

Esame di Calcolo delle Probabilità

(2 ore e 30 minuti)

Si prega di scrivere in maniera chiara, risposte non leggibili non saranno corrette. In tutti gli esercizi si richiede di illustrare il proprio lavoro.

Esercizio 1 (10 punti). *Ipotizziamo di lanciare due dadi separatamente più volte ed indichiamo con X_1 il numero di lanci del primo dado necessari per ottenere 1 e con X_2 il numero di lanci del secondo dado necessari per ottenere un numero pari.*

(a) *Si trovino le leggi di X_1 e X_2 .*

(b) *Si calcoli $\mathbb{P}(X_1 > X_2)$.*

Soluzione. (a) *Per costruzione X_1 e X_2 sono delle geometriche modificate di parametri rispettivamente $1/6$ e $1/2$, quindi*

$$\mathbb{P}(X_1 = k) = \frac{1}{6} \left(\frac{5}{6}\right)^{k-1} \quad e \quad \mathbb{P}(X_2 = k) = \frac{1}{2} \left(\frac{1}{2}\right)^{k-1} = \frac{1}{2^k}.$$

(b) *Osserviamo che*

$$\mathbb{P}(X_1 > X_2) = \sum_{k=1}^{\infty} \mathbb{P}(X_1 > X_2, X_2 = k) = \sum_{k=1}^{\infty} \mathbb{P}(X_1 > k, X_2 = k) = \sum_{k=1}^{\infty} \mathbb{P}(X_1 > k) \mathbb{P}(X_2 = k).$$

Calcoliamo $\mathbb{P}(X_1 > k)$

$$\mathbb{P}(X_1 > k) = \sum_{j=k+1}^{\infty} p(1-p)^{j-1} = (1-p)^k,$$

dove abbiamo posto $p = 1/6$. Di conseguenza, ponendo $q = 1/2$, si ha

$$\begin{aligned} \mathbb{P}(X_1 > X_2) &= \sum_{k=1}^{\infty} (1-p)^k \mathbb{P}(X_2 = k) = \sum_{k=1}^{\infty} (1-p)^k q(1-q)^{k-1} \\ &= q(1-p) \sum_{k=1}^{\infty} [(1-p)(1-q)]^{k-1} \\ &= \frac{q(1-p)}{1 - (1-p)(1-q)}. \end{aligned}$$

Sostituendo i valori di p e q abbiamo

$$\mathbb{P}(X_1 > X_2) = \frac{5/12}{1 - 1/12} = 5/11.$$

Esercizio 2 (10 punti). Nello studiare un test diagnostico si è determinato che se una persona è malata il test è positivo nel 96% dei casi e se una persona è sana il test è negativo nel 94% dei casi. Una persona si sottopone al test e questo risulta positivo. Qual è la probabilità che tale persona sia effettivamente malata supponendo che il 5% della popolazione sia infetta? Si descriva un algoritmo per verificare il risultato ottenuto su R .

Soluzione. Chiamiamo P l'evento il paziente è positivo e N l'evento il paziente è negativo, ovviamente $P^C = N$. Possiamo assumere che, dato che il 5% della popolazione è infetta, si abbia $\mathbb{P}(P) = 5\%$ (e ovviamente $\mathbb{P}(N) = 95\%$). Definiamo ora gli eventi $T_+ =$ "il test risulta positivo" e $T_- =$ "il test risulta negativo", ovviamente $T_+^C = T_-$. Siamo interessati a calcolare $\mathbb{P}(P|T_+)$. Sappiamo che

$$\mathbb{P}(T_+|P) = 96\% \quad \text{e} \quad \mathbb{P}(T_-|N) = 94\%.$$

Usiamo il Teorema di Bayes

$$\mathbb{P}(P|T_+) = \frac{\mathbb{P}(T_+|P)\mathbb{P}(P)}{\mathbb{P}(T_+)} = \frac{\mathbb{P}(T_+|P)\mathbb{P}(P)}{\mathbb{P}(T_+|P)\mathbb{P}(P) + \mathbb{P}(T_+|N)\mathbb{P}(N)}.$$

L'unica quantità non nota è $\mathbb{P}(T_+|N)$, ma questa si ottiene osservando che

$$\mathbb{P}(T_+|N) = 1 - \mathbb{P}(T_-|N).$$

Si ottiene quindi

$$\mathbb{P}(P|T_+) = \frac{0.96 \cdot 0.05}{0.96 \cdot 0.05 + 0.06 \cdot 0.95} \approx 46\%.$$

Per verificare il risultato in R sarà sufficiente simulare l'esperimento molte volte e calcolare la frequenza empirica del risultato desiderato. Per il teorema dei grandi numeri questa convergerà alla probabilità teorica. Per esempio si può utilizzare il seguente codice

```

1  n=10000 #numero tentativi
2  pop=c('sano','infetto'); #popolazione
3  test=c('positivo','negativo'); #esiti test
4  n_positivi=0 #numero di test positivi
5  s=0; #numero di positivi con test positivo
6
7  for (i in c(1:n)){
8    persona=sample(pop,size=1,replace=TRUE,prob=c(0.95,0.05))
9    if (persona=='sano'){
10     #Se la persona e' sana la prob. di avere un test negativo e' 0.94
11     risultato_test=sample(test,size=1,replace=TRUE,prob=c(0.06,0.94))
12     if (risultato_test=='positivo'){
13       n_positivi=n_positivi+1
14     }
15   } else {
16     #Se la persona e' infetta la prob. di avere un test positivo e' 0.96
17     risultato_test=sample(test,size=1,replace=TRUE,prob=c(0.96,0.04))
18     if (risultato_test=='positivo'){
19       n_positivi=n_positivi+1
20     }
21   }
22   if (persona=='infetto' &&& risultato_test=='positivo'){
23     s=s+1
24   }
25 }
26 p_empirica=s/n_positivi

```

Esercizio 3 (10 punti). Si consideri la funzione

$$f(x) = \begin{cases} cx^3(1-x^2)^2 & 0 < x < 1, \\ 0 & \text{altrimenti,} \end{cases}$$

con $c \in \mathbb{R}$.

- (a) Si determini c in modo che f sia una densità
(Suggerimento: si ricordi che $\int_0^1 t^{\alpha-1}(1-t)^{\beta-1}dt = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$);
- (b) Fissato c come sopra, sia X una variabile aleatoria assolutamente continua con densità f .
Se ne calcoli la speranza matematica.

Soluzione. (a) Affinché f sia una densità serve che $\int_{-\infty}^{\infty} f(x)dx = 1$ e $f(x) \geq 0$ quasi ovunque.
Per la seconda basta $c \geq 0$, cerchiamo c che soddisfi la prima, cioè tale che

$$1 = \int_{-\infty}^{\infty} f(x)dx = c \int_0^1 x^3(1-x^2)^2 dx.$$

Con la sostituzione $t = x^2$ (da qui $dt = 2xdx$) otteniamo

$$1 = \frac{c}{2} \int_0^1 t(1-t)^2 dt = \frac{c}{2} \int_0^1 t^{2-1}(1-t)^{3-1} dt = \frac{c}{2} \frac{\Gamma(2)\Gamma(3)}{\Gamma(5)} = \frac{1 \cdot 2 \cdot c}{2 \cdot 4!},$$

da cui $c = 24$.

- (b) Calcoliamo ora la speranza matematica di X

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} xf(x)dx = 24 \int_0^1 x^4(1-x^2)^2 dx,$$

sostituendo x^2 con t e osservando che, di conseguenza, $x = \sqrt{t}$ per $t \geq 0$, otteniamo

$$\begin{aligned} \mathbb{E}(X) &= 24 \int_0^1 \frac{t^2}{2\sqrt{t}}(1-t)^2 dt = 12 \int_0^1 t^{3/2}(1-t)^2 dt = 12 \int_0^1 t^{5/2-1}(1-t)^{3-1} dx \\ &= \frac{12\Gamma(5/2)\Gamma(3)}{\Gamma(11/2)} = \frac{24\Gamma(5/2)}{9/2 \cdot 7/2 \cdot 5/2 \cdot \Gamma(5/2)} = \frac{24}{109/8} = \frac{192}{315} = \frac{64}{105}. \end{aligned}$$

Esercizio 4 (10 punti). Sia $(X_n)_n$ una successione di variabili aleatorie indipendenti e identicamente distribuite su uno spazio di probabilità $(\Omega, \mathcal{A}, \mathbb{P})$. Assumiamo che X_1 abbia legge uniforme su $[0, 1]$. Definiamo ora le seguenti successioni di variabili aleatorie

$$X^{(n)} = \max\{X_1, \dots, X_n\}, \quad Y^{(n)} = n(1 - X^{(n)}).$$

- (a) Si dimostri che $X^{(n)}$ e $Y^{(n)}$ sono effettivamente variabili aleatorie (per ogni n);
- (b) Si calcolino le densità e funzioni di ripartizione di $X^{(n)}$ e $Y^{(n)}$;
- (c) Si dimostri che $Y^{(n)} \xrightarrow{L} Z$ per $n \rightarrow \infty$, i.e., che $Y^{(n)}$ converge in legge ad una variabile Z quando n va all'infinito, e si calcoli la funzione di ripartizione di Z . È una legge nota?

Soluzione. (a) Se $X^{(n)}$ è una variabile aleatoria allora lo è anche $Y^{(n)}$ in quanto funzione continua di variabile aleatoria. Basta quindi verificare che l'insieme $\{X^{(n)} \geq t\}$ è un evento per ogni $t \in \mathbb{R}$. Per l'indipendenza delle X_j si ha

$$\{X^{(n)} \leq t\} = \{\max\{X_1, \dots, X_n\} \leq t\} = \{X_1 \leq t, \dots, X_n \leq t\} = \bigcap_{j=1}^n \{X_j \leq t\}.$$

Dunque $\{X^{(n)} \leq t\}$ è un evento in quanto intersezione finita di eventi $(\{X_j \leq t\} \in \mathcal{A}$ in quanto X_j è a sua volta una variabile aleatoria)¹.

(b) Calcoliamo prima le funzioni di ripartizione.

$$\begin{aligned} F_{X^{(n)}}(t) &= \mathbb{P}(X^{(n)} \leq t) = \mathbb{P}(\max\{X_1, \dots, X_n\} \leq t) = \mathbb{P}(X_1 \leq t, \dots, X_n \leq t) \\ &= \prod_{j=1}^n \mathbb{P}(X_j \leq t) = \begin{cases} 0 & t \leq 0, \\ t^n & 0 < t < 1, \\ 1 & t \geq 1, \end{cases} \end{aligned}$$

$$\begin{aligned} F_{Y^{(n)}}(t) &= \mathbb{P}(Y^{(n)} \leq t) = \mathbb{P}(n(1 - X^{(n)}) \leq t) = \mathbb{P}\left(X^{(n)} \geq 1 - \frac{t}{n}\right) = 1 - \mathbb{P}\left(X^{(n)} \leq 1 - \frac{t}{n}\right) \\ &= 1 - \begin{cases} 0 & 1 - \frac{t}{n} \leq 0, \\ \left(1 - \frac{t}{n}\right)^n & 0 < 1 - \frac{t}{n} < 1, \\ 1 & 1 - \frac{t}{n} \geq 1, \end{cases} = 1 - \begin{cases} 0 & t \geq n, \\ \left(1 - \frac{t}{n}\right)^n & 0 < t < n, \\ 1 & t \leq 0, \end{cases} \\ &= \begin{cases} 0 & t \leq 0, \\ 1 - \left(1 - \frac{t}{n}\right)^n & 0 < t < n, \\ 1 & t \geq n. \end{cases} \end{aligned}$$

Possiamo ora calcolare le densità derivando le funzioni di ripartizione

$$\begin{aligned} f_{X^{(n)}}(t) &= \begin{cases} 0 & t \leq 0, \\ nt^{n-1} & 0 < t < 1, \\ 0 & t \geq 1, \end{cases} \\ f_{Y^{(n)}}(t) &= \begin{cases} 0 & t \leq 0, \\ \frac{1}{n} \left(1 - \frac{t}{n}\right)^{n-1} & 0 < t < n, \\ 0 & t \geq n. \end{cases} \end{aligned}$$

(c) Calcoliamo per un t generico $\lim_{n \rightarrow \infty} F_{Y^{(n)}}(t)$. Ricordando che $\lim_{n \rightarrow \infty} \left(1 - \frac{t}{n}\right)^n = e^{-t}$ si ha

$$\lim_{n \rightarrow \infty} F_{Y^{(n)}}(t) = \begin{cases} 0 & t \leq 0, \\ 1 - e^{-t} & 0 < t. \end{cases}$$

Ovviamente questa è una funzione di ripartizione di una variabile Esponenziale di parametro $\lambda = 1$, allora $Y^{(n)} \xrightarrow{\mathcal{L}} Z$ per $n \rightarrow \infty$ con $Z \sim \text{Esp}(1)$.

¹Si osservi anche che, poiché le X_j sono identicamente distribuite si ha che $\{X_j \leq t\} = \{X_1 \leq t\}$ per ogni j e quindi $\{X^{(n)} \leq t\} = \{X_1 \leq t\}$.

Tabella 2: Legge, funzione di ripartizione, media, e varianza più importanti distribuzioni.

	Densità	F. di rip.	Media	Varianza
$B(n, p)$	$p_k = \binom{n}{k} p^k (1-p)^{n-k}$		np	$np(1-p)$
Ipergeometrica	$p_k = \frac{\binom{r}{k} \binom{b}{n-k}}{\binom{b+r}{n}}$		$\frac{nr}{b+r}$	$\frac{nr b(b+r-n)}{(b+r)^2(b+r-1)}$
$P(\lambda)$	$p_k = e^{-\lambda} \frac{\lambda^k}{k!}$		λ	λ
$G(p)$	$p_k = p(1-p)^k$	$1 - (1-p)^{k+1}$	$\frac{1-p}{p}$	$\frac{1-p}{p^2}$
$Exp(\lambda)$	$f(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & x \leq 0 \end{cases}$	$F(x) = \begin{cases} 1 - e^{-\lambda x} & x > 0 \\ 0 & x \leq 0 \end{cases}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
$N(\mu, \sigma^2)$	$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)}$		μ	σ^2
$\Gamma(\alpha, \lambda)$	$f(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} & x > 0 \\ 0 & x \leq 0 \end{cases}$		$\frac{\alpha}{\lambda}$	$\frac{\alpha}{\lambda^2}$