdot 3 = 1.595$$

mentre

ora con $\alpha = 5\% (0.05)$

$$1 - \frac{\alpha}{2} = 1 - 0.025 = 0.975$$

$$\text{e } t_{0.975(9)} = 2.2628$$

ipotesi, e quindi si accetta H_0

Se si accetta H_0 con $\alpha = 5\%$, a maggior ragione lo si accetta anche con $\alpha = 1\%$.

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

Università degli studi di Bari Aldo Moro
Docente: Stefano Rossi, Simone Del Vecchio

24-06-2022

Esercizio 1. Si lancia una moneta equa N volte, dove N è una variabile geometrica di parametro $p = \frac{1}{2}$. Sia X la variabile aleatoria che conta il numero di teste ottenute

- (1) Calcolare $P[X \geq 1]$.
- (2) Per ogni naturale $n \geq 1$, calcolare $P[N = n | X = 0]$.
- (3) Calcolare $P[X = 1]$.

Esercizio 2. Per ogni valore del parametro $\lambda > 0$ si considera la funzione

$$f(x) := 2\lambda x e^{-\lambda x^2} \chi_{[0, \infty)}(x), x \in \mathbb{R}.$$

- (1) Verificare che f è la densità di una certa variabile aleatoria X .
- (2) Determinare lo stimatore di massima verosimiglianza di λ relativo a un campione (X_1, X_2, \dots, X_n) di rango n distribuito come X .
- (3) Esibire una statistica sufficiente per il parametro λ .
- (4) Determinare la legge di X^2 e dire se si tratta di una legge notevole.
- (5) Dire se $\frac{X_1^2 + X_2^2 + \dots + X_n^2}{n}$ è uno stimatore corretto di $\frac{1}{\lambda}$.

Esercizio 3. Dare la definizione di distribuzione di Fisher di parametri (n, m) e richiamare le proprietà dei suoi quantili.

Si devono confrontare le varianze di due popolazioni normali indipendenti. Un campione di rango $n_1 = 9$ estratto dalla prima fornisce una varianza campionaria pari a $S_1^2 = 1.5278$, mentre un campione di rango $n_2 = 10$ estratto dalla seconda fornisce una varianza campionaria pari a $S_2^2 = 1.6556$. Verificare l'ipotesi che le due varianze siano uguali con un livello di significatività del 5% e del 10%.

Esercizio 1

$$1) P[N=n] = p(1-p)^{n-1} \quad n \geq 1$$

$$= \frac{1}{2} \cdot \left(\frac{1}{2}\right)^{n-1} = \left(\frac{1}{2}\right)^n \quad n \geq 1$$

$$2) P[X \geq 1] = 1 - P[X=0]$$

$$P[X=0] = \sum_{n=1}^{\infty} P[X=0 | N=n] \cdot P[N=n] \quad \text{(stesse formule
sulle probabilità
totali)}$$

$$= \sum_{n=1}^{\infty} \underbrace{\left(\frac{1}{2}\right)^n}_{\substack{\uparrow \\ \text{è la probabilità} \\ \text{di ottenere} \\ \text{0 test in } n \text{ lenz}}} \cdot \left(\frac{1}{2}\right)^n = \sum_{n=1}^{\infty} \left(\frac{1}{4}\right)^n = \frac{1}{1-\frac{1}{4}} - 1$$

$$= \frac{4}{3} - 1 = \frac{4-3}{3} = \frac{1}{3}$$

$$P[X \geq 1] = 1 - \frac{1}{3} = \frac{2}{3}$$

$$2) P[N=n | X=0] = \frac{P[X=0 | N=n] \cdot P[N=n]}{P[X=0]}$$

$$= \frac{\left(\frac{1}{2}\right)^n \cdot \left(\frac{1}{2}\right)^n}{\frac{1}{3}} = 3 \cdot \frac{1}{4^n}$$

$$3) P[X=1] = \sum_{n=1}^{\infty} P[X=1 | N=n] \cdot P[N=n] = \sum_{n=1}^{\infty} \binom{n}{1} \left(\frac{1}{2}\right)^n \left(\frac{1}{2}\right)^n =$$

$$= \sum_{n=1}^{\infty} n \cdot \left(\frac{1}{4}\right)^n =$$

Ora $\sum_{n=1}^{\infty} n q^n$ si calcola con il solito trucco:

$$\text{Da } \sum_{n=0}^{\infty} q^n = \frac{1}{1-q} \quad \text{per } |q| < 1, \text{ si fia}$$

$$\sum_{n=1}^{\infty} n q^{n-1} = \frac{1}{(1-q)^2} \Rightarrow \sum_{n=1}^{\infty} n q^n = \frac{q}{(1-q)^2}$$

Sostituendo $q = 1/L$, troviamo

$$\sum_{n=1}^{\infty} n \left(\frac{1}{L}\right)^{n-1} = \frac{1}{L} \cdot \frac{1}{\left(1 - \frac{1}{L}\right)^2} = \frac{1}{L} \cdot \frac{16}{9} = \frac{4}{9}.$$

Esercizio n° 2

$$f(x) = \begin{cases} 0 & x < 0 \\ 2\lambda x e^{-\lambda x^2} & x \geq 0 \end{cases}$$

1) f è una densità di probabilità. Infatti, $f(x) \geq 0 \forall x \in \mathbb{R}$ e

$$\int_{\mathbb{R}} f(x) dx = \int_0^\infty 2\lambda x e^{-\lambda x^2} dx = -e^{-\lambda x^2} \Big|_0^\infty = 1$$

2) La densità condizionata è

$$h(x_1, x_2, \dots, x_n; \lambda) = 2^n \lambda^n x_1 x_2 \dots x_n e^{-2 \sum_{i=1}^n x_i^2}$$

Sì ha

$$\ln h = n \ln \lambda - 2 \sum_{i=1}^n x_i^2 \quad (\text{e meno di costanti additive che non dipendono da } \lambda)$$

$$\frac{\partial}{\partial \lambda} \ln h = \frac{n}{\lambda} - \sum_{i=1}^n x_i^2 \stackrel{!}{=} 0$$

$$\frac{1}{n} = \frac{1}{\sum_{i=1}^n x_i^2} \Rightarrow \lambda = \frac{1}{\frac{1}{n} \sum_{i=1}^n x_i^2}$$

$$\hat{\lambda}_n = \frac{1}{\frac{1}{n} \sum_{i=1}^n x_i^2} \quad \begin{matrix} \text{è lo stima} \\ \text{to massime} \\ \text{verosimiglianza} \end{matrix}$$

3) Dal teorema di fattorizzazione di Fisher si vede che

$$T = \sum_{i=1}^n x_i^2 \quad \text{è una statistica sufficiente per } \lambda$$

$$4) P[X^2 \leq t] = P[X \leq \sqrt{t}] = \int_0^{\sqrt{t}} 2\lambda x e^{-\lambda x^2} dx$$

Segue che la densità di X^2 è

$$g(t) = \frac{d}{dt} P[X^2 \leq t] = \frac{d}{dt} \int_0^{\sqrt{t}} 2\lambda x e^{-\lambda x^2} dx =$$

= ~~result~~

~~diff~~

$$\frac{2\lambda \sqrt{t} e^{-\lambda t}}{2\sqrt{t}} = \lambda e^{-\lambda t}$$

che è la densità esponenziale

5) Dal punto 4 segue che $\frac{1}{n}(X_1^2 + X_2^2 + \dots + X_n^2)$ è un stimatore corretto di $\frac{1}{\lambda}$ (perché $E[X^2] = \frac{1}{\lambda}$).

Esercizio 3

La distribuzione di Fisher di parametri $n \times m$ è la legge di una variabile aleatoria $F_{n,m} = \frac{\frac{X}{n}}{\frac{Y}{m}}$ dove $X \sim \chi^2_n$ e $Y \sim \chi^2_m$.

Sono due variabili indipendenti distribuite come una $\chi^2(n)$

e una $\chi^2(m)$, rispettivamente.

Detto $F_d(n, m)$ il quantile corrispondente (cioè

$$P\left[\frac{X}{Y} \leq F_d(n, m)\right] = d$$

si ha

$$F_d(n, m) \cdot F_{1-d}(m, n) = 1$$

$$n_1 = 9$$

$$n_2 = 10$$

$$H_0: \bar{G}_1^2 = \bar{G}_2^2$$

Sotto l'ipotesi nulla

$\frac{S_1^2}{S_2^2}$ è una V.a. di Fisher $F(8, 9)$

La regione di rifiuto è

$$\frac{S_1^2}{S_2^2} > F_{\frac{\alpha}{2}}(8, 9) \quad o \quad \frac{S_1^2}{S_2^2} < F_{\frac{1-\alpha}{2}}(8, 9)$$

Nel nostro caso $\frac{S_1^2}{S_2^2} = \frac{1.5278}{1.6556} \approx 0.92281$

Con $\alpha = 5\% (0.05)$ $\frac{\alpha}{2} = 0.025$ e $1 - \frac{\alpha}{2} = 0.475$

(con $\alpha = 10\% (0.1)$) $\frac{\alpha}{2} = 0.05$ e $1 - \frac{\alpha}{2} = 0.95$

$$F_{0.05}(8, 9) = 3.23$$

$$F_{0.95}(8, 9) = 3.39$$

Quindi al 10% rifiuto ilo. A maggior regime ilo. vuchi accettate anche al 5%.