Variabili casuali di Poisson

VARIABILI CASUALI DI POISSON

$$X \in \mathbb{N} \qquad p(k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

$$X \sim P_o(\lambda) \qquad \lambda \in \mathbb{R}^+$$

$$\phi(t) = e^{\lambda(e^t - 1)}$$

$$E[X] = Var(X) = \lambda$$

In un certo senso, più grande è il valor medio, più grande è la varianza.

La poissoniana e la binomiale sono collegate, nel senso che una variabile casuale binomiale sotto certe condizioni nel limite per un numero di esperimenti molto grande, tendente a infinito, può convergere alla poissoniana. Riprendiamo questo passaggio.

Consideriamo $Y \sim B(n, p)$ con $n >> 1$

$$np = \lambda \Rightarrow p = \frac{\lambda}{n} << 1$$

$$Y \in \{0, 1, \dots n\}$$

$$P_Y(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Nulla di nuovo. Ricordiamoci che possiamo sviluppare il coefficiente binomiale, p alla k e 1 – p alla n – k possiamo riscriverli sfruttando la relazione evidenziata.

$$\frac{n!}{k!(n-k)!} \left(\frac{\lambda}{n}\right)^k \underline{\left(1 - \frac{\lambda}{n}\right)^{n-k}} =$$

Raccogliamo.

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

Semplifichiamo.

$$= \frac{\lambda^k}{k!} \frac{\overbrace{n(n-1)(n-2)\dots(n-k+1)\cancel{(n-k)!}}^{n!}}{n^k \cancel{(n-k)!}} \frac{\left(1 - \frac{\lambda}{n}\right)^n}{\left(1 - \frac{\lambda}{n}\right)^k}$$

Adesso supponiamo di mandare n a infinito, una binomiale con numero infinito di prove.

$$\lim_{n\to+\infty} p_X(k) = \lim_{n\to+\infty} \frac{\lambda^k}{k!} \frac{n\cdot(n-1)(n-2)\cdots(n-k+1)}{n^k} \frac{\left(1-\frac{\lambda}{n}\right)^n ⊛}{\left(1-\frac{\lambda}{n}\right)^k} =$$

$$= \frac{\lambda^k}{k!} \quad 1 \quad \frac{e^{-\lambda}}{1}$$

Riprendiamo il limite.

$$⊛ \lim_{n\to+\infty} \left(1-\frac{\lambda}{n}\right)^n = \lim_{n\to+\infty} e^{\ln\left(\left(1-\frac{\lambda}{n}\right)^n\right)} = \lim_{n\to+\infty} e^{n\ln\left(1-\frac{\lambda}{n}\right)} =$$

$$= \lim_{n\to+\infty} e^{n\left(\ln(1)-\frac{\lambda}{n}+\dots\right)} = e^{-\lambda}$$

Esercizi

es. 1

Il n° medio di incidenti (settimanali) in un tratto di autostrada è pari a 3, quale è la prob. che la prossima settimana si verifichi almeno un incidente?

In prima battuta assomiglia molto a un problema con la disugugalianza di Cebicev, si costruisce una variabile, la si ambienta in una data situazione e si fornisce il valore medio. Viene chiesto almeno un incidente quindi la probabilità di X maggiore o uguale di 1. Solo che in questo specifico caso otterremo che la probabilità è maggiore di 3, vero ma poco utile.

Come facciamo a riconoscere di poter usare la poissoniana? Abbiamo un tratto di autostrada, quindi alta utenza, un numero relativamente basso di incidenti, quindi è un evento raro rispetto al numero di utenti che passano per l'autostrada. Possiamo considerare X una poissoniana di parametro 3 (il valor medio della poissoniana è uguale al parametro).

$$X \sim Po(3)$$

Possiamo vedere la probabilità direttamente come la sommatoria oppure come l'evento complementare.

$$P(X\geq 1) = \sum_{k=1}^{+\infty} P(X=k) = \dots$$

oppure

$$P(X\geq 1) = 1 - P(X<1) = 1 - P(X=0)$$

La seconda strada è più semplice.

$$= 1 - \frac{3^0}{0!} e^{-3} = 1 - e^{-3}$$

Questo è un esempio in cui bisogna decidere se usare o meno la poissoniana.

Prendiamo un esempio limite, che viene più spontaneo usare una binomiale anziché una poissoniana.

es. 2
Una catena di produzione produce oggetti che presentano difetti con probabilità 0.1 indipendentemente dagli altri oggetti.
Si preleva un campione di 10 oggetti, qual è la prob. che il campione contenga al più un oggetto difettoso?

Abbiamo 10 oggetti, vediamoli come 10 esperimenti per scoprire se ognuno è difettoso o no, indipendenti tra loro per come sono fatti. Quindi viene in mente una binomiale con parametro 10 (oggetti) e 0.1 (probabilità di essere difettoso).

$$X = \text{'n° oggetti difettosi'} \sim B\left(10, \frac{1}{10}\right)$$

$$P(X \leq 1) = P(X=0) + P(X=1) =$$

$$= \binom{10}{0} \left(\frac{1}{10}\right)^0 \left(\frac{9}{10}\right)^{10} + \binom{10}{1} \frac{1}{10} \left(\frac{9}{10}\right)^9 \simeq 0,7361..$$

Cosa succederebbe se decidessimo di approssimare questa variabile casuale ad una poissoniana?

Pensiamo di approssimare X con $Y \sim Po(\lambda)$ anche se n NON È MOLTO GRANDE $(n=10)$ e $p = \frac{1}{10}$ NON È MOLTO PICCOLO

Non siamo nelle condizioni di farlo, però vedremo che l'approssimazione funziona.

$$\lambda = np = 10 \cdot \frac{1}{10} = 1$$

$$P(Y \leq 1) = P(Y=0) + P(Y=1) =$$

$$= \frac{1^0}{0!} e^{-1} + \frac{1^1}{1!} e^{-1} = 2e^{-1} \simeq 0.7358$$

Effettivamente la poissoniana approssima abbastanza bene la binomiale nonostante n non sia così grande e p così piccolo.

es. 3

In media arrivano ogni giorno 5 richieste di rimborso ad una compagnia di assicurazioni

Quale è la prob. che lunedì arrivino 3 richieste?

X = 'n° richieste in un giorno' $\sim Po(5)$

$$P(X=3) = \frac{5^3}{3!} e^{-5}$$

Un aspetto tipico in cui si utilizza per modellare il problema di una variabile casuale di Poisson è il numero di accessi in un ospedale. In base a questa stima viene costruita una certa ampiezza dei locali, senza essere esageratamente grandi o troppo piccoli. Quando succede qualcosa di particolare, succede che la struttura non è pronta al numero di accessi superiore alla variabile di Poisson per cui è stata programmata e lì iniziano i problemi.

Con quale probabilità in una settimana di 5 giorni lavorativi in esattamente 3 giorni arrivano 4 richieste?

Calcoliamo questa probabilità prima per il singolo giorno.

$$P(X=4) = \frac{5^4}{4!} e^{-5} \simeq 0.1755$$

Poi introduciamo una variabile casuale.

Y = 'n° di giorni con 4 richieste sui 5 lavorativi'

Sarà una binomiale, perché abbiamo 5 giorni e in 5 giorni osserviamo quel determinato fenomeno.

$$Y \sim B(5, 0.1755)$$

$$P(Y = 3) = \binom{5}{3} (0.1755)^3 (1 - 0.1755)^2$$

es. 4

Il numero di apparecchi difettosi prodotti in un giorno è una v.c. di Poisson di media 4, quale è la prob. che nell'arco di due giorni non vengano prodotti apparecchi difettosi?

$X_1 \sim Po(4)$ $\quad$ X_1 = 'n° apparecchi difettosi del giorno 1'

$X_2 \sim Po(4)$ $\quad$ X_2 = 'n° app. difettosi prodotti nel giorno 2'

$X_1 + X_2$ = 'n° complessivo di prodotti difettosi nelle due giornate'

Il numero di prodotti difettosi di un giorno è indipendente dal numero di prodotti difettosi di un altro, per ipotesi, anche perché altrimenti si sarebbe dovuto comunicare come dipende il numero di oggetti difettosi di una giornata dalla giornata successiva.

$X_1 + X_2 \sim Po(\lambda_1 + \lambda_2)$ $\quad$ per la riproducibilità delle v.c. di Poisson

$$X_1 + X_2 \sim Po(4 + 4)$$

$$P(X_1 + X_2 = 0) = \frac{8^0}{0!} e^{-8}$$

esercizio su coppie di v.c.

X e Y sono v.c. di Bernoulli:

$X \sim Be\left(\frac{1}{2}\right)$

$P(Y=1|X=0)=\varepsilon \qquad P(Y=0|X=1)=2\varepsilon \qquad \left(0<\varepsilon<\frac{1}{2}\right)$

1) $p(i,j)$? , $p_X(i)$, $p_Y(j)$, $E[X]$, $E[Y]$, $Cov(X,Y)$

2) $\exists\, \varepsilon : Cov(X,Y)=0$?

In base al testo:

$Y \backslash X$	0	1	
0			
1			
	$\frac{1}{2}$	$\frac{1}{2}$	

$$p(0,0)=P(X=0, Y=0)=$$
$$=P(Y=0|X=0)\,P(X=0)$$

$$p(1,0)=P(X=1 \cap Y=0)=$$
$$=P(Y=0|X=1)\,P(X=1)=$$
$$=2\varepsilon\frac{1}{2}=\varepsilon$$

$Y \backslash X$	0	1	
0		ε	
1	$\frac{\varepsilon}{2}$		
	$\frac{1}{2}$	$\frac{1}{2}$	

Gli altri vanno ricavati per differenza con la funzione marginale.

$Y \backslash X$	0	1	
0	$\frac{1}{2}-\frac{\varepsilon}{2}$	ε	
1	$\frac{\varepsilon}{2}$	$\frac{1}{2}-\varepsilon$	
	$\frac{1}{2}$	$\frac{1}{2}$	

Calcoliamo anche la funzione marginale della Y.

$Y \backslash X$	0	1	
0	$\frac{1}{2}-\frac{\varepsilon}{2}$	ε	$\frac{1}{2}+\frac{\varepsilon}{2}$
1	$\frac{\varepsilon}{2}$	$\frac{1}{2}-\varepsilon$	$\frac{1}{2}-\frac{\varepsilon}{2}$
	$\frac{1}{2}$	$\frac{1}{2}$	

Da qui, ricaviamo il parametro della bernoulliana che rappresenta Y.

$$Y \sim Be\left(\frac{1}{2}-\frac{\varepsilon}{2}\right)$$

$$X \sim Be\left(\frac{1}{2}\right) \qquad E[X] = p = \frac{1}{2}, \quad Var(X) = pq = \frac{1}{2}\cdot\frac{1}{2} = \frac{1}{4}$$

$$Y \sim Be\left(\frac{1}{2}-\frac{\varepsilon}{2}\right) \quad E[X] = p = \frac{1}{2}-\frac{\varepsilon}{2} \quad Var(x) = \left(\frac{1}{2}-\frac{\varepsilon}{2}\right)\left(\frac{1}{2}+\frac{\varepsilon}{2}\right) = \frac{1}{4}-\frac{\varepsilon^2}{4}$$

Adesso calcoliamo la covarianza.

$$Cov(X,Y) = E[XY] - E[X]E[Y]$$

$$E[XY] = \cancel{0\cdot 0\cdot\left(\frac{1}{2}-\frac{\varepsilon}{2}\right)} + \cancel{0\cdot 1\cdot\frac{\varepsilon}{2}} + \cancel{1\cdot 0\cdot\varepsilon} + 1\cdot 1\cdot\left(\frac{1}{2}-\varepsilon\right) = \frac{1}{2}-\varepsilon$$

$$\sum_{i=1}^{m}\sum_{j=1}^{n} x_i y_j\, p(x_i, y_j)$$

$$Cov(X,Y) = E[XY] - E[X]\,E[Y] = \frac{1}{2}-\varepsilon - \frac{1}{2}\left(\frac{1}{2}-\frac{\varepsilon}{2}\right) =$$

$$= \frac{1}{2} - \varepsilon - \frac{1}{4} + \frac{\varepsilon}{4} = \frac{1}{4} - \frac{3}{4}\varepsilon$$

$$Cov(X,Y) = 0 \text{ se } \frac{1}{4} - \frac{3}{4}\varepsilon = 0 \Rightarrow \varepsilon = \frac{1}{3}$$

per $\varepsilon = \frac{1}{3}$
X e Y sono
INDIP?

$Y \backslash X$	0	1	
0	$\frac{1}{2}-\frac{\varepsilon}{2}$	ε	$\frac{1}{2}+\frac{\varepsilon}{2}$
1	$\frac{\varepsilon}{2}$	$\frac{1}{2}-\varepsilon$	$\frac{1}{2}-\frac{\varepsilon}{2}$
	$\frac{1}{2}$	$\frac{1}{2}$	

Sostituiamo 1/3 nella tabella.

Poi si confronta il prodotto delle marginali con la funzione di massa congiunta su tutte le coppie ordinate.

$Y \backslash X$	0	1	
0	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{2}{3}$
1			

$$\left.\begin{aligned}
p(0,0) &= \tfrac{1}{3} \overset{si}{=} p_X(0)\, p_Y(0) = \tfrac{1}{2} \cdot \tfrac{2}{3} \\
p(0,1) &= \tfrac{1}{6} = p_X(0)\, p_Y(1) = \tfrac{1}{2} \cdot \tfrac{1}{3} \\
p(1,0) &= \tfrac{1}{3} = p_X(1)\, p_Y(0) = \tfrac{1}{2} \cdot \tfrac{2}{3} \\
p(1,1) &= \tfrac{1}{6} = p_X(1)\, p_Y(1) = \tfrac{1}{2}\, \tfrac{1}{3}
\end{aligned}\right\} \text{X e Y sono INDIPENDENTI}$$

es. $\left.\begin{array}{l} X \sim Po(\lambda_1) \\ Y \sim Po(\lambda_2) \end{array}\right\}$ INDIPENDENTI

$$P(X=k \mid X+Y=n) = ?$$

$k = 1, 2, \dots n$
$n \in \mathbb{N}$

Per definizione.

$$P(X=k \mid X+Y=n) = \frac{P(X=k \cap X+Y=n)}{P(X+Y=n)}$$

$$= \frac{P(X=k \cap Y=n-k)}{P(X+Y=n)}$$

$$\underset{INDIP.}{=} \frac{P(X=k)\, P(Y=n-k)}{P(X+Y=n)}$$

$X \sim Po(\lambda_1)$

$$P(X=k) = \frac{\lambda_1^k}{k!} e^{-\lambda_1}$$

$Y \sim Po(\lambda_2)$

$$P(Y=n-k) = \frac{\lambda_2^{n-k} e^{-\lambda_2}}{(n-k)!}$$

$X+Y \sim Po(\lambda_1+\lambda_2)$ per riproducibilità di Poissoniana

$$P(X+Y=n) = \underline{\frac{(\lambda_1+\lambda_2)^n}{n!} e^{-(\lambda_1+\lambda_2)}}$$

$$P(X=k \mid X+Y=n) = \frac{\frac{\lambda_1^k e^{-\lambda_1}}{k!} \frac{\lambda_2^{n-k} e^{-\lambda_2}}{(n-k)!}}{\frac{(\lambda_1+\lambda_2)^n}{n!} e^{-(\lambda_1+\lambda_2)}}$$

$$P(X=k \mid X+Y=n) = \frac{\frac{\lambda_1^k \cancel{e^{-\lambda_1}}}{k!} \frac{\lambda_2^{n-k} \cancel{e^{-\lambda_2}}}{(n-k)!}}{\frac{(\lambda_1+\lambda_2)^n}{n!} \cancel{e^{-(\lambda_1+\lambda_2)}}}$$

$$= \frac{n!}{k!(n-k)!} \frac{\lambda_1^k \lambda_2^{n-k}}{(\lambda_1+\lambda_2)^n} = \binom{n}{k} \left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^k \left(\frac{\lambda_2}{(\lambda_1+\lambda_2)}\right)^{n-k}$$

$$\boxed{\frac{\lambda_1}{\lambda_1+\lambda_2} = p, \quad q = \frac{\lambda_2}{\lambda_1+\lambda_2}} \qquad = \underline{\binom{n}{k} p^k q^{n-k}}$$

$0 < \frac{\lambda_1}{\lambda_1+\lambda_2} < 1$

funzione di massa di una binomiale $B\left(n, \frac{\lambda_1}{\lambda_1+\lambda_2}\right)$

Modellazione di variabili casuali continue

La prima è stata già incontrata.

MODELLI DI V.C. CONTINUE

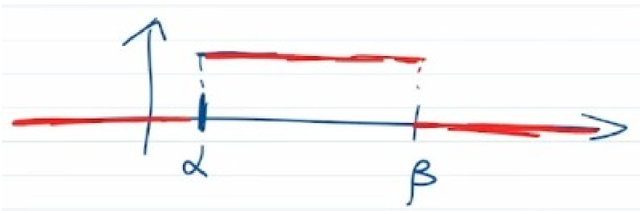
1) VARIABILE CASUALE UNIFORME

Ci abbiamo a che fare quando abbiamo a che fare con una variabile casuale definita in un intervallo che può assumere qualunque valore di quell'intervallo con la stessa probabilità. Non è totalmente corretto (perché la probabilità per una variabile casuale continua è sempre a 0), ma la funzione di densità di questa variabile è sempre costante, non c'è un valore che spicca su altri.

$X \sim U(\alpha, \beta)$ con $\alpha < \beta$ e $\alpha, \beta \in \mathbb{R}$

$$f(x) = \begin{cases} \frac{1}{\beta - \alpha} & \text{se } x \in [\alpha, \beta] \\ 0 & \text{altrove} \end{cases}$$

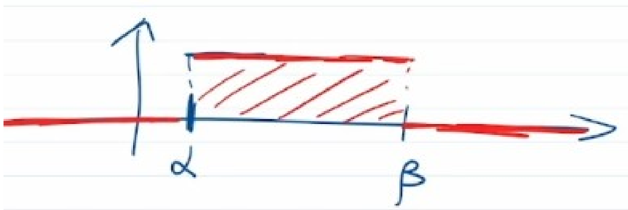
Immaginiamo di disegnare la funzione di densità.



ovviamente

$$1 = \int_{-\infty}^{+\infty} f(x)\,dx = \int_{\alpha}^{\beta} \frac{1}{\beta - \alpha}\,dx = \frac{\beta - \alpha}{\cancel{\beta - \alpha}}$$

L'integrale si può fare a occhio.



È un rettangolo che ha una base lunga beta – alfa e un'altezza uguale a 1 / (beta – alfa) affinché l'area sia unitaria.

Il valor medio di questa variabile sarà il punto medio dell'intervallo ragionevolmente. La varianza misura quanto i dati sono sparpagliati rispetto al valor medio, quindi dipenderà dalla lunghezza dell'intervallo.

$$E[X] = \int_{-\infty}^{+\infty} x f(x)\,dx = \cancel{\int_{-\infty}^{\alpha} 0 x\,dx} + \int_{\alpha}^{\beta} \frac{x}{\beta-\alpha}\,dx + \cancel{\int_{\beta}^{+\infty} 0 . x\,dx} =$$

$$= \frac{1}{\beta-\alpha}\left[\frac{x^2}{2}\right]_{\alpha}^{\beta} = \frac{1}{2}\frac{(\beta^2-\alpha^2)}{\beta-\alpha} = \frac{1}{2}(\alpha+\beta)$$ punto medio dell'intervallo

Facile da ricordare.

dell'intervallo

$$E[X^2] = \int_{-\infty}^{+\infty} x^2 f(x)\,dx = \cancel{\int_{-\infty}^{\alpha} x^2 . 0\,dx} + \int_{\alpha}^{\beta} \frac{x^2}{\beta-\alpha}\,dx + \cancel{\int_{\beta}^{+\infty} x^2 . 0\,dx} =$$

$$= \frac{1}{\beta-\alpha}\left[\frac{x^3}{3}\right]_{\alpha}^{\beta} = \frac{1}{3}\,\frac{\beta^3-\alpha^3}{\beta-\alpha} =$$

$$= \frac{1}{3}\left(\beta^2+\alpha^2+\alpha\beta\right)$$

$$Var(X) = E[X^2] - E^2[X] = \frac{1}{3}\left(\beta^2+\alpha^2+\alpha\beta\right) - \left(\frac{\alpha+\beta}{2}\right)^2 =$$

$$= \frac{4\beta^2+4\alpha^2+4\alpha\beta-3\alpha^2-3\beta^2-6\alpha\beta}{12} = \frac{\alpha^2+\beta^2-2\alpha\beta}{12} =$$

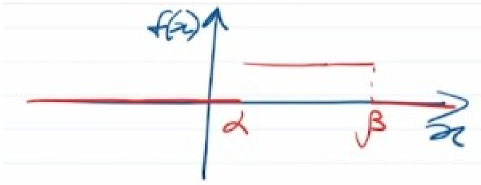
$$= \frac{(\beta-\alpha)^2}{12}$$

In un qualche modo sapevamo che la varianza fosse legata all'ampiezza dell'intervallo, in particolare sappiamo che la varianza ha questo ruolo di quadrato rispetto alla variabile casuale e anche questo ruolo viene rispettato.

Calcoliamo la funzione di ripartizione. La funzione di densità:

$$f(x) = \begin{cases} \dfrac{1}{\beta-\alpha} & \text{se } \alpha \le x \le \beta \\ 0 & \text{altrove} \end{cases}$$

Riprendiamo il grafico.

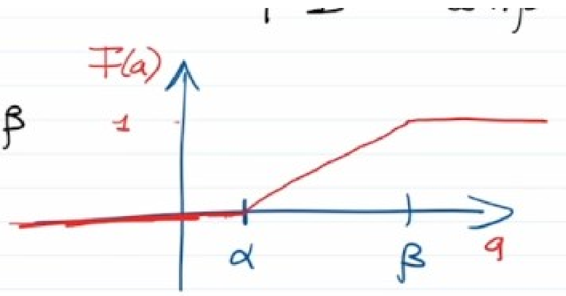


Se consideriamo un valore minore di alfa, la probabilità è 0 perché non c'è nessun valore della variabile casuale. Viceversa, considerando un valore più grande di beta, la probabilità è una certezza (1). Nella situazione intermedia, la probabilità è l'integrale della funzione di densità.

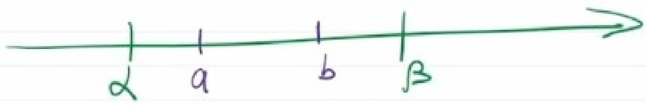
$$F(a) = P(X \le a) = \begin{cases} 0 & \text{se } a < \alpha \\ \cancel{\displaystyle\int_{-\infty}^{\alpha} 0\,dx} + \displaystyle\int_{\alpha}^{a} \frac{1}{\beta-\alpha}\,dx = \frac{a-\alpha}{\beta-\alpha} & \text{se } \alpha \le a \le \beta \\ 1 & \text{se } a > \beta \end{cases}$$

Quindi abbiamo una funzione continua con due tratti costanti e un tratto lineare.

$$F(a) = \begin{cases} 0 & \text{se } a < \alpha \\ \dfrac{a-\alpha}{\beta-\alpha} & \text{se } \alpha \le a \le \beta \\ 1 & \text{se } a > \beta \end{cases}$$



Dati: a e $b \in \mathbb{R}$ con $a < b$ e
$\alpha < a < b < \beta$



$$P(X \in [a,b]) = \frac{b-a}{\beta-\alpha} \quad \text{infatti:}$$

Ci sono due modi per dimostrarlo.

$$P(X\in[a,b]) = \int_a^b f(x)\,dx = \int_a^b \frac{1}{\beta-\alpha}\,dx = \frac{1}{\beta-\alpha}\Big[x\Big]_a^b = \frac{b-a}{\beta-\alpha}$$

È un rapporto di lunghezza degli intervalli (sottointervallo e intervallo di definizione).

oppure questa proprietà si può dimostrare

$$P(a\le X\le b) = P(X\le b) - P(X<a)$$

Non escludere a perché compreso nella disuguaglianza iniziale.

$$\underset{\text{v.c. continua}}{=} P(X\le b) - P(X\le a) =$$

Aggiungere o togliere un singolo valore nella variabile casuale continua non cambia le probabilità.

$$= F(b) - F(a) = \frac{b-\alpha}{\beta-\alpha} - \frac{a-\alpha}{\beta-\alpha} = \frac{b-a}{\beta-\alpha}$$

Riepilogo del calcolo della funzione di ripartizione

RIEPILOGO DEL CALCOLO DI $F(a)$

$$F(a) = P(X\le a) \underset{\text{def. di v.c. continua}}{=} \int_{x\le a} f(x)\,dx = \int_{-\infty}^{a} f(x)\,dx$$

Devo separare le situazioni diverse che possono capitare.

$$= \begin{cases} ① & \text{se } a<\alpha \\ ② & \text{se } \alpha\le a\le\beta \\ ③ & \text{se } a>\beta \end{cases}$$

Vediamo tutte le casistiche.

$$① = \int_{-\infty}^{a} 0 \, dx$$

$$② = \cancel{\int_{-\infty}^{\alpha} 0 \, dx} + \int_{\alpha}^{a} \frac{1}{\beta-\alpha} \, dx = \frac{[x]_{\alpha}^{a}}{\beta-\alpha} = \frac{a-\alpha}{\beta-\alpha}$$

La funzione tra -infinito e alfa è 0, quindi da contributo nullo.

$$③ = \cancel{\int_{-\infty}^{\alpha} 0 \, dx} + \int_{\alpha}^{\beta} \frac{1}{\beta-\alpha} \, dx + \cancel{\int_{\beta}^{a} 0 \, dx} = \frac{[x]_{\alpha}^{\beta}}{\beta-\alpha} = \frac{\beta-\alpha}{\beta-\alpha} = 1$$

Ho verificato che la variabile casuale binomiale fosse riproducibile guardando la funzione generatrice dei momenti della X e della Y, avendo p in comune. Moltiplicandole insieme mi accorgo che hanno la stessa struttura della funzione generatrice di una binomiale. Ma dal punto di vista del senso perché la binomiale è riproducibile? Se io considero una prima variabile casuale binomiale che conta il numero di successi in N prove e poi considero un altro set di esperimenti di M prove e conto il numero di successi, posso anche immaginare di fare N + M prove e contare il numero complessivo di successi, è una binomiale più grande che abbraccia i primi N esperimenti più gli altri M esperimenti. Diciamo che la riproducibilità della binomiale è non ovvia, ma sensata. Nelle variabili casuali poissoniane non è così ovvio che sia riproducibile, è ereditata dalla binomiale, perché è un limite particolare della binomiale.

C'è un'altra variabile casuale introdotta per approssimare la binomiale, la gaussiana. All'inizio si pensava fosse solo una tecnica per velocizzare i calcoli, poi si è scoperto che la proprietà di approssimare la binomiale è qualcosa di più generale. La gaussiana è una variabile casuale continua, riproducibile anch'essa perché eredita questa proprietà dalla binomiale. Le poissoniane, le binomiali e le guassiane sono in un qualche modo collegate.

Un'altra variabile casuale riproducibile è la binomiale negativa perché è una somma di geometriche, non occorre dimostrare la riproducibilità, a patto che il parametro p sia uguale alla somma delle due variabili.

Esercizi

es. 9 FOGLIO VI

Sia X una variabile casuale con valor medio 5 e varianza 4. Determinare il più piccolo intervallo in cui cade almeno il 75% delle osservazioni, usare la disuguaglianza di Cebicev.

$$P(x_1 < X < x_2) \geq 75\%$$

Conoscendo solo valor medio, varianza e usando Cebicev si può riscrivere così:

$$P(-\delta < X - E[X] < \delta) = P(|X - E[X]| < \delta) =$$

Calcoliamo la probabilità col valore complementare.

$$= 1 - \underline{P(|X - E[X]| \geq \delta)}$$

La disuguaglianza evidenziata rientra nella disuguaglianza di Cebicev.

DISUG. DI ČEBYČEV

$$P(|X - E[X]| \geq \delta) \leq \frac{Var(X)}{\delta^2}$$

$$= 1 - \underline{P(|X - E[X]| \geq \delta)} \geq 1 - \frac{Var(X)}{\delta^2} = 1 - \frac{4}{\delta^2} = 75\%$$

$$1 - \frac{4}{\delta^2} = \frac{3}{4} \Rightarrow \frac{4}{\delta^2} = 1 - \frac{3}{4} \Rightarrow \frac{4}{\delta^2} = \frac{1}{4} \Rightarrow \delta^2 = 16$$

Vuol dire che:

$$|X - 5| < 4 \Rightarrow -4 < X - 5 < 4$$

$$\underline{1 < X < 9}$$

es. 6 FOGLIO VI

Una moneta equilibrata viene lanciata un certo numero N di volte. Si considerano gli eventi:

A = "Esce testa al più una volta"

B = "Testa e croce escono almeno una volta"

Si chiede la probabilità di A, di B, dell'intersezione e della probabilità di B condizionato da A.

es. 6 FOGLIO VI

A = 'esce T al più una volta'

B = 'escono T e C almeno una volta'

n lanci di moneta equilibrata

$X =$ 'n° T su n lanci' $\Rightarrow X \sim B\left(n, \frac{1}{2}\right)$

$$P(A) = P(X \leq 1) = P(X=0) + P(X=1) =$$

$$= \binom{n}{0}\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^n + \binom{n}{1}\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^{n-1} =$$

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

$$P(B) = P(1 \leq X \leq n-1) = \sum_{k=1}^{n-1}\binom{n}{k}\left(\frac{1}{2}\right)^k\left(\frac{1}{2}\right)^{n-k}$$

oppure $P(B) = 1 - P(X=0) - P(X=n) =$

$$= 1 - \binom{n}{0}\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^n - \binom{n}{n}\left(\frac{1}{2}\right)^n\left(\frac{1}{2}\right)^0$$

$$= 1 - \frac{1}{2^n} - \frac{1}{2^n} = 1 - \frac{2}{2^n} =$$

$$= 1 - \frac{1}{2^{n-1}} = \frac{2^{n-1}-1}{2^{n-1}}$$

$$P(A \cap B) = P(\text{'al più una T'} \cap \text{'almeno 1 T e 1 C'}) =$$

$$= P(X \leq 1 \cap 1 \leq X \leq n-1) = P(X=1) =$$

$$= \binom{n}{1}\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^{n-1} = \frac{n}{2^n}$$

$$P(B|A) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{n}{2^n}}{\frac{2^{n-1}-1}{2^{n-1}}} = \frac{n}{2(2^{n-1}-1)} = \frac{n}{2^n-2}$$

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{\frac{n}{2^n}}{\frac{n+1}{2^n}} = \frac{n}{n+1}$$