Elementi di Probabilità e Statistica, anno 2018

Gli esercizi che seguono sono stati proposti (da me o da Franco Flandoli) durante le lezioni e le prove d'esame del corso "Elementi di Probabilità e Statistica" negli anni accademici a partire dal 2010.

1 Parte 1

Esercizio 1.1. Alle origini del Calcolo delle Probabilità si colloca il *Cavaliere De Méré* che, da matematico amatoriale ed incallito giocatore, ha posto diversi problemi (la maggior parte dei quali risolti da Pascal).

Un giorno egli evidenziò un apparente paradosso: aveva osservato empiricamente che, lanciando 3 dadi, era più facile ottenere come somma dei numeri usciti il numero 11 piuttosto che il 12, e questo nonostante il numero 11 si possa ottenere con 6 combinazioni possibili (6–4–1, 6–3–2, 5–5–1, 5–4–2, 5–3–3 e 4–4–3), tante quante quelle con cui si può ottenere il 12 (quali?).

Tuttavia Pascal mostrò che c'era un errore nel suo ragionamento: sapresti ricostruire quale?

Esercizio 1.2. Si lancia tre volte una moneta equilibrata, e si considerino gli eventi A *"le facce uscite non sono tutte eguali"* e B *" al più una faccia è testa"*.

Gli eventi A e B sono indipendenti?

Qual è la risposta se la moneta non è equilibrata?

Esercizio 1.3. Un dado equilibrato, con le facce numerate da 1 a 6, viene lanciato n volte: qual è la probabilità che il numero 6 esca esattamente 2 volte?

Per quale valore di n questa probabilità è massima?

Esercizio 1.4. Quante volte almeno si deve lanciare un dado affinché ci sia una probabilità superiore al 99% che esca almeno un 6?

Esercizio 1.5. In una città, il 17% della popolazione si è vaccinato contro l'influenza : all'apice dell'epidemia di influenza, le persone non vaccinate si ammalano con probabilità 0,12 e quelle vaccinate invece con probabilità 0,02.

Qual è la probabilità di ammalarsi? Qual è la probabilità che una persona ammalata si sia vaccinata ?

Esercizio 1.6. Una fabbrica produce dei componenti elettronici che vende in scatole di 10 pezzi. Prima di essere messa in vendita, ogni scatola viene controllata nel modo seguente: si scelgono a caso 5 pezzi e se almeno 4 risultano funzionanti la scatola passa alla vendita.

a) Qual è la probabilità che una scatola con esattamente 8 pezzi funzionanti passi alla vendita?

b) Stessa domanda per una scatola con 4 pezzi funzionanti.

2 Parte 2

Esercizio 2.1. Si lanciano due dadi equilibrati, e consideriamo i due eventi A "*la somma dei due numeri usciti è 7*" , e B "*il numero comparso nel primo dado è 4*".

1) Provare che gli eventi A e B sono indipendenti.

2) Più in generale, indicando rispettivamente con le v.a. X ed Y il numero uscito nel primo e nel secondo dado, si considerino gli eventi $\{ X+Y=n \}$ e $\{ X=k \}$: esaminare per quali valori di n e di k questi due eventi risultano indipendenti.

Esercizio 2.2. In un concorso la prova scritta consiste in un test composto di 5 domande: per ciascuna di esse il test richiede di scegliere una tra 4 possibili risposte, delle quali una sola è corretta. Il test viene valutato attribuendo punteggio 1 ad ogni risposta esatta e $(-\frac{1}{4})$ ad ogni risposta errata.

Si consideri un concorrente che risponda a caso ad ogni domanda:

a) Qual è la probabilità che ottenga in totale 2,5 punti ?

b) Quale punteggio ottiene in media?

Esercizio 2.3. Sia X una variabile aleatoria a valori interi (relativi): trovare delle condizioni necessarie e sufficienti affinchè X e X^2 siano indipendenti tra loro.

Esercizio 2.4. Si consideri una v.a. discreta X che prende i valori 2^i, per $i = 0, 1, \ldots$ ed è tale che $\mathbf{P}(X=1) = \mathbf{P}(X=2^0) = \frac{3}{4}$; e, per $i \geq 1$, $\mathbf{P}(X=2^i) = 5^{-i}$.

a) Dire se la v.a ammette valore atteso ed in tal caso calcolarlo.

b) Dire se ammette momento secondo, terzo, e più in generale per quali valori $p \geq 1$ ammette momento di ordine p.

Esercizio 2.5. Sia X una variabile aleatoria con distribuzione geometrica di parametro p : qual è la probabilità che X assuma un valore pari, e quella che assuma un valore dispari?

Sia ora Y una variabile con distribuzione di *Poisson* di parametro λ : rispondere alle stesse domande.

Attenzione: il numero 0 non è considerato né pari né dispari. Per rispondere alla seconda domanda si suggerisce di scrivere il numero $(e^{\lambda} + e^{-\lambda})$ come somma di una serie.

Esercizio 2.6. Vi sono due monete, indistinguibili al tatto, delle quali una è perfettamente equilibrata e l'altra è truccata: per quest'ultima è $\frac{2}{3}$ la probabilità che esca testa. Se ne sceglie una a caso e la si lancia 2 volte di seguito, e siano X_1 e X_2 i risultati dei due lanci (X_i vale 1 oppure 0 a seconda che all'i–mo lancio sia uscita testa oppure croce).

1) Le variabili X_1 e X_2 sono indipendenti?

2) Se è uscita testa in entrambi i lanci, qual è la probabilità che sia stata scelta la moneta truccata?

Esercizio 2.7. Consideriamo la misura $\mathbf{m}$ definita sull'insieme dei naturali strettamente positivi tale che $\mathbf{m}(k) = k^{-1}$ e consideriamo, per ogni n, la funzione f_n definita da:

$$f_n(k) = \begin{cases} k^{-\frac{1}{n}} & \text{se}\, k \geq n \\ 0 & \text{se}\, k < n \end{cases}$$

1) Le funzioni f_n sono integrabili rispetto a $\mathbf{m}$?
2) Convergono ad un limite f, e questo limite è integrabile?
3) Si può *passare al limite* sotto il segno d'integrale?

Esercizio 2.8. Da una moneta truccata se ne può ottenere una equilibrata nel modo seguente. Si lancia due volte, se esce TC si decide che è uscita “testa”, se esce CT si decide per “croce” e se i due risultati sono eguali si riprova altre due volte e di seguito fino ad ottenere due risultati diversi.

a) Provare che la “moneta” ottenuta in questo modo è effettivamente equilibrata.

b) Quanti lanci in media è necessario effettuare per arrivare al risultato?

Esercizio 2.9. Provare che, se X_n è una variabile Binomiale di parametri n e $\frac{\lambda}{n}$ ed X una variabile di Poisson di parametro λ, per ogni intero positivo k si ha

$$\lim_{n\to\infty} \mathbf{P}\big\{X_n = k\big\} = \mathbf{P}\big\{X = k\big\}$$

Esercizio 2.10. Siano X e Y due variabili aleatorie *geometriche* di parametro p, indipendenti: calcolare $\mathbf{P}(X = Y)$ e $\mathbf{P}(X < Y)$.

Esercizio 2.11. Siano X e Y due variabili di *Poisson* di parametri rispettivamente λ e μ, indipendenti.

a) Determinare la distribuzione di probabilità di $S = X + Y$.

b) Determinare la distribuzione condizionata di X sapendo che $S = n$.

Esercizio 2.12. Siano X ed Y due variabili indipendenti, equidistribuite, che prendono i valori 1 e -1 con probabilità p e $1-p$ $(0<p<1)$, sia poi $U = XY$.

Provare che X e U sono indipendenti se e solo se $p = \frac{1}{2}$.

Esercizio 2.13. Il numero di clienti che si presentano in un giorno in un grande magazzino è rappresentato da una v.a. X con distribuzione di *Poisson* di parametro λ; inoltre ogni cliente, indipendentemente dagli altri, ha probabilità p di essere derubato. Indichiamo con Y la v.a. che indica il numero di clienti che sono stati derubati.

1) Calcolare la distribuzione di probabilità della v.a. Y (Suggerimento: osservare che $\{Y = n\} = \cup_{k=n}^{+\infty}\{Y = n\,,\, X = k\}$)

2) Calcolare la distribuzione di probabilità della v.a. $Z = X - Y$ (numero di clienti non derubati).

3) Le v.a. X e Y sono indipendenti? Le v.a. Y e Z sono indipendenti?

Esercizio 2.14. Tra tutte le variabili aleatorie discrete che prendono solo i valori 1, 2 e 3 e che hanno valore atteso $\mathbf{E}[X] = 2$, trovare quelle che hanno varianza rispettivamente massima e minima.

Esercizio 2.15 (La variabile multinomiale). Consideriamo un esperimento che ha k possibili esiti, ciascuno con probablità p_i $(p_i > 0\,,\, p_1 + \cdots + p_k = 1)$: di esso si fanno n prove in condizioni di indipendenza ed indichiamo, per ogni i, con X^i la variabile che indica quante volte si è realizzato l'esito i.

Si chiama *variabile multinomiale* (di parametri $(n; p_1, \ldots, p_k)$) la variabile vettoriale $\mathbf{X} = (X^1, \ldots, X^k)$.

Determinare la distribuzione di probabilità della variabile $\mathbf{X}$ e specificare se le componenti X^i sono indipendenti.

Notiamo che, se X è *binomiale* di parametri (n, p), la coppia (X,n-X) è multinomiale di parametri $(n; p, 1-p)$.

3 Parte 3

Esercizio 3.1. Sia X una v.a. con densità a valori positivi: provare che vale la formula

$$\mathbf{E}[X] = \int_0^{+\infty} \mathbf{P}\big\{X > x\big\}\, \mathrm{d}x$$

Esercizio 3.2. Dire se le seguenti funzioni possono essere funzioni di ripartizione, ed in tal caso se la probabilità associata è definita da una densità.

Specificare inoltre se una v.a. che abbia quella legge di probabilità ammette valore atteso.

$$F(x) = \frac{1}{2} + \frac{1}{\pi}\operatorname{arctg} x$$

$$G(x) = \begin{cases} 0 & x < 0 \\ 1 - \frac{1}{2(1+x)} & x \geq 0 \end{cases}$$

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

$$K(x) = \begin{cases} 0 & x < 0 \\ 1 - (x-1)^2 & 0 \leq x < 1 \\ 1 & x \geq 1 \end{cases}$$

Esercizio 3.3. Supponiamo che la densità congiunta di una variabile doppia (X, Y) si possa scrivere nella forma $f(x, y) = h(x).k(y)$, dove h e k sono due funzioni boreliane positive: provare che X e Y sono indipendenti. Chi sono rispettivamente le densità di X e di Y ?

Esercizio 3.4. Siano X e Y due variabili indipendenti con densità esponenziale di parametro 1, e siano $U = X + Y$ e $V = \frac{X}{X+Y}$.

a) Calcolare la densità congiunta di U, V.

b) U e V sono indipendenti?

c) Verificare che vale l'eguaglianza

$$\mathbf{E}\left[\frac{X}{X+Y}\right] = \frac{\mathbf{E}[X]}{\mathbf{E}[X] + \mathbf{E}[Y]}$$

Esercizio 3.5. Sia (X, Y) una variabile doppia avente densità

$$f(x, y) = \begin{cases} \mathrm{e}^{-x} & \text{per } x > y > 0 \\ 0 & \text{altrimenti} \end{cases}$$

e sia $Z = X - Y$.

a) Qual è la densità di Z ? Si tratta di una densità nota ?

b) Le variabili Y e Z sono indipendenti ?

Esercizio 3.6. Sia (X, Y) una variabile doppia uniformemente distribuita sul cerchio unitario $\{(x, y) \mid x^2 + y^2 \leq 1\}$.

a) Calcolare le densità marginali di X e di Y. Le componenti sono indipendenti?

b) Calcolare le densità del *modulo* e dell'*argomento*; più precisamente le densità delle variabili aleatorie $R = \sqrt{X^2 + Y^2}$ e $T = \operatorname{arctg}\left(\frac{Y}{X}\right)$.

Esercizio 3.7. Sia (X, Y) una variabile aleatoria doppia con densità congiunta

$$f(x,y) = \begin{cases} x\,\mathrm{e}^{-(x+y)} & \text{se} \quad x>0 \quad \text{e} \quad y>0 \\ 0 & \text{altrimenti} \end{cases}$$

a) X e Y sono indipendenti ?
b) Calcolare (se esiste) il valore atteso di $\frac{Y}{X}$.
c) Calcolare la densità della variabile $X+Y$.
d) Calcolare la densità della variabile $U = \min(X, Y)$.

Esercizio 3.8. Consideriamo una variabile aleatoria doppia (X, Y) avente come densità la funzione

$$f(x,y) = \begin{cases} 10\,x^2 y & \text{se} \quad 0<y<x<1 \\ 0 & \text{altrimenti} \end{cases}$$

a) Le componenti X e Y sono indipendenti?
b) Poniamo $U = X$ e $V = \frac{X}{Y}$: calcolare la densità del vettore (U, V) .
c) Le variabili U e V sono indipendenti?
d) Calcolare $\mathbf{P}\left\{X > 2Y \,\middle|\, Y < \frac{1}{2}\right\}$.

4 Parte 4

Esercizio 4.1. Sia $(X_n)_{n\geq 1}$ una successione di variabili aleatorie indipendenti uniformemente distribuite sull'intervallo $[0, 1]$ e siano rispettivamente $M_n = \max(X_1, \ldots, X_n)$ e $V_n = \min(X_1, \ldots, X_n)$.

a) Calcolare le densità di M_n e V_n.

b) Indagare sulla convergenza in probabilità delle due successioni $(M_n)_{n\geq 1}$ e $(V_n)_{n\geq 1}$.

Esercizio 4.2. Sia, per $n \geq 1$, f_n la funzione definita da

$$f_n(x) = \begin{cases} c_n . x^{2n} & -1 < x < 1 \\ 0 & \text{altrove} \end{cases}$$

Dopo aver identificato la costante c_n che rende la funzione sopra scritta una densità, si consideri per ogni n una v.a. X_n avente densità f_n : esaminare se la successione $(X_n)_{n\geq 1}$ converge in legge ed eventualmente a quale limite.

Esercizio 4.3. Sia data una successione $(X_n)_{n\geq 1}$ di v.a. indipendenti equidistribuite, con densità

$$f(x) = \begin{cases} 2x & \text{se } 0 \leq x \leq 1 \\ 0 & \text{altrove} \end{cases}$$

Indagare sulla convergenza in legge della successione di variabili aleatorie $(Y_n)_{n \geq 1}$ definite da

$$Y_n = 2\sqrt{n}\min(X_1, \ldots, X_n).$$

Esercizio 4.4. Una compagnia aerea dispone di un aeromobile con 160 posti e, sapendo per esperienza che in media 4 passeggeri su 100 che hanno prenotato il volo poi non si presentano, accetta fino a 164 prenotazioni per ogni volo. Supponiamo che per un dato volo siano state effettuate esattamente 164 prenotazioni.

a) Qual è il numero medio di passeggeri che si presentano all'imbarco?

b) Qual è (approssimativamente) la probabilità che la compagnia sia costretta a rifiutare l'imbarco a qualche passeggero per esaurimento dei posti sull'aereo?

5 Parte 5

Esercizio 5.1. Consideriamo un campione $X_1, \ldots, X_n$ di variabili di *Poisson* di parametro $\theta, \theta > 0$: che cosa si può dire circa un intervallo di fiducia per θ ed un test dell'ipotesi $\mathcal{H}_0\big)\ \theta \leq \theta_0$ contro l'alternativa $\mathcal{H}_1\big)\ \theta > \theta_0$?

Esercizio 5.2. Si vuole verificare con quale frequenza si presenta tra i neonati una certa malformazione, più precisamente si vuole verificare l'ipotesi $\mathcal{H}_0\big)\ \theta \leq 0,02$, essendo θ la probabilità (sconosciuta) con la quale si presenta questa malformazione: per fare questo si controllano delle cartelle cliniche di neonati fino a quando se ne trova una nella quale compare questa malformazione. Pianificare un test per decidere al livello 0,1 , sulla base del numero di cartelle che è stato necessario verificare, se l'ipotesi può essere accettata.

Un procedimento più preciso si otterrebbe continuando a verificare cartelle fino a quando non se ne trovano 10 contenenti questa malformazione: esaminare se i calcoli sono agevoli in questo caso.

Esercizio 5.3. Sia $X_1, \ldots, X_n$ un campione di taglia n e legge *geometrica* di parametro θ $\big(0 < \theta < 1\big)$. Determinare un riassunto esaustivo. Esiste una stima di massima verosimiglianza, una stima consistente?

Esercizio 5.4. Consideriamo come insieme dei parametri gli interi strettamente positivi $k \geq 1$ e sia $\mathbf{m}^k$ la distribuzione di probabilità uniforme su $\big\{1, \ldots, k\big\} \subseteq \mathbb{N}$.

Sia dato un campione di taglia n e legge $\mathbf{m}^k$: considerare le stesse domande dell'esercizio precedente.

Esercizio 5.5. Si consideri, per $\theta > 1$, la distribuzione di probabilità $\mathbf{m}^\theta$ sugli interi strettamente positivi $\mathbb{N}^* = \{1, 2, \dots\}$ tale che $\mathbf{m}^\theta(k) = \zeta(\theta)^{-1}\, k^{-\theta}$, essendo

$$\zeta(s) = \sum_{n=1}^{+\infty} \frac{1}{n^s}$$

Dato un campione di taglia n e legge $\mathbf{m}^\theta$, considerare le stesse domande degli esercizi precedenti.

Osservazione: la funzione ζ sopra definita è la celebre *funzione zeta di Riemann*, molto importante in teoria dei numeri. Questa funzione è stata studiata approfonditamente, ma di essa non si può dare un'espressione esplicita in termini di funzioni elementari.

Esercizio 5.6. Consideriamo un campione di taglia n di v.a. con densità

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

dove $0 < \theta < +\infty$.

a) Indagare se esiste una statistica esaustiva T e la stima di massima verosimiglianza di θ.

b) Esaminare se tale stima è corretta.

Si vuole esaminare ora il test dell'ipotesi

$$\mathcal{H}_0)\quad \theta \leq 1 \qquad \text{contro l'alternativa} \qquad \mathcal{H}_1)\quad \theta > 1$$

utilizzando come regione critica $D = \{T > c\}$: determinare la costante c in modo tale che il test sopra indicato abbia livello α.

Esercizio 5.7. Siano $X_1, \dots, X_n$ variabili aleatorie indipendenti equidistribuite, dotate di momento primo e secondo. Tra tutte le stime *lineari* del valore atteso (cioè della forma $U = \sum_{i \leq n} a_i X_i$) , trovare quella corretta di varianza minima.

Esercizio 5.8. Viene condotto un sondaggio telefonico per determinare la percentuale di famiglie che vedono un certo programma televisivo: se si desidera che, nella determinazione di tale percentuale, l'errore non sia superiore a 0,02 con un grado di fiducia del 90%, quante famiglie almeno devono essere intervistate?

Esercizio 5.9. Sia $X_1, \dots, X_n$ un campione di taglia n con densità uniforme sull'intervallo $[0, \theta]$, $0 < \theta < +\infty$.

a) Indagare se esiste una *statistica esaustiva* e trovare una stima *corretta* di θ.

b) Trovare un intervallo di fiducia al livello $1-\alpha$ per il parametro θ (si suggerisce di cercare un intervallo di fiducia della forma $\big[T\,,\,T(1+d)\big]$ con d da calcolare opportunamente).

c) Trovare la regione critica di un test dell'ipotesi

$$\mathcal{H}_0)\quad \theta \geq 1 \qquad \text{contro l'alternativa} \qquad \mathcal{H}_1)\quad \theta < 1$$

ad un livello α prefissato.

6 Parte 6

Esercizio 6.1. Vengono prodotti artigianalmente dei manufatti che dovrebbero essere lunghi 120 cm, e si considera che la produzione e' buona se almeno il 90 % hanno una lunghezza compresa tra i 118 ed i 122 cm : assumendo che la variabile aleatoria che rappresenta la lunghezza dei manufatti sia gaussiana, imporre delle limitazioni sulla varianza affinchè tale condizione sia soddisfatta.

Se vengono misurati 27 pezzi e si trova $\sum_{i\leq 27}\big(x_i-120\big)^2 = 54,86$, si può accettare al livello 0,05 l'ipotesi che la produzione sia di buona qualità ?

Esercizio 6.2. Una ditta farmaceutica, che propaganda un nuovo farmaco, sostiene che abbrevia di almeno 4 giorni la durata di una malattia infettiva rispetto ai farmaci tradizionali: a sostegno di questa tesi riporta i dati di un esperimento condotto, come si usa dire, *in doppio cieco*, somministrando a un gruppo di 11 pazienti il nuovo farmaco e ad un gruppo di controllo di 7 pazienti i farmaci tradizionali.

I dati delle durate delle malattie, espresse in giornate, sono i seguenti: $\sum_{i\leq 11} x_i = 91, \qquad \sum_{i\leq 11} x_i^2 = 827, \qquad \sum_{j\leq 7} y_j = 80, \qquad \sum_{j\leq 7} y_j^2 = 976.$

Pianificare un test per verificare, al livello 0,05 , se l'affermazione della ditta può essere ritenuta corretta.

Esercizio 6.3. Siano $X_1,\ldots,X_n,X_{n+1}$ indipendenti con densità gaussiana $N(\mu,\sigma^2)$ (μ e σ^2 sconosciuti): siamo interessati ad utilizzare i valori osservati di $X_1,\ldots,X_n$ per determinare un intervallo di fiducia (detto *intervallo di previsione*) per X_{n+1} al livello $1-\alpha$. A tale scopo, denotiamo $\overline{X}_n = \frac{X_1+\ldots+X_n}{n}$ e $S_n^2 = \frac{\sum_{i\leq n}(X_i-\overline{X}_n)^2}{n-1}$.

a) Determinare la distribuzione di probabilità di $\frac{X_{n+1}-\overline{X}_n}{S_n}$;

b) ottenere un intervallo di fiducia per la variabile sopra scritta e dedurne un *intervallo di previsione* per X_{n+1} dati $X_1,\ldots,X_n$.

Esercizio 6.4. Siano $X_1,\dots,X_n$ e $Y_1,\dots,Y_m$ due campioni indipendenti con densità gaussiana $N(m,\sigma^2)$ (identica per entrambi) , e definiamo $\overline{X} = \frac{X_1+\cdots+X_n}{n}$ e $S^2(X) = \frac{\sum_{i\leq n}(X_i-\overline{X})^2}{n-1}$ ed in modo analogo $\overline{Y}$ e $S^2(Y)$.

a) Provare che, scelto comunque $0 < t < 1$, la v.a. $tS^2(X)+(1-t)S^2(Y)$ è una stima corretta della varianza σ^2.

b) Tra le stime sopra indicate, individuare quella di *rischio quadratico* minimo.

Esercizio 6.5. Consideriamo un campione $X_1,\dots,X_{17}$ di variabili gaussiane con media m sconosciuta e varianza 2 : supponiamo che la somma dei valori osservati $(x_1+\cdots+x_{17})$ sia eguale a 11,82. Qual è la soglia di accettazione per il test dell'ipotesi

$\mathcal{H}_0)\, m = 0$ contro l'alternativa $\mathcal{H}_1)\, m \neq 0$?