CORSO DI STATISTICA (parte 2) - ESERCITAZIONE 3

Dott.ssa Antonella Costanzo

a.costanzo@unicas.it

Esercizio 1. La v.c. Uniforme Continua

Secondo alcuni sondaggi sul sito della Apple (technical support site, *www.apple.com/support/itunes)*, il tempo di download di un gioco per l' iPod utilizzando una connessione a banda larga va dai 3 ai 6 minuti. Si assuma che il tempo segua una distribuzione uniforme continua compresa tra 3 e 6 minuti. Calcolare:

1. la probabilità di impiegare meno di 3 minuti per scaricare il gioco
2. la probabilità di impiegare meno di 4 minuti per scaricare il gioco
3. la probabilità che si impieghi un tempo compreso tra 4 e 5 minuti
4. Media e varianza del tempo di download

Sol.

Indichiamo con X la variabile casuale "tempo di download" X ~ U(a=3, b=6).

Se X è una variabile casuale uniforme continua:

distribuzione di probabilità

$$f(x) = \begin{Bmatrix} \dfrac{1}{b-a} & se\ x \in [a,b] \\ 0 & altrove \end{Bmatrix}$$

Valore atteso e varianza della v.c. uniforma continua

$E(X) = \frac{a+b}{2}$ $\qquad$ $Var(X) = \frac{(a-b)^2}{12}$

Funzione di ripartizione

$$F(X) = P(X \leq x) = \frac{x-a}{b-a}$$

1) $P(X \leq 3) = \frac{3-3}{6-3} = 0$

2) $P(X \leq 4) = \frac{4-3}{6-3} = \frac{1}{3}$

3) $P(4 \leq X \leq 5) = P(X \leq 5) - P(X \leq 4) = \frac{5-3}{6-3} - \frac{4-3}{6-3} = \frac{2}{3} - \frac{1}{3} = \frac{1}{3}$

4) $E(X) = \frac{a+b}{2} = \frac{3+6}{2} = \frac{9}{2}$ $\qquad Var(X) = \frac{(a-b)^2}{12} = \frac{9}{12}$

Esercizio 2. La v.c. Esponenziale (negativa)

Una banca ha il problema di ottimizzare la ripartizione di risorse umane tra il front-office (che è a contatto con il cliente), e il back-office (che svolge il relativo lavoro d'ufficio). A tal fine risulta importante conoscere i flussi di arrivo della clientela. Sapendo che si verificano in media 0.2 arrivi all'ora, ci si chiede qual è la probabilità che fino al prossimo arrivo trascorrano tra i 30 e i 45 minuti a partire da un istante qualsiasi.

Sol.

Si noti che in questo problema la probabilità del tempo fino al prossimo arrivo non dipende da quando c'è stato l'ultimo arrivo, quindi se si inizia a contare l'attesa dall'ultimo arrivo o da qualsiasi altro momento non cambia nulla. In questo senso allora risulta corretto applicare un modello senza memoria come l'esponenziale.

$X \sim Exp(\lambda)$ dove il parametro λ è dato dal reciproco della media $E(X) = \frac{1}{\lambda}$

Distribuzione di probabilità

$f(x) = \lambda e^{-\lambda x}$ con $\lambda > 0$

Funzione di ripartizione:

$F(x) = 1 - e^{-\lambda x}$ con $\lambda > 0$

Valore atteso e varianza della v.c. esponenziale

$$E(X) = \frac{1}{\lambda}$$

$$Var(X) = \frac{1}{\lambda^2}$$

Nel nostro caso:

$$E(X) = \frac{1}{\lambda} = 0.2$$

Il parametro λ è dato dal reciproco della media quindi λ =5

30 minuti corrispondono a 0.5 ore e 45 minuti corrispondono a 0.75 ore

$$P(0.5 < X < 0.75) = F_X(0.75) - F_X(0.5) = \left[1 - e^{-5(0.75)}\right] - \left[1 - e^{-5(0.5)}\right] = 0.9765 - 0.9179 = 0.059$$

Esercizio 3. La v.c Normale

Nell'agenzia di una grande banca il tempo X (in minuti) necessario a un impiegato di cassa per fornire al cliente il servizio richiesto si distribuisce normalmente con media 4.5 e scarto 1.1.

a) Qual è la probabilità che un servizio preso a caso dall'insieme dei servizi resi dalla banca ai clienti, richieda:

1) più di 6 minuti o meno di 5;
2) fra i 3.5 e i 6 minuti;
3) al massimo 3 minuti?

b) Qual è il tempo necessario per cliente se soltanto il 5% delle richieste necessita di un tempo superiore?

Sol.

Essendo $X \sim N(4.5; 1.1)$

a)

1. $P(X > 6 \cup X < 5) = P(X > 6) \cup P(X < 5) = P(X > 6) + P(X < 5)$

$$P(X > 6 \cup X < 5) = P\left(Z > \frac{6 - 4.5}{1.1}\right) + P\left(Z < \frac{5 - 4.5}{1.1}\right) = P(Z > 1.36) + P(Z < 0.45) = 0.7605$$

2. $P(3.5 < X < 6) = P\left(\frac{3.5-4.5}{1.1} < Z < \frac{6-4.5}{1.1}\right) = P(0.91 < Z < 1.36) = 0.9131 - [1 - 0.8186] = 0.7317$

3. $P(X < 3) = P\left(Z < \frac{3-4.5}{1.1}\right) = P(Z < -1.36) = 1 - P(Z < 1.36) = 0.0896$

b)

Si tratta di individuare il valore x tale che $P(X > x) = 0.05$ ovvero tale che:

$$P(X \leq x) = 0.95$$

$$P\left(Z \leq \frac{x - 4.5}{1.1}\right) = 0.95$$

Dalle tavole risulta che il valore più vicino a 0.95 è compreso tra 1.64 e 1.65, per cui approssimando il valore

$$z_{0.95} = 1.645$$

Partendo dalla versione standardizzata di X otteniamo che :

$$X_{0.05} = \mu + \sigma(z_{0.95}) = 4.5 + 1.1(1.645) = 6.3095$$

Esercizio 4. Applicazioni del Teorema di Chebyshev

Supponiamo che una banca riceva in media 50 clienti al giorno con uno scarto di 10. Si vuole calcolare:

1) la probabilità che nella giornata di domani, la banca riceverà tra i 30 ed i 70 clienti.

2) la probabilità che nella giornata di domani, la banca riceverà meno di 40 clienti oppure più di 60 clienti

Supponiamo ora che il numero di clienti che riceve la banca si distribuisca secondo una legge normale con stessa media e stesso scarto. Calcolare:

3) la probabilità che nella giornata di domani, la banca riceverà tra i 30 e i 70 clienti

Sol

1) A tale scopo sia X il numero dei clienti che la banca riceverà domani. Dobbiamo determinare:

$$P(\mu - k\sigma \leq X \leq \mu + k\sigma) = P(30 \leq X \leq 70)$$

Per il teorema di Chebyshev se una distribuzione di probabilità ha media μ e scarto quadratico medio σ, allora la probabilità che il valore di una variabile casuale con tale distribuzione differisca da μ per più di $k\sigma$ è inferiore a $\frac{1}{k^2}$. In particolare:

$$P(X < \mu - k\sigma \; e \; X > \mu + k\sigma) \leq \frac{1}{k^2}$$

In forma compatta: $P(|X - \mu| > k\sigma) \leq \frac{1}{k^2}$

Equivalentemente:

$$P(|X - \mu| > k) \leq \frac{\sigma^2}{k^2}$$

A partire dal risultato del Teorema di Chebyshev, la probabilità cercata si ricava come segue:

$$P(\mu - k\sigma \leq X \leq \mu + k\sigma) \geq 1 - \frac{1}{k^2}$$

$$P(|X - \mu| < k\sigma) \ \geq 1 - \frac{1}{k^2}$$

Ossia per complemento a uno del risultato del Teorema:

$$1 - P(|X - \mu| > k) \geq 1 - \frac{\sigma^2}{k^2}$$

Sapendo che: $\mu = 50$, $\sigma^2 = 100$ e che $|50 - 30| = |70 - 50| = 20 = k$

Allora:

$$P(|X - 50| > 20) \leq \frac{100}{20^2}$$

Da cui:

$$1 - P(|X - 50| > 20) \geq 1 - \frac{100}{20^2} \geq 0.75$$

Con una probabilità del 75% la banca riceverà domani tra i 30 e i 70 clienti

2) In questo caso dobbiamo determinare:

$$P(X < \mu - k\sigma \; e \; X > \mu + k\sigma) = P(X < 40 \; e \; X > 60)$$

Per il Teorema di Chebyshev

$$P(|X - \mu| > k) \leq \frac{\sigma^2}{k^2}$$

Sapendo che: $\mu = 50$, $\sigma^2 = 100$ e che $|50 - 40| = |60 - 50| = 10 = k$

Allora:

$$P(|X - 50| > 10) \leq \frac{100}{10^2}$$

Se il numero di clienti che riceve la banca si distribuisca secondo una legge normale con media pari a 50 e scarto quadratico medio 10 allora:

$$X \sim N(50, 10)$$

la probabilità che nella giornata di domani, la banca riceverà tra i 30 e i 70 clienti è pari a:

$$P(300 \le X \le 700) = P\left(\frac{30-50}{10} \le Z \le \frac{70-50}{10}\right) = P(-2 \le Z \le 2) = P(Z \le 2) - P(Z \le -2) =$$

$$= 0.9772 - [1 - 0.9772] = 0.9772 - 0.0228 = 0.9544$$

Nota: nel caso di distribuzione nota (X segue una legge normale) allora la probabilità di osservare tra i 30 e i 70 clienti nella giornata successiva è del 95%.

Esercizio 5. Verifica delle proprietà degli stimatori (piccoli campioni): efficienza e non distorsione

Da una popolazione su cui è definita una variabile X con media incognita μ, si estrae un campione casuale di numerosità *n*. Si definiscano le proprietà dei seguenti stimatori per μ e si valuti il "migliore" stimatore per la media incognita in termini di correttezza ed efficienza:

1. $T_1 = \frac{1}{n}\sum_{i=1}^{n} X_i$
2. $\boldsymbol{T_2 = \frac{1}{2}(X_1 + X_n)}$
3. $T_3 = X_2$
4. $\boldsymbol{T_4 = \frac{1}{3}(X_2 + X_8)}$

Uno stimatore T_n si dice non distorto per θ se $E(T_n) = \theta$ ossia quando si escludono deviazioni sistematiche nella stima di θ per cui la media dello stimatore coincide col parametro.

Non distorsione

- $E(T_1) = \frac{1}{n}E(\sum_{i=1}^{n} X_i) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}n\mu = \mu$

- $E(T_2) = \frac{1}{2}[E(X_1) + E(X_n)] = \frac{1}{2}[\mu + \mu] = \mu$

- $E(T_3) = E(X_2) = \mu$

- $E(T_4) = \frac{1}{3}[E(X_2) + E(X_8)] = \frac{1}{3}[\mu + \mu] = \frac{2}{3}\mu \neq \mu$

Efficienza (relativa)

- $Var(T_1) = \frac{1}{n^2}Var(\sum_{i=1}^{n} X_i) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n}$

- $Var(T_2) = \frac{1}{4}[Var(X_1) + Var(X_n)] = \frac{1}{4}[\sigma^2 + \sigma^2] = \frac{\sigma^2}{2}$

- $Var(T_3) = Var(X_2) = \sigma^2$

Tra i tre stimatori tutti corretti si sceglierà quello più efficiente, ossia quello con varianza più piccola. In questo caso, per campioni con n>2 si sceglierà lo stimatore T1.