Variabile casuale lognormale

APPLICAZIONE DELLA TEORIA DELLE FUNZIONI DI UNA V.C.: VARIABILE CASUALE LOGNORMALE

$$W \sim N(\mu, \sigma^2) \qquad X = e^{W} \qquad X \sim \text{Lognormal}(\mu, \sigma^2)$$

$$F_X(a) = P(X \le a) = \begin{cases} 0 & \text{se } a \le 0 \\ F_Z\left(\frac{\ln(a)-\mu}{\sigma}\right) & \text{se } a > 0 \end{cases}$$

È una variabile casuale positiva (non può assumere valori negativi), è l'esponenziale di una variabile casuale gaussiana.

$$P(X \le a) = P(e^{W} \le a) = P(\underline{W \le \ln(a)}) =$$

$Z \sim N(0,1)$

$$= P\left(\frac{W-\mu}{\sigma} \le \frac{\ln(a)-\mu}{\sigma}\right) = P\left(Z \le \frac{\ln(a)-\mu}{\sigma}\right) = F_Z\left(\frac{\ln(a)-\mu}{\sigma}\right)$$

Dalla lognormale, possiamo ricondurci alla gaussiana W, per poi passare alla normale standard tramite la normalizzazione.

$$f_X(a) = \frac{d F_X(a)}{da} = \begin{cases} 0 & \text{se } a \le 0 \\ \frac{d F_Z\left(\frac{\ln(a)-\mu}{\sigma}\right)}{da} & \text{se } a > 0 \end{cases}$$

$$\frac{d F_Z\left(\frac{\ln(a)-\mu}{\sigma}\right)}{da} = \frac{d}{da}\left(\int_{-\infty}^{\frac{\ln(a)-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx\right) =$$

$$= \frac{1}{\sqrt{2\pi}} e^{-\left(\frac{\ln(a)-\mu}{\sigma}\right)^2 \frac{1}{2}} \frac{1}{\sigma a} = \frac{1}{a\sigma\sqrt{2\pi}} e^{-\frac{(\ln(a)-\mu)^2}{2\sigma^2}}$$

$$\Downarrow$$

$$f_X(a) = \begin{cases} 0 & \text{se } a \leq 0 \\ \frac{1}{a\sigma\sqrt{2\pi}} e^{-\frac{(\ln(a)-\mu)^2}{2\sigma^2}} & \text{se } a > 0 \end{cases}$$

$$E[X] = E[e^W] = \phi_W(1) = \left(e^{t\mu + \frac{t^2}{2}\sigma^2}\right)\Big|_{t=1} = e^{\mu + \frac{\sigma^2}{2}}$$

$$E[X^2] = E[e^{2W}] = \phi_W(2) = \left(e^{t\mu + \frac{t^2}{2}\sigma^2}\right)\Big|_{t=2} = e^{2\mu + 2\sigma^2}$$

$$Var(X) = E[X^2] - E^2[X] = e^{2\mu + 2\sigma^2} - e^{2\mu + \sigma^2} =$$

$$= e^{2\mu + \sigma^2}\left(e^{\sigma^2} - 1\right)$$

$$Y = e^{W} \longrightarrow W = W_1 + W_2 \sim N(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$$

$$Y \sim \text{lognormal}(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$$

La proprietà si può generalizzare al prodotto di n v.c. lognormali indipendenti;

$$Y = e^{W_1} e^{W_2} = e^{(W_1+W_2)} \qquad \text{INDIP} \begin{cases} W_1 \sim N(\mu_1, \sigma_1^2) \\ W_2 \sim N(\mu_2, \sigma_2^2) \end{cases}$$

$$\Downarrow \qquad\qquad\qquad \Downarrow$$

$$Y = e^{W} \longrightarrow W = W_1 + W_2 \sim N(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$$

$$Y \sim \text{lognormal}(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$$

La proprietà si può generalizzare al prodotto di n v.c. lognormali indipendenti;

Funzioni di coppie di variabili casuali

FUNZIONI DI COPPIE DI V.C.

(X, Y) coppie di v.c. note

$$Z = h(X, Y)$$

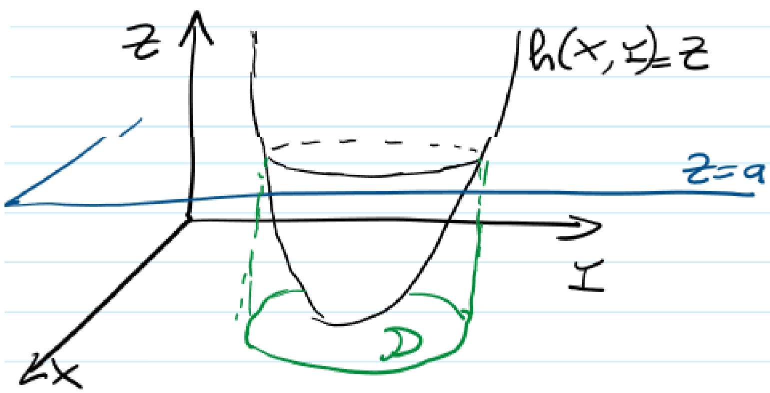
$$E[Z^n] = E[h^n(X,Y)] = \begin{cases} \displaystyle\sum_{i=1}^{m}\sum_{j=1}^{n} h^n(x_i, y_j)\, p(x_i, y_j) & \text{se } (X,Y) \text{ v.c. discrete con } X \in \{x_1, \dots x_m\} \text{ e } Y \in \{y_1, \dots y_n\} \\ \displaystyle\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} h^n(x,y) f(x,y)\, dx\, dy & \text{se } (X,Y) \text{ coppia di v.c. congiuntamente continue} \end{cases}$$

con $n \in \mathbb{N} \setminus \{0\}$

$$Var(Z) = E[Z^2] - E^2[Z]$$

$$F_Z(a) = P(Z \leq a) = P(h(X,Y) \leq a) = P((X,Y) \in D) =$$

$$= \iint_D f(x,y)\, dx\, dy$$



CASO PARTICOLARE: $Z = X + Y$

$$Z = h(X,Y) = X + Y$$

$$F_Z(a) = P(X+Y \le a) = P(Y \le -X+a) = $$
$$= \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{-x+a} f(x,y)\, dy \right) dx$$

oppure (equivalentemente)

$$F_Z(a) = P(X+Y \le a) = P(X \le -Y+a) =$$
$$= \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{-y+a} f(x,y)\, dx \right) dy$$

$$f_Z(a) = \frac{d F_Z(a)}{d a} = \frac{d}{da} \left(\int_{-\infty}^{+\infty} \left(\int_{-\infty}^{-x+a} f(x,y)\, dy \right) dx \right) =$$
$$= \int_{-\infty}^{+\infty} \frac{d}{da} \left(\int_{-\infty}^{-x+a} f(x,y)\, dy \right) dx = \int_{-\infty}^{+\infty} f(x, a-x) \cdot 1 \, dx =$$

si dimostra che vale anche
$$\int_{-\infty}^{+\infty} f(a-y, y)\, dy$$

Se X e Y sono INDIPENDENTI

$$f(x,y) = f_X(x)\, f_Y(y)$$

$$f_Z(a) = \int_{-\infty}^{+\infty} f_X(x)\, f_Y(a-x)\, dx$$

oppure $$= \int_{-\infty}^{+\infty} f_X(a-y)\, f_Y(y)\, dy$$

PRODOTTO DI CONVOLUZIONE

es. 1 $X \sim E(1)$, $Y \sim E(1)$

X e Y INDIPENDENTI

$$Z = X + Y$$

$$\boxed{\begin{array}{l} f_X(x) = \begin{cases} e^{-x} & \text{se } x \geq 0 \\ 0 & \text{altrove} \end{cases} \\ f_Y(y) = \begin{cases} e^{-y} & \text{se } y \geq 0 \\ 0 & \text{altrove} \end{cases} \end{array}}$$

$$f_Z(a) = \int_{-\infty}^{+\infty} f_X(\underline{x})\, f_Y(\underline{a-x})\, dx = \int_0^a e^{-x} e^{-(a-x)}\, dx = *$$

$f(x, a-x) = f_X(x)\, f_Y(a-x) \neq 0$ se

$$\begin{cases} x \geq 0 \\ a - x \geq 0 \end{cases} \qquad \begin{cases} x \geq 0 \\ x \leq a \end{cases} \Rightarrow a \geq 0$$

$$* = \int_0^a e^{-(\cancel{x}+a-\cancel{x})}\, dx = \left[e^{-a} x\right]_0^a = a\, e^{-a}$$

$$f_Z(a) = \begin{cases} a\, e^{-a} & \text{se } a \geq 0 \\ 0 & \text{altrove} \end{cases}$$

$$F_Z(a) = P(Z \leq a) = \int_{-\infty}^{a} f_Z(z)\, dz = \begin{cases} \cancel{\int_{-\infty}^{0} 0\, dz} + \int_{0}^{a} z\, e^{-z}\, dz & \text{se } a \geq 0 \\ 0 & \text{se } a < 0 \end{cases}$$

$$\int_0^a z\, e^{-z}\, dz = \left[z\left(-e^{-z}\right) \right]_0^a + \int_0^a e^{-z}\, dz = -a\, e^{-a} + \left[-e^{-z} \right]_0^a =$$

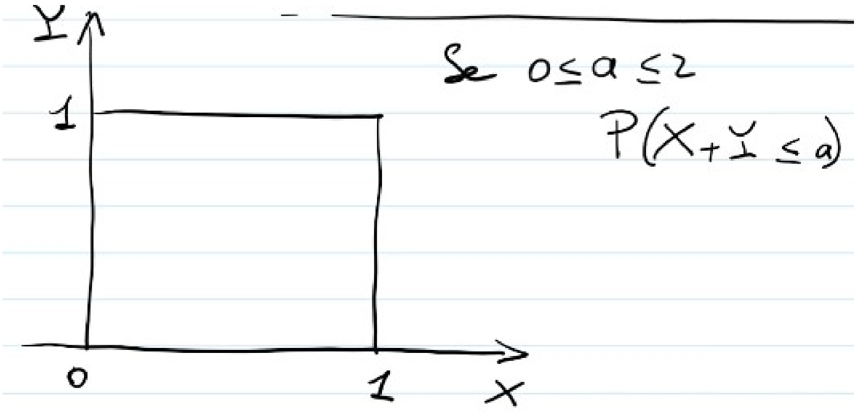
$$= 1 - a\, e^{-a} - e^{-a} = 1 - (a+1)\, e^{-a}$$

es. 2 $\underbrace{X \sim U(0,1), \quad Y \sim U(0,1)}_{\text{INDIP.}}$

$$Z = X + Y \qquad F_Z(a)? \quad f_Z(a)?$$

$$F_Z(a) = P(X + Y \leq a) = \begin{cases} 0 & \text{se } a < 0 \\ ? & \\ \cdot & \\ 1 & \text{se } a > 2 \end{cases}$$

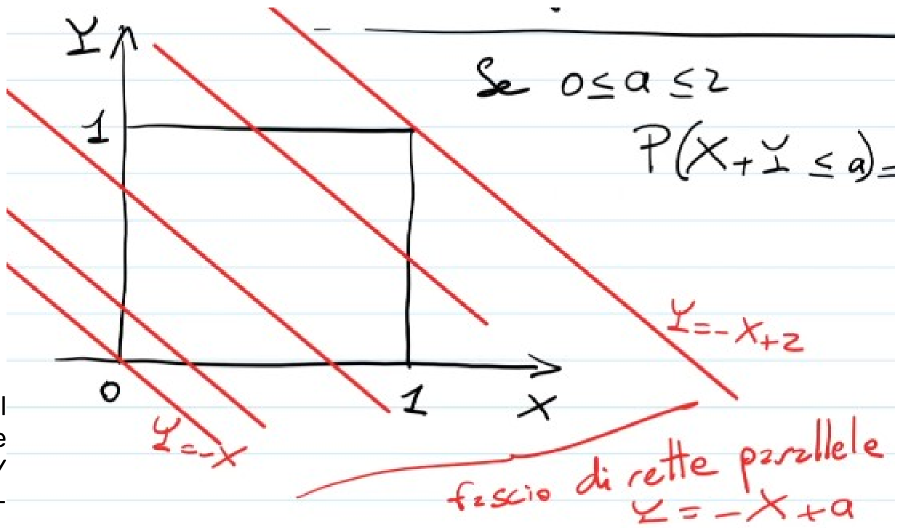
Dobbiamo stabilire come si comporta la funzione tra 0 e 2. Usiamo il metodo grafico.



Questa probabilità, se ha senso, si può scrivere:

$$=P(Y \le -X+a)$$

Già fatto nel caso generale. Possiamo disegnare il fascio di rette dell'equazione sopra riportata, parallele alle bisettrici del secondo e quarto quadrante.

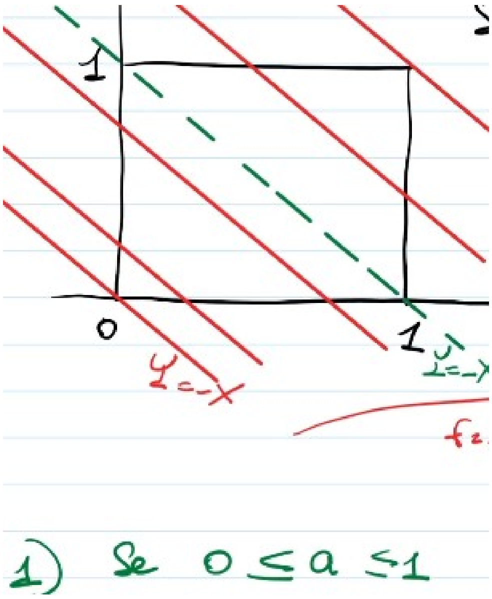


Ci interessa il fascio di rette compreso tra Y = -X e Y = -X + 2, i casi estremi.

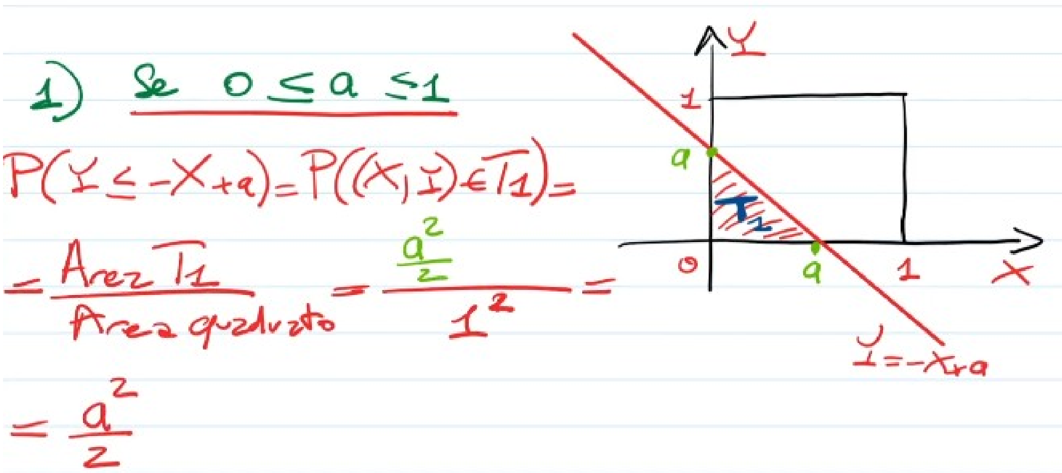
Guardando la consegna, la regione che la soddisfa sta sotto la retta Y = -X + a.

$$\boxed{\begin{array}{l} Y \le -X+a \\ \text{semipiano "sotto"} \\ \text{la retta} \end{array}}$$

Così ci si accorge che l'intersezione tra il semipiano sotto la retta e il quadrato di definizione della X e della Y assume forme diverse a seconda che la retta del fascio considerato passi per la diagonale del quadrato, sotto (triangolo) o sopra (pentagono). Per questo motivo, separiamo il caso triangolo dal caso pentagono.



Sotto questa retta è un triangolo.

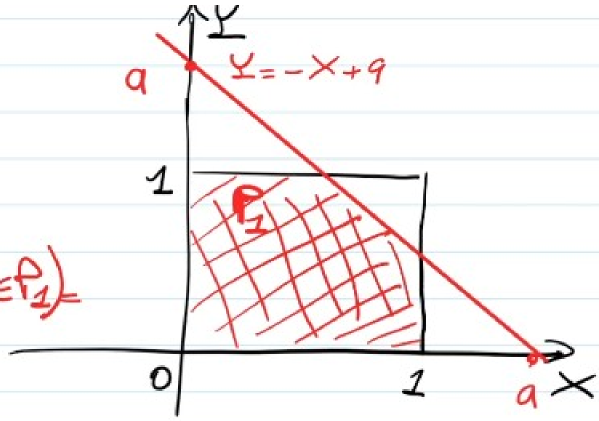


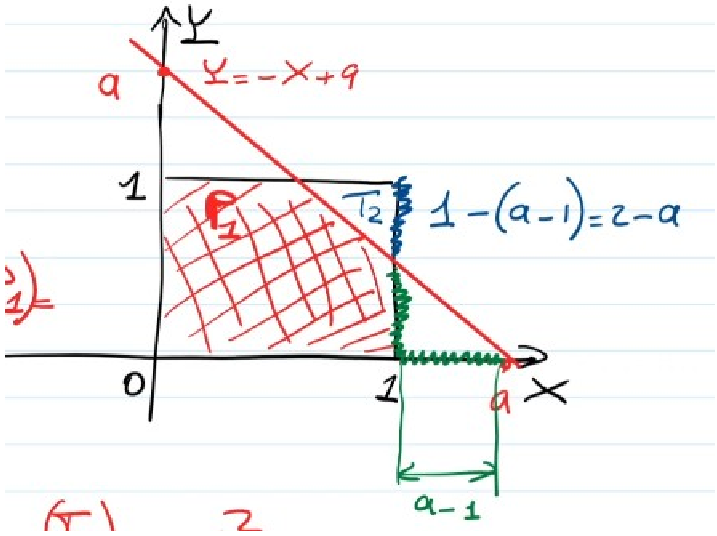
Nel secondo caso, trattiamo un pentagono.

$= \frac{a}{2}$

2) Se $1 < a \leq 2$

$$P(Y \leq -X+a) = P((X,Y) \in P_1) = \frac{\text{Area}(P_1)}{\text{Area quadrato}}$$



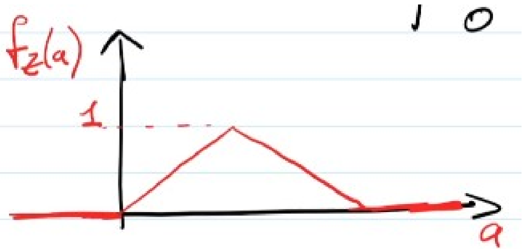


In base a questo, consideriamo l'area del pentagono come la differenza tra l'area del quadrato e l'area del triangolo.

$$= \frac{\text{Area quadrato} - \text{Area}(T_2)}{\text{Area quadrato}} = \frac{1^2 - \frac{(2-a)^2}{2}}{1^2} = 1 - \frac{(2-a)^2}{2}$$

$$F_Z(a) = \begin{cases} 0 & \text{se } a < 0 \\ \frac{a^2}{2} & \text{se } 0 \leq a \leq 1 \\ 1 - \frac{(2-a)^2}{2} & \text{se } 1 < a \leq 2 \\ 1 & \text{se } a > 2 \end{cases}$$

$$f_Z(a) = \frac{dF_Z(a)}{da} = \begin{cases} 0 & \text{se } a < 0 \\ a & \text{se } 0 \leq a \leq 1 \\ 2-a & \text{se } 1 < a \leq 2 \\ 0 & \text{se } a > 2 \end{cases}$$



È una funzione tenda. Sommando due variabili casuali uniformi non ottengo una variabile casuale uniforme. C'è un picco in 1, quindi questo dice che il range di valori probabili sta attorno al valore centrale (che è anche il valor medio). Continuando a sommare variabili uniformi, si ottiene sempre più una funzione di densità simile alla gaussiana, grazie al teorema del limite centrale.

Teorema del limite centrale

Ha avuto prima un nome straniero, poi un nome italiano. La parola centrale si riferisce a teorema, non a limite. È un teorema importante della probabilità, ha una valenza centrale.

TEOREMA DEL LIMITE CENTRALE
O CENTRALE DEL LIMITE

$X_1, X_2 \dots X_n$ V.C. INDIPENDENTI E IDENTICAMENTE DISTRIBUITE

$$\mu = E[X_k], \quad \sigma^2 = Var(X_k) \qquad k = 1, \dots n$$

La prima idea è fare la somma di tutte le variabili casuali.

$$Y = X_1 + X_2 \dots + X_n$$

È un teorema limite, n va mandato a infinito, ma c'è un problema: media e varianza tendono a infinito. Allora si prende la standardizzazione di questa somma.

$$Y_n = \frac{X_1 + X_2 \dots + X_n - n\mu}{\sqrt{n\sigma^2}}$$

In questa maniera, la variabile casuale ha valor medio nullo e varianza 1.

$$\forall a \in \mathbb{R} \quad \lim_{n \to +\infty} F_{Y_n}(a) = F_Z(a)$$

La sua funzione di ripartizione diventa la funzione di ripartizione della normale standard.

ovvero

$$\lim_{n} \Phi(t) = \Phi(t)$$

CONVERGENZA IN DISTRIBUZIONE ALLA NORMALE STANDARD

La convergenza in distribuzione si dimostra tramite la funzione generatrice dei momenti. È un risultato sorprendente, viene applicato in un'approssimazione dove sommo tante variabili casuali, ma non infinite. Quante ne servono? Almeno 30. Dalle 30, in poi la variabile viene considerata gaussiana,

OSS. Se $n \geq 30$ $Y_n \mathrel{\dot\sim} N(0,1)$

"si comporta "circa" come una normale standard"

Se n è finito, anche la media e la varianza saranno finite. Nel caso in cui non si faccia il limite tendente a infinito, possiamo dedurre che:

$$\Downarrow$$

$$S_n = X_1 + X_2 \ldots + X_n \mathrel{\dot\sim} N(n\mu, n\sigma^2)$$

Questo è vero solo con n finito.

In questa ottica rientra la binomiale, perché è una somma di bernoulliane.

APPLICAZIONI

1) $X_1, X_2 \dots X_n \sim Be(p)$

$S_n \sim B(n, p)$ Se $n \geq 30$ $S_n \overset{\circ}{\sim} N(np, npq)$

Storicamente, è stato scoperto prima questo, poi tutto il resto.

Come posso rappresentare una variabile casuale discreta con una variabile casuale continua? Ci sarà una tecnica per rendere l'approssimazione più elegante e precisa nel passaggio da discrete a continue.

2) $\overline{X} = \dfrac{X_1 + X_2 \dots + X_n}{n}$ Se $n \geq 30$

$\overline{X} \overset{\circ}{\sim} N\left(\mu, \dfrac{\sigma^2}{n}\right)$