Modelli

Variabile aleatoria di Bernoulli

Definizione

Astrazione del lancio di una moneta
$X \sim \mathrm{B}(p)$ assume valori $\{0, 1\}$

$$P_X(x) = \begin{cases} p, & x = 1 \\ 1 - p = q, & x = 0 \end{cases}$$

con $p \in (0, 1)$

Formule

$\Phi_X(t) = q + e^t p$

Dimostrazione

$\ldots = \mathbb{E}[e^{tX}] = \sum_{i=0}^{1} e^{t \cdot i} \cdot P_X(x_i) = \sum_{i=0}^{1} e^{t \cdot i} \cdot P_X(i) = e^{0t}(1-p) + e^{1t} p = \ldots$

$\mathbb{E}[X] = p$

Dimostrazione

$\ldots = \frac{d\Phi_X(t)}{dt}\Big|_{t=0} = e^t p\big|_{t=0} = \ldots$
oppure
$\cdots = \sum_{i=0}^{1} x_i \cdot P_X(x_i) = \sum_{i=0}^{1} i \cdot P_X(i) = 0 \cdot q + 1 \cdot p = \ldots$

$\mathbb{E}[X^2] = p$

Dimostrazione

$\ldots = \frac{d^2\Phi_X(t)}{dt^2}\Big|_{t=0} = \frac{d}{dt} e^t p\big|_{t=0} = e^t p\big|_{t=0} = \ldots$
oppure
$\cdots = \sum_{i=0}^{1} x_i^2 \cdot P_X(x_i) = \sum_{i=0}^{1} i^2 \cdot P_X(i) = 0^2 \cdot q + 1^2 \cdot p = \ldots$

$\mathrm{Var}(X) = p(1-p) = p \cdot q$

Dimostrazione

$\ldots = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = p - p^2 = \ldots$

Variabile aleatoria binomiale

Definizione

Esperimenti ripetuti e indipendenti con due possibili esiti con probabilità p e $1-p$
Abbinando 1 al successo e 0 all'insuccesso, ripetendo n volte l'esperimento, la variabile che conta il numero di successi è detta binomiale ed è la somma di n variabili aleatorie Bernoulliane stocasticamente indipendenti
$X_1, \ldots, X_n \sim \mathrm{B}(p),\ X_i \perp\!\!\!\perp X_j \ \ i \neq j$
$Y = \sum_{i=1}^n X_i \sim \mathrm{Bin}(n, p)$

$$P_Y(y) = \binom{n}{y} (p)^y (1-p)^{n-y} \ \ y \in \{0, \ldots, n\}$$

Formule

$\Phi_Y(t) = (e^t p + q)^n$

Dimostrazione ›

$$\ldots = \mathbb{E}[e^{tY}] = \mathbb{E}\left[e^{t\sum_{i=1}^n X_i}\right] = \mathbb{E}\left[\prod_{i=1}^n e^{tX_i}\right] = \prod_{i=1}^n \mathbb{E}[e^{tX_i}] = \ldots$$

$\mathbb{E}[Y] = n \cdot p$

Dimostrazione ›

$\ldots = \mathbb{E}\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n \mathbb{E}[X_i] = \ldots$

$\mathrm{Var}(Y) = n \cdot p \cdot (1-p) = n \cdot p \cdot q$

Approfondimento ›

$Y_1 \sim \mathrm{Bin}(n_1, p),\ Y_2 \sim \mathrm{Bin}(n_2, p),\ Y_1 \perp\!\!\!\perp Y_2,\ Z = Y_1 + Y_2$
$Z \sim \mathrm{Bin}(n_1 + n_2, p)$
$\Phi_Z(t) = \mathbb{E}[e^{tZ}] = \mathbb{E}[e^{tY_1 + tY_2}] = \mathbb{E}[e^{tY_1} e^{tY_2}] = \mathbb{E}[e^{tY_1}] \cdot \mathbb{E}[e^{tY_2}] = (e^t p + q)^{n_1} (e^t p + q) = (e^t p + q)^{n_1 + n_2}$

Esempio ›

> # Variabile aleatoria binomiale
>
> $Y \sim \mathrm{Bin}\left(20, \frac{1}{3}\right)$ conta il numero di teste su 20 lanci
> La media di teste è $\mathbb{E}[Y] \approx 6.667$

La probabilità che esca testa esattamente 3 volte è $P_Y(3) \approx 0.0429$
La probabilità che esca testa tra 4 e 6 volte è $\sum_{y=4}^{6} P_Y(y) \approx 0.4189$

Variabile aleatoria geometrica

Definizione

Esperimenti indipendenti ripetuti con probabilità di successo p costante
Conta il numero di ripetizioni dell'esperimento fino al primo insuccesso
$X \sim \text{Geo}(p)$
$P_X(x) = p^x \cdot (1-p) \quad x \in \mathbb{N}$

$$\sum_{i=0}^{+\infty} P_X(x) = \sum_{i=0}^{+\infty} p^x \cdot (1-p) = (1-p) \cdot \sum_{i=0}^{+\infty} p^x = 1$$

Formule

$$\Phi_X(t) = \frac{1-p}{1-e^t p}$$

$$\mathbb{E}[X] = \left.\frac{d\Phi_X(t)}{dt}\right|_{t=0} = \frac{p}{1-p}$$

è il numero medio di successi prima di un insuccesso

Esempio

Variabile aleatoria geometrica

$X \sim \text{Geo}(0.85)$ conta il numero di lanci prima che un razzo esploda
La media di lanci prima che il razzo esploda è $\mathbb{E}[X] \approx 5.667$
La probabilità che il razzo esploda al secondo lancio è $P_X(2) \approx 0.1084$
La probabilità che esploda prima del quarto lancio è $\sum_{x=1}^{3} P_X(x) \approx 0.3280$

Variabile aleatoria di Poisson

Definizione

$Y \sim \text{Bin}(n, p)$, $(n \to +\infty, p \to 0)$, $\frac{y}{n} \to 0$ $(y \ll n)$
Poniamo $\mu = n \cdot p$

$$P_Y(y) \approx \frac{\mu^y}{y!} \cdot e^{-\mu}$$

Dimostrazione

$$
\begin{aligned}
\dots &= \binom{n}{y} \cdot p^y \cdot (1-p)^{n-y} = \frac{n!}{y!(n-y)!} \cdot p^y \cdot (1-p)^{n-y} \\
&= \frac{n(n-1)\cdot \ldots \cdot (n-y)}{y!} \cdot p^y \cdot (1-p)^{n-y} \approx \frac{n^y}{y!} \cdot p^y \cdot (1-p)^{n-y} \\
&\approx \frac{n^y}{y!} \cdot p^y \cdot (1-p)^n = \frac{(np)^y}{y!} \cdot \sum_{i=0}^{n} \binom{n}{i} \cdot 1^{n-i} \cdot (-p)^i \\
&= \frac{(np)^y}{y!} \cdot \left[1 - np + \frac{n(n-1)}{2!}p^2 - \frac{n(n-1)(n-2)}{3!}p^3 + \ldots\right] \\
&\approx \frac{(np)^y}{y!} \cdot \left[1 - np + \frac{n^2}{2!}p^2 - \frac{n^3}{3!}p^3 + \ldots\right] \\
&= \frac{\mu^y}{y!}\left[1 - \mu + \frac{\mu^2}{2!} - \frac{\mu^3}{3!} + \ldots\right] = \ldots
\end{aligned}
$$

$Z \sim \text{Pois}(\mu)$, $\mu > 0$, $z \in \mathbb{N}$

$$P_Z(z) = \frac{\mu^z}{z!} \cdot e^{-\mu}$$

Approssima la probabilità di eventi simili che si possono verificare in un intervallo di tempo/spazio/...

Formule

$$\Phi_Z(t) = e^{\mu \cdot (e^t - 1)}$$

Dimostrazione

$$\dots = \mathbb{E}[e^{tZ}] = \sum_{z=0}^{+\infty} e^{tz} \cdot \frac{\mu^z}{z!} \cdot e^{-\mu} = e^{-\mu} \sum_{z=0}^{+\infty} \frac{(e^t\mu)^z}{z!} = e^{-\mu} e^{e^t\mu} = \dots$$

Si può notare che la media è uguale a quella del Bin
$\mathbb{E}[Z] = \mu = n \cdot p$

Dimostrazione

$$\dots = \left.\frac{d\Phi_Z(t)}{dt}\right|_{t=0} = e^{\mu \cdot (e^t - 1)} \cdot (\mu e^t) = \dots$$

$$\mathbb{E}[Z^2] = \left.\frac{d^2\Phi_Z(t)}{dt^2}\right|_{t=0} = \mu^2 + \mu$$

$$\text{Var}[Z] = \mu$$

Dimostrazione

$$\dots = \mathbb{E}[Z^2] - \mathbb{E}[Z]^2 = \mu^2 + \mu - \mu^2 = \dots$$

Approfondimento

Inoltre è riproducibile
$Z_1 \sim \text{Pois}(\mu_1) \perp\!\!\!\perp Z_2 \sim \text{Pois}(\mu_2)$, $Z = Z_1 + Z_2$
$\Phi_Z(t) = \Phi_{Z_1}(t) \cdot \Phi_{Z_2}(t) = e^{\mu_1 \cdot (e^t - 1)} \cdot e^{\mu_2 \cdot (e^{t-1})} = e^{(\mu_1 + \mu_2)(e^t - 1)} \implies Z \sim \text{Pois}(\mu_1 + \mu_2)$

Esempio

> # Variabile aleatoria di Poisson
>
> $Z \sim \text{Pois}(5)$ conta il numero di pizze prodotte in un'ora
> La media di pizze per ora è esattamente $\mathbb{E}[X] = 5$

La probabilità che siano state prodotte 6 pizze è $P_Z(6) \approx 0.1462$

La probabilità che siano state prodotte tra 3 e 8 pizze è $\sum_{z=3}^{8} P_Z(z) \approx 0.8073$

Variabile aleatoria ipergeometrica

Definizione

Estrazione senza reimmissione da un lotto contenente N elementi di cui D difettosi
L'estrazione di un pezzo difettoso vale 1 e di un pezzo non difettoso 0
Indica la probabilità di estrarre $k \leq D$ pezzi difettosi da un campione di $n \leq N$ elementi
$K \sim \mathrm{Iper}(N, n, D)$, $k \in [\max\{0, n - N + D\}, \min\{n, D\}]$

$$P_K(k) = \frac{\binom{D}{k}\binom{N-D}{n-k}}{\binom{N}{n}}$$

Ogni estrazione è una Bernoulliana con $p = \frac{D}{N}$

Approfondimento

La probabilità di estrarre un pezzo difettoso alla seconda estrazione è

$$\begin{aligned}\mathbb{P}(D\mathrm{II}) = \mathbb{P}(D\mathrm{II}|\mathrm{DI}) \cdot \mathbb{P}(D\mathrm{I}) + \mathbb{P}(D\mathrm{II}|\bar{\mathrm{D}}\mathrm{I}) \cdot \mathbb{P}(\bar{D}\mathrm{I}) &= \frac{D-1}{N-1} \cdot \frac{D}{N} + \frac{D}{N-1} \cdot \left(1 - \frac{N}{D}\right) \\ &= \frac{D^2 - D + DN - D^2}{(N-1)(N)} = \frac{D(N-1)}{N(N-1)} = \frac{D}{N}\end{aligned}$$

Formule

$$\mathbb{E}[K] = n \cdot \frac{D}{N}$$

Esempio

> # Variabile aleatoria ipergeometrica
>
> $K \sim \mathrm{Iper}(120, 7, 35)$ conta i pezzi difettosi estratti su un campione di 7 da un totale di 120 di cui 35 difettosi
> La media di pezzi difettosi estratti è $\mathbb{E}[K] \approx 2.042$
> La probabilità di estrarre 5 pezzi difettosi è $P_K(5) \approx 0.0195$
> La probabilità di estrarre tra 4 e 7 pezzi difettosi è $\sum_{k=4}^{7} P_K(k) \approx 0.1089$

Variabile aleatoria di Pascal

Definizione

Anche chiamata **negativa binomiale**
$Y \sim \mathrm{NegBin}(m, p)$
Conta il numero di insuccessi y prima di m successi, quindi per definizione l'ultimo è un successo, p è la

probabilità di successo

$$P_Y(y) = \binom{y+m-1}{y} \cdot p^m \cdot (1-p)^y$$

Si può anche scrivere come somma di m geometriche $X_i \sim \text{Geo}(1-p)$, $X_i \perp\!\!\!\perp X_j \;\; i \neq j$, che contano gli insuccessi

Formule

$$\Phi_Y(t) = \frac{(1-p)^m}{(1-e^t p)^m}$$

Dimostrazione

$$\dots = \mathbb{E}[e^{tY}] = \mathbb{E}\left[e^{t\sum_{i=1}^{y} X_i}\right] = \prod_{i=1}^{y} \mathbb{E}[e^{tX_i}] = \dots$$

$$\mathbb{E}[Y] = \frac{m(1-p)}{p}$$

Dimostrazione

$$\dots = \mathbb{E}\left[\sum_{i=1}^{m} X_i\right] = \sum_{i=1}^{y} \mathbb{E}[X_i] = \dots$$

Esempio

Variabile aleatoria di Pascal

$Y \sim \mathrm{NegBin}(15, 0.75)$ conta il numero di partite perse prima di riuscire a vincerne 15
La media di partite perse è $\mathbb{E}[Y] = 5$
La probabilità di perdere 3 partite è $P_Y(3) \approx 0.1420$
La probabilità di perdere tra 4 e 7 partite è $\sum_{y=4}^{7} P_Y(y) \approx 0.5328$

Variabile aleatoria uniforme

Definizione

$X \sim \mathrm{Unif}(a, b)$

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & x \in [a,b] \\ 0, & x \notin [a,b] \end{cases}$$

Formule

$$\Phi_X(t) = \frac{e^{tb} - e^{ta}}{(b-a) \cdot t}$$

Dimostrazione

$$\ldots = \mathbb{E}[e^{tX}] = \int_a^b e^{tx} \cdot \frac{1}{b-a}\, dx = \ldots$$

Approfondimento

Si può definire una trasformazione
$\varphi : [a,b] \to [0,1]$, $\varphi(y) = \frac{y-a}{b-a}$

$$\frac{X-a}{b-a} = Y \sim \mathrm{Unif}(0,1)$$

$$f_Y(y) = \begin{cases} 1, & y \in [0,1] \\ 0, & y \notin [0,1] \end{cases}$$

$$\Phi_Y(t) = \mathbb{E}[e^{tY}] = \frac{e^t - 1}{t}$$

$$\mathbb{E}[Y] = \frac{d\Phi_Y(t)}{dt}\bigg|_{t=0} = \frac{d}{dt}\left(\frac{e^t-1}{t}\right)\bigg|_{t=0} = \frac{te^t - (e^t-1)}{t^2}\bigg|_{t=0} = \lim_{t\to 0} \frac{t \cdot e^t}{2t} = \frac{1}{2}$$

$$\mathrm{Var}[Y] = \frac{1}{12}$$

Ora si possono riutilizzare i calcoli per $X = Y \cdot (b-a) + a$

Formule

$$\mathbb{E}[X] = \frac{b+a}{2}$$

Dimostrazione

$$\ldots = \mathbb{E}[Y \cdot (b-a) + a] = (b-a) \cdot \mathbb{E}[Y] + a = \frac{b-a}{2} + a = \ldots$$

$$\mathrm{Var}[X] = \frac{(b-a)^2}{12}$$

Dimostrazione

$$\ldots = \mathrm{Var}[Y \cdot (b-a) + a] = (b-a)^2 \cdot \mathrm{Var}[Y] = \ldots$$

Esempio

Variabile aleatoria uniforme

$X \sim \text{Unif}(0, 30)$ approssima la distribuzione di probabilità che l'autobus arrivi alla fermata tra le 10:00 e le 10:30
La media di minuti di attesa è $\mathbb{E}[X] = 15$
La probabilità di aspettare più di 10 minuti è $\int_{10}^{30} f_X(x)\, dx = 0.75$

Variabile aleatoria esponenziale

Definizione

$\mu = \lambda x$, $Y \sim \text{Pois}(\mu)$

$$P_Y(y) = \frac{(\lambda x)^y}{y!} \cdot e^{-\lambda x}$$

con $y \in \mathbb{N}$
x indica il tempo di funzionamento di un sistema riparabile, Y conta gli eventi guasto
La probabilità di non osservare guasti in $[0, x]$ è $\mathbb{P}(Y = 0) = P_Y(0) = e^{-\lambda x}$, $x, \lambda > 0$
Allo stesso modo non osservare guasti entro un tempo x equivale al corretto funzionamento per tale tempo
$X \sim \text{Exp}(\lambda)$ conta il tempo di funzionamento
$\mathbb{P}(X > x) = e^{-\lambda x}$ e $F_X(x) = \mathbb{P}(X \leq x) = 1 - e^{-\lambda x}$
$f_X(x) = \lambda \cdot e^{-\lambda x}$, $x > 0$
Viene utilizzata per misurare attese, code, decadimenti e rotture improvvise

Formule

$$\Phi_X(t) = \frac{\lambda}{\lambda - t}$$

con $t < \lambda$

Dimostrazione

$$\ldots = \mathbb{E}[e^{tX}] = \int_0^{+\infty} e^{-(\lambda - t)\cdot x}\, dx = \ldots$$

$$\mathbb{E}[X] = \frac{d\Phi_X(t)}{dt}\bigg|_{y=0} = \frac{1}{\lambda}$$

$$\text{Var}[X] = \frac{1}{\lambda^2}$$

Approfondimento

Vige della proprietà di assenza di memoria, dati $x_2 > x_1$, $x_2 = x_1 + x$ ($x > 0$)

$$\begin{aligned}\mathbb{P}(X > x_2 | X > x_1) &= \frac{\mathbb{P}((X > x_2) \cap (X > x_1))}{\mathbb{P}(X > x_1)} = \frac{\mathbb{P}(X > x_2)}{\mathbb{P}(X > x_1)} = \frac{1 - F_X(x_2)}{1 - F_X(x_1)} \\ &= \frac{e^{-\lambda(x_1 + x)}}{e^{-\lambda x_1}} = \frac{e^{-\lambda x_1} \cdot e^{-\lambda x}}{e^{-\lambda x_1}} = e^{-\lambda x} = \mathbb{P}(X > x)\end{aligned}$$

Esempio

Variabile aleatoria esponenziale

$X \sim \mathrm{Exp}\left(\frac{1}{50}\right)$ conta il tempo di funzionamento di una batteria di un'auto
La media del tempo di vita prima della rottura è $\mathbb{E}[X] = 50$ mila chilometri
La probabilità che duri meno di 45 mila chilometri è $\mathbb{P}(X \leq 45) \approx 0.5934$
La probabilità che si rompa dopo più di 40 mila chilometri è $\mathbb{P}(X > 40) \approx 0.4493$

Variabile aleatoria gamma

Definizione

$X_i \sim \mathrm{Exp}(\lambda),\, i = 1, \dots, n$
$X_n = \sum_{i=1}^{n} X_i \sim \mathrm{Gamma}(n, \lambda)$ può essere pensato come periodo complessivo di funzionamento avendo a disposizione n elementi, utilizzati uno dopo l'altro appena si verifica il guasto del precedente
Y conta il numero di elementi rotti (quindi di guasti), la probabilità che il periodo di funzionamento sia $> x$ è

$$\mathbb{P}(X > x) = 1 - F(x) = \sum_{y=0}^{n-1} \frac{(\lambda x)^y}{y!} e^{-\lambda x} = \mathbb{P}(Y < n)$$

ovvero la probabilità che si verifichino $n-1$ guasti

Formule

$$F_{X_n}(x) = \sum_{y=n}^{+\infty} \frac{(\lambda x)^y}{y!} \cdot e^{-\lambda x}$$

$$f_{X_n}(x) = \lambda e^{-\lambda x} \cdot \frac{(\lambda x)^{n-1}}{(n-1)!}$$

Dimostrazione

$$\begin{aligned} \dots &= \frac{dF_{X_n}(x)}{dx} = \sum_{y=n}^{+\infty} \frac{\lambda(\lambda x)^{y-1}}{(y-1)!} \cdot e^{-\lambda x} - \sum_{y=n}^{+\infty} \frac{\lambda(\lambda x)^y}{(y)!} \cdot e^{-\lambda x} \\ &= \lambda e^{-\lambda x} \left[\sum_{z=n-1}^{+\infty} \frac{(\lambda x)^z}{z!} - \sum_{y=n}^{+\infty} \frac{(\lambda x)^y}{y!} \right] = \dots \end{aligned}$$

$$\Phi_{X_n}(t) = \left(\frac{\lambda}{\lambda - t} \right)^n$$

con $t < \lambda$

Definizione

Si definisce una funzione gamma anche per n non interi tale che

$$f_{X_n}(x) = \lambda e^{-\lambda x} \cdot \frac{(\lambda x)^{n-1}}{\Gamma(n)}$$

Se $n \in N$, da prima, $\Gamma(n) = (n-1)!$
Se invece $\alpha \in \mathbb{R}$

$$\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1} \cdot e^{-x} \, dx$$

Esempio

Alcuni valori noti: $\Gamma(\alpha+1) = \alpha \cdot \Gamma(\alpha)$, $\Gamma(1) = 1$, $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$

Esempio

> ## Variabile aleatoria gamma
>
> $X \sim \text{Gamma}\left(8, \frac{1}{50}\right)$ conta la strada percorsa da un'auto durante una gara avendo a disposizione 8 pieni di carburante che si consumano ogni 50 chilometri
> La probabilità che l'auto riesca a percorrere 250 chilometri è $\mathbb{P}(X > 250) \approx 0.8666$

Variabile aleatoria gaussiana

Definizione

Anche chiamata **normale**, modello di interpretazione di errori o scostamenti
Lo scostamento $X - \mu$, con μ valore vero, accompagna le misure sperimentali di un certo valore X effettuate nelle stesse condizioni

$$U = \frac{X - \mu}{\sigma}$$

sono gli errori di misura come multipli della loro ampiezza tipica, $\sigma = \sqrt{\text{Var}[X]}$
$U \sim \text{N}(0, 1)$ è variabile normale standard con media 0 e deviazione 1, senza errori sistematici

$$f_U(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2}$$

Dimostrazione

Se le misure non sono affette da errori sistematici:

- $\mathbb{E}[U] = 0$
- $f_U(u) = f_U(-u)$ e $\lim_{u\to+\infty} f_U(u) = \lim_{u\to-\infty} f_U(u) = 0$
- $f_U(0) > f_U(u) \ \ \forall u \in \mathbb{R} \setminus \{0\}$

Ovvero:

$$\frac{df_U(u)}{du} = -f_U(u) \cdot u$$

e $f_U(u) > 0 \ \ \forall u \in \mathbb{R}$
Una soluzione all'equazione differenziale è $e^{-u^2/2}$
$f_U(u) = k \cdot e^{-u^2/2}$, per essere densità

$$1 = \int_{-\infty}^{+\infty} k \cdot e^{-u^2/2}\, du = 2k \int_0^{+\infty} e^{-u^2/2}\, du$$

$$t = \frac{u^2}{2} \quad u = \sqrt{2t} \quad du = \frac{1}{\sqrt{2}} t^{(1/2)-1}$$

$$= \frac{2}{\sqrt{2}} k \int_0^{+\infty} t^{(1/2)-1} \cdot e^{-t}\, dt = \frac{2}{\sqrt{2}} k \cdot \Gamma\left(\frac{1}{2}\right) = \frac{2\sqrt{\pi}}{\sqrt{2}} k \implies k = \frac{1}{\sqrt{2\pi}}$$

$X = \mu + \sigma U \sim \mathrm{N}(\mu, \sigma^2)$ è variabile aleatoria normale (non standard)

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-(1/2)((X-\mu)/\sigma)} \left| \frac{du}{dx} \right| = \frac{1}{\sigma\sqrt{2\pi}} e^{-(1/2)((x-\mu)/\sigma)^2}$$

con $\sigma > 0$

Formule

$\Phi_U(t) = e^{t^2/2}$

$\Phi_X(t) = e^{t\mu + t^2\sigma^2/2}$

$\mathbb{E}[X] = \mu$

Dimostrazione

$\ldots = \mathbb{E}[\mu + \sigma U] = \mu + \sigma\mathbb{E}[U] = \ldots$

$\mathrm{Var}[X] = \sigma^2$

Approfondimento

$X_1, X_2 \sim \mathrm{N}(0,1)$, $X_1 \perp\!\!\!\perp X_2$, $Y = X_1 + X_2$

$\Phi_Y(t) = \Phi_{X_1}(t) \cdot \Phi_{X_2}(t) = e^{t^2/2} \cdot e^{t^2/2} = e^{2t^2 2}$

Quindi $Y \sim \mathrm{N}(0, 2)$

$X_i \sim \mathrm{N}(\mu_i, \sigma_i^2)$, $X_i \perp\!\!\!\perp X_j \;\; i \neq j$, $a_i \in \mathbb{R}$ $(i, j = 1, \ldots, n)$
$Y = \sum_{i=1}^n a_i X_i$
$\mathbb{E}[Y] = \sum_{i=1}^n a_i \mu_i = \bar{\mu}$
$\mathrm{Var}[X] = \sum_{i=1}^n (a_i \sigma_i)^2 = \bar{\sigma}^2$

$$\Phi_Y(t) = \prod_{i=1}^n \Phi_{X_i}(t) = e^{t\bar{\mu} + t^2\bar{\sigma}^2/2}$$

Quindi $Y \sim \mathrm{N}(\bar{\mu}, \bar{\sigma}^2)$

$X_i \sim \mathrm{N}(\mu_i, \sigma_i^2)$, $X_i \perp\!\!\!\perp X_j \;\; i \neq j$, $a_i, b_i \in \mathbb{R}$ $(i, j = 1, \ldots, n)$
$Y = \sum_{i=1}^n a_i X_i$, $Z = \sum_{i=1}^n b_i X_i$
$\mathrm{Cov}(Y, Z) = \mathbb{E}[(Y - \mathbb{E}[Y])(Z - \mathbb{E}[Z])] = \sum_{i=1}^n a_i b_i \sigma_i^2$
Se $\sigma_i = \sigma \;\; \forall i \in \{1, \ldots, n\}$
$\mathrm{Cov}(Y, Z) = 0 \iff \sum_{i=1}^n a_i b_i = 0$, Y e Z sono dette ortogonali

Esempio

Variabile aleatoria gaussiana

$X \sim \mathrm{N}(0.45, 0.05^2)$ conta i litri erogati da un distributore automatico con media 0.45 e deviazione 0.05
$U \sim \mathrm{N}(0, 1)$ normale standard dalla quale si ottengono i valori tabulati
La probabilità che siano erogati più di 0.5 litri è $\mathbb{P}(X > 0.5) = 1 - \mathbb{P}(U \leq 1) \approx 0.1587$

Variabile aleatoria normale bivariata

Definizione

$X \sim \mathrm{N}(\mu_1, \sigma_1^2)$, $Y \sim \mathrm{N}(\mu_2, \sigma_2^2)$, $X \perp\!\!\!\perp Y$, per semplicità $\mu_1 = \mu_2 = 0$
Allora si può definire

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2} e^{-(1/2)(x^2/\sigma_1^2 + y^2/\sigma_{2^2})}$$

Dimostrazione

$$\ldots = f_X(x) \cdot f_Y(y) = \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-(1/2)(x^2/\sigma_1^2)} \cdot \frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-(1/2)(y^2/\sigma_2^2)} = \ldots$$

(X, Y) si dice normale bivariata

$X_i \sim \mathrm{N}(0, 1)$, $X_i \perp\!\!\!\perp X_j \;\; i \neq j \; (i, j = 1, \ldots, n)$
Allora

$$f(x_1, \ldots, x_n) = \frac{1}{(\sqrt{2\pi})^n} e^{-(1/2)(x_1^2 + \cdots + x_n^2)}$$

Il vettore $(X_1, \ldots, X_n)$ si dice gaussiana multivariata standard

Teorema centrale del limite

Teorema

X_i variabile aleatoria, $X_i \perp\!\!\!\perp X_j \;\; i \neq j$, tutte seguenti la stessa distribuzione con media μ e varianza σ^2 ($i, j = 1, \dots, n$)
$Y = \sum_{i=1}^n X_i$ converge in distribuzione a $\mathrm{N}(n\mu, n\sigma^2)$

Dimostrazione

Si dice che una successione $\{X_n\}_{n\in\mathbb{N}}$ di variabili aleatorie converge in distribuzione a X se $\lim_{n\to+\infty} F_{X_n}(x) = F_X(x) \;\; \forall x$ in cui $F_X(x)$ è continua

$$Y_i = \frac{X_i - \mu}{\sigma}$$

$$Z = \frac{\left(\sum_{i=1}^n X_i\right) - n\mu}{\sqrt{n}\sigma} = \sum_{i=1}^n \frac{X_i - \mu}{\sqrt{n}\sigma} = \frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i \sim \mathrm{N}(0,1)$$

$$\begin{aligned}
\Phi_Z(t) &= \mathbb{E}[e^{tZ}] = \mathbb{E}\left[e^{t\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n Y_i\right)}\right] = \mathbb{E}\left[\prod_{i=1}^n e^{\frac{t}{\sqrt{n}}Y_i}\right] = \prod_{i=1}^n \mathbb{E}\left[e^{\frac{t}{\sqrt{n}}Y_i}\right] \\
&= \prod_{i=1}^n \mathbb{E}\left[1 + \frac{t}{\sqrt{n}}Y_i + \left(\frac{t}{\sqrt{n}}\right)^2 \cdot \frac{Y_i^2}{2!} + \dots\right] \\
&= \prod_{i=1}^n \left[\mathbb{E}[1] + \frac{t}{\sqrt{n}}\mathbb{E}[Y_i] + \left(\frac{t}{\sqrt{n}}\right)^2 \cdot \frac{\mathbb{E}[Y_i^2]}{2!} + \dots\right] \\
&\quad \mathbb{E}[1] = 1 \quad \mathbb{E}[Y_i] = \frac{\mathbb{E}[X_i] - \mu}{\sigma} = \frac{\mu - \mu}{\sigma} = 0 \quad \mathbb{E}[Y_i^2] = 1 \\
&= \prod_{i=1}^n \left[1 + 0 + \frac{1}{n}\left(\frac{t^2}{2!} + \frac{t^3}{\sqrt{n}} \cdot \frac{\mathbb{E}[Y_i^3]}{3!} + \dots\right)\right]
\end{aligned}$$

$\mathbb{E}[Y_i] = \mathbb{E}[Y_j] \;\; \forall i, j = 1, \dots, n$ poiché hanno la stessa distribuzione

$$\dots = \left[1 + \frac{1}{n}\left(\frac{t^2}{2!} + \frac{t^3}{\sqrt{n}} \cdot \frac{\mathbb{E}[Y_i^3]}{3!} + \dots\right)\right]^n$$

$$\lim_{n\to+\infty} \frac{n}{f(n)} = \pm\infty \implies \lim_{n\to+\infty}\left[1 + \frac{f(n)}{n}\right]^n = e^{\lim_{n\to+\infty} f(n)}$$

$$n \to +\infty \implies \dots = e^{\lim_{n\to+\infty}\left(\frac{t^2}{2!} + \frac{t^3}{\sqrt{n}} \cdot \frac{\mathbb{E}[Y_i^3]}{3!} + \dots\right)} = e^{t^2/2}$$

Ovvero è la funzione di generazione dei momenti della gaussiana standard
$Y = Z\sqrt{n}\sigma + n\mu \implies \dots$

Approssimazione normale della distribuzione binomiale

Teorema di De Moivre-Laplace

Una variabile aleatoria $\mathrm{Bin}(n, p)$ per n grandi ha approssimativamente la stessa distribuzione di una $\mathrm{N}(np, np(1-p))$

$X_i \sim \mathrm{Bin}(n, p)$, $X_i \perp\!\!\!\perp X_j \;\; i \neq j$
$S_n = \sum_{i=1}^{n} X_i$ è il numero di successi
Se $n \to +\infty$

$$\frac{S_n - np}{\sqrt{np(1-p)}} \sim \mathrm{N}(0, 1)$$

$U \sim \mathrm{N}(0, 1)$
Se $a < b$ e $np(1-p) > 10$

$$\lim_{n \to \infty} \mathbb{P}\left(a \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq b\right) = F_U(b) - F_U(a)$$

è una buona approssimazione

Correzione di continuità

Nell'approssimazione di una variabile aleatoria discreta con una continua per t si considera l'intervallo $t - \frac{1}{2} \leq y \leq t + \frac{1}{2}$