5 Modelli di variabili casuali

5.1 Modelli discreti

Nel mondo reale per descrivere diversi fenomeni si utilizzano delle variabili casuali discrete le quali, per essere caratterizzate, necessitano di poche distribuzioni di probabilità.
Per esempio, per testare l'efficacia di un farmaco, il numero di pazienti che guariscono dopo l'assunzione di tale farmaco può essere descritto da una variabile casuale *binomiale*; a livello industriale, se si seleziona un campione da un lotto di produzione, il numero di unità difettose presenti nel campione può essere descritto da una variabile casuale *ipergeometrica*; in un problema di controllo della qualità il numero di campioni richiesti per produrre un falso allarme può essere descritto da una variabile casuale *geometrica*; il numero di globuli bianchi in un campione di sangue può essere descritto da una variabile casuale *poissoniana*.

5.1.1 Distribuzione uniforme

È una distruibuzione che è uniforme su un insieme, cioè che attribuisce la stessa probabilità ad ogni elemento dell'insieme su cui è definita.

$$f(x;n) = \begin{cases} \dfrac{1}{n} & x = 1, 2, ..., n\,, \\ 0 & \text{altrove}\,. \end{cases}$$

Un esempio è fornito dal lancio di un dado equilibrato, in cui ogni faccia ha probabilità $\frac{1}{6}$ di presentarsi ad ogni lancio.

- $E[X] = \frac{n+1}{2}$
- $\text{var}[X] = \frac{n^2-1}{12}$

Questa distribuzione è quella che fornisce la classica definizione di probabilità "casi favorevoli" su "casi possibili".

5.1.2 Distribuzione binomiale

Spesso un esperimento consiste in una serie di prove ripetute, ognuna con due soli esiti possibili (si/no; vero/falso; successo/insuccesso; 0/1). Ogni prova il cui esito è dicotomico è detta *prova di Bernoulli*.

Un processo di Bernoulli ha le seguenti caratteristiche:

1. l'esito è dicotomico
2. la probabilità di successo, indicata con p, è la stessa in ogni prova
3. l'esperimento consiste in prove ripetute
4. le prove sono indipendenti

La variabile casuale X che conta il numero di successi che si registrano in n prove di Bernoulli è chiamata binomiale e la sua funzione di densità è:

$$B(x, n, p) = \binom{n}{x} p^x q^{n-x}, \qquad x = 0, 1, ..., n$$

dove $\binom{n}{x} = \dfrac{n!}{x!(n-x)!}$, $p \in [0, 1]$ è la probabilità di successo e $q = 1 - p$ è la probabilità di insuccesso. Il nome nasce dal fatto che nello sviluppo binomiale di $(p+q)^n$ i termini corrispondono ai valori di $B(x, n, p)$ per $x = 0, ..., n$.

Nota bene: $\sum_{i=0}^{n} B(i, n, p) = 1$; $(p + q = 1)$.

- $E[X] = np$
- $\mathrm{var}[X] = npq$

La distribuzione binomiale viene utilizzata in ambito industriale, in ambito medico, in ambito militare, nelle estrazioni *con reinserimento dell'oggetto* (per avere l'indipendenza delle prove).

Esempi

1) La probabilità che un componente resista ad una prova d'urto è del 75%. Calcolare la probabilità che esattamente due dei 4 componenti sottoposti alla prova, la superino con successo.
$x = 2, n = 4, p = \frac{3}{4}$.
$B(2, 4, \frac{3}{4}) = \binom{4}{2}\left(\frac{3}{4}\right)^2\left(\frac{1}{4}\right)^2 = \frac{4!}{2!2!}\frac{3^2}{4^4} = \frac{27}{128}$

2) Il 5% dei chip di memoria prodotti da una fabbrica è difettoso. Determinare la probabilità che su 4 chip scelti a caso:
a) uno sia difettoso;
b) nessuno sia difettoso;
c) meno di due siano difettosi.
Infine, se X è la variabile casuale che indica il numero dei chip difettosi su un totale di 400 chip, determi-

nare $E[X]$ e $\mathrm{var}[X]$.
Per la prima parte dell'esercizio:
$n = 4, \quad p = 0.05$
a) $P[X = 1] = \binom{4}{1}(0.05)^1 \cdot (0.95)^3 = 0.171$
b) $P[X = 0] = \binom{4}{0}(0.05)^0 \cdot (0.95)^4 = 0.814$
c) $P[X < 2] = P[X = 0] + P[X = 1] = 0.985$
Per la seconda parte dell'esercizio:
$n = 400, \quad p = 0.05$
$E[X] = np = 20 \quad \mathrm{var}[X] = npq = 19$

5.1.3 Distribuzione ipergeometrica

Si distingue dalla distribuzione binomiale perchè non viene richiesta l'indipendenza delle prove e si basa sul campionamento *senza reinserimento* dell'oggetto.
Viene utilizzata nel campionamento in accettazione e nella certificazione di qualità.
Lo scopo è determinare la probabilità di selezionare x successi da k unità indicate come successi ed $n-x$ insuccessi da $N-k$ unità indicate come insuccessi, dato un campione casuale di dimensione n selezionato da N unità.
Il numero x di successi dipende dal numero di successi k presenti nell'insieme N da cui vengono sele-

zionate n unità. La funzione di densità è:

$$f(x, N, n, k) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}, \qquad x = 0, 1, ..., n$$

con $k > n$ e $N - k > n$.

- $E[X] = \frac{nk}{N}$
- $\text{var}[X] = \frac{nk(N-k)(N-n)}{N^2(N-1)}$

Esempio

Un rivenditore acquista chip a lotti di 10. In ogni lotto controlla a caso 3 chip.
Decide di accettare il lotto solo se nessuno dei 3 chip è difettoso.
Sapendo che il 30% dei lotti ha 4 pezzi difettosi su 10 e il 70% dei lotti ha 1 pezzo difettoso su 10, calcolare la percentuale di rifiuto del rivenditore.
Chiamiamo:
A l'evento "il lotto viene accettato"
$R = \overline{A}$ l'evento "il lotto viene rifiutato"
D_i il lotto ha i pezzi difettosi.

$$P[A] = P[A|D_4]P[D_4] + P[A|D_1]P[D_1]$$ per il th. delle prob. totali

$P[D_4] = \frac{3}{10}, \quad P[D_1] = \frac{7}{10};$

$$P[A|D_4] = \frac{\binom{4}{0}\binom{6}{3}}{\binom{10}{3}} = \frac{1 \cdot \frac{6!}{3!3!}}{\frac{10!}{3!7!}} = \frac{1}{6}$$

$$P[A|D_1] = \frac{\binom{1}{0}\binom{9}{3}}{\binom{10}{3}} = \frac{1 \cdot \frac{9!}{3!6!}}{\frac{10!}{3!7!}} = \frac{7}{10}$$

Perciò:
$P[A] = \frac{1}{6} \cdot \frac{3}{10} + \frac{7}{10} \cdot \frac{7}{10} = \frac{54}{100} = 54\%$
$P[R] = P[\overline{A}] = 1 - P[A] = 46\%$

5.1.4 Distribuzione geometrica

Si consideri un esperimento in cui le proprietà sono le stesse descritte in un esperimento binomiale, ma con l'eccezione che le prove indipendenti vengano ripetute fino al raggiungimento del primo successo.
Perciò l'obiettivo non è calcolare la probabilità di ottenere x successi in n prove, con n noto, bensì la probabilità che si verifichi il primo successo nella x-esima prova.
La variabile casuale geometrica conta il numero x di

prove necessarie per ottenere il primo successo.

$$G(x,p) = pq^{x-1} \quad x = 1, 2, 3, \ldots \text{ (non finito)}$$

dove p è la probabilità di successo, $q = 1 - p$ la probabilità di insuccesso.

- $E[X] = \dfrac{1}{p}$
- $\mathrm{var}[X] = \dfrac{q}{p^2} = \dfrac{1-p}{p^2}$

X è detta tempo di attesa (es. il ritardo di un numero nel gioco del lotto).
$E[X] = \frac{1}{p}$ è detta tempo di ritorno, è la media del numero di ripetizioni da tentare per avere un successo.
La variabile casuale geometrica gode della proprietà “mancanza di memoria”.
In uno schema successo-insuccesso supponiamo di non aver ottenuto alcun successo nelle prime k prove.
Qual è la probabilità di dover attendere ancora m prove per avere il primo successo?
È la stessa che si avrebbe se le prime k prove non fossero state eseguite.

Ciò è ovvio perchè in uno schema a prove ripetute indipendenti i risultati delle prime non influenzano le successive.

Esempi

1) Di un certo processo manifatturiero si conosce che, in media, 1 prodotto su 100 è difettoso.
Qual è la probabilità che il quinto prodotto controllato sia il primo identificato come difettoso?
$p = 0.01 =$ probabilità di avere un difetto
$x = 5$
$G(5, 0.01) = 0.01 \cdot (0.99)^4 = 0.0096$
2) Ad un centralino telefonico tutte le linee sono occupate. Sapendo che $p = 0.05$ è la probabilità di ottenere una linea libera in un momento di massima affluenza di telefonate, calcolare la probabilità che siano necessari 5 tentativi prima di trovare la linea libera.
$p = 0.05 =$ probabilità di avere la linea libera
$x = 5$
$G(5, 0.05) = 0.05 \cdot (0.95)^4 = 0.041$

5.1.5 Distribuzione di Poisson

La distribuzione di probabilità poissoniana si utilizza negli esperimenti in cui è possibile fare un qualche tipo di conteggio, in un certo intervallo di tempo o di spazio. La variabile casuale X che rappresenta il numero di eventi che si vericano in un intervallo di tempo o di spazio è non finita e segue la legge:

$$P(x, \lambda) = \frac{e^{-\lambda}\lambda^{x}}{x!}, \quad x = 0, 1, 2, ...$$

con $\lambda > 0$.
Il parametro λ rappresenta il numero medio di realizzazioni dell'evento nell'intervallo considerato.

- $E[X] = \lambda$
- $\text{var}[X] = \lambda$

Esempi di esperimenti in cui si utilizza la variabile casuale di Poisson sono:

– il numero di chiamate telefoniche ricevute da un centralino in un intervallo di tempo fissato

– il numero di clienti che entrano in un ufficio in un giorno

– il numero degli incidenti stradali in una regione in una settimana

– il numero di refusi in una pagina di un libro

L'evento che si verifica in un certo intervallo è un *evento raro*.

Esempi

1) In un esperimento, il numero medio di particelle radioattive individuate da un contatore in 1 millisecondo è pari a 4.
Qual è la probabilità che 6 particelle vengano individuate in un dato millisecondo?

$x = 6$ e $\lambda = 4$

$$P[6, 4] = \frac{e^{-4}4^6}{6!} = 0.1042$$

2) Il numero medio di navi che attraccano in un porto, in un giorno, è pari a 10 e la struttura può gestire 15 navi al giorno. Qual è la probabilità che in un dato giorno le navi non possano attraccare?

$\lambda = 10$

$$P[X > 15] = 1 - P[X \leq 15] = 1 - \sum_{x=0}^{15} \frac{e^{-10}10^x}{x!} =$$

$$1 - \left(e^{-10} + e^{-10} \cdot 10 + \frac{e^{-10} \cdot 10^2}{2!} + \ldots + \frac{e^{-10}10^{15}}{15!} \right)$$

$$= 0.0487$$

5.2 Modelli continui

Le distribuzioni continue nascono da un processo di misurazione del fenomeno di interesse. Quando si dispone di un'espressione matematica adatta alla rappresentazione di un fenomeno continuo, siamo in grado di calcolare la probabilità che la variabile casuale assuma valori compresi in intervalli. Infatti la probabilità che la variabile casuale continua assuma un *particolare valore* è pari a zero. Questa caratteristica distingue appunto i fenomeni continui, che sono il risultato di un processo di misurazione, da quelli discreti, ottenuti mediante un processo di conteggio.

I modelli continui hanno importanti applicazioni in ingegneria, fisica, economia, nelle scienze sociali. Alcuni tipici fenomeni continui sono l'altezza, il peso, le variazioni dei prezzi azionari giornalieri, il tempo che intercorre fra gli arrivi di aerei in un aeroporto.

5.2.1 Distribuzione uniforme continua

Tra tutte le distribuzioni continue, la più semplice è la distribuzione uniforme o rettangolare, caratterizzata da una funzione di densità "piatta". La variabile casuale continua uniforme X è definita su un intervallo chiuso $[a, b]$ e la sua probabilità è uniforme in $[a, b]$:

$$U(x; a, b) = \begin{cases} \dfrac{1}{b-a} & a \leq x \leq b\,, \\ 0 & \text{altrove}\,. \end{cases}$$

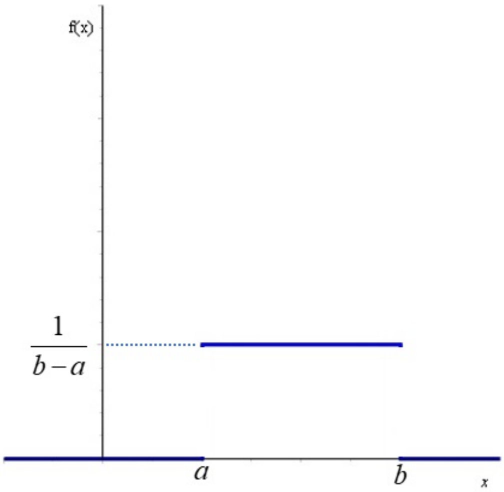

Figura 1: Distribuzione uniforme

- $E[X] = \frac{a+b}{2}$
- $\text{var}[X] = \frac{(b-a)^2}{12}$

I precedenti valori si ottengono applicando la definizione di valore atteso e varianza di una variabile casuale continua.

Esempi

1) Una sala per congressi può essere prenotata per al massimo 4 ore. La durata di una conferenza può essere rappresentata da una variabile casuale uniforme X su $[0, 4]$.

– Scrivere la funzione densità

– Calcolare la probabilità che una qualunque conferenza duri almeno 3 ore.

$$f(x) = \begin{cases} \dfrac{1}{4} & 0 \leq x \leq 4\,, \\ 0 & \text{altrove}\,. \end{cases}$$

$P[X \geq 3] = \int_3^4 \frac{1}{4} dx = \frac{1}{4}$

2) Gli autobus passano nei pressi dell'università ogni ora tra le 8:30 e le 13:30. Calcolare la probabilità che una persona debba aspettare almeno un quarto d'ora durante tale periodo.

La variabile casuale che indica "tempo mancante al prossimo autobus" segue un modello uniforme.
$P[X \geq 15] = 1 - P[X < 15] = 1 - P[X \leq 15] = 1 - F(15)$

$$F(x) = P[a \leq X \leq x] = \int_a^x \frac{1}{b-a} dt = \frac{x-a}{b-a}$$

e $b - a = 60\text{min} - 0\text{min}$

$F(15) = \frac{15-0}{60-0} = \frac{1}{4} \Rightarrow P[X \geq 15] = \frac{3}{4}$

5.2.2 Distribuzione normale o gaussiana

È la distribuzione di probabilità continua più importante. Descrive molti fenomeni che si verificano in natura, si usa per misurazioni fisiche e approssima molto bene gli errori di misurazione. Gioca un ruolo fondamentale in statistica ed è il cardine del teorema

del limite centrale.
Una variabile casuale continua X è normale se:

$$N(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad x \in \mathbb{R}$$

con $\mu \in \mathbb{R}$, $\sigma > 0$.

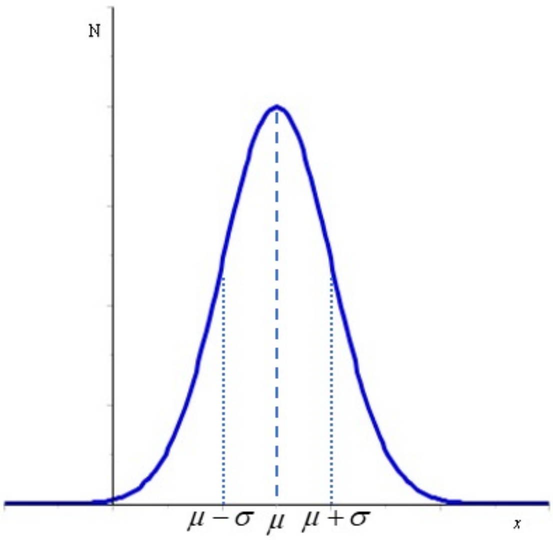

Figura 2: Distribuzione normale

- $E[X] = \mu$
- $\mathrm{var}[X] = \sigma^2$

Proprietà

1. N è simmetrica rispetto alla retta $x = \mu$ (mediana=media)

2. N ha un punto di massimo in $x = \mu$ (moda = media)

3. N ha due punti di flesso in $x = \mu \pm \sigma$

4. $\lim_{x \to \pm\infty} N(x, \mu, \sigma) = 0$

5. modificare il valore di μ equivale a traslare su Ox il grafico di N senza deformarlo

6. cambiare il valore di σ equivale a modificare la forma del grafico di N senza traslarlo.

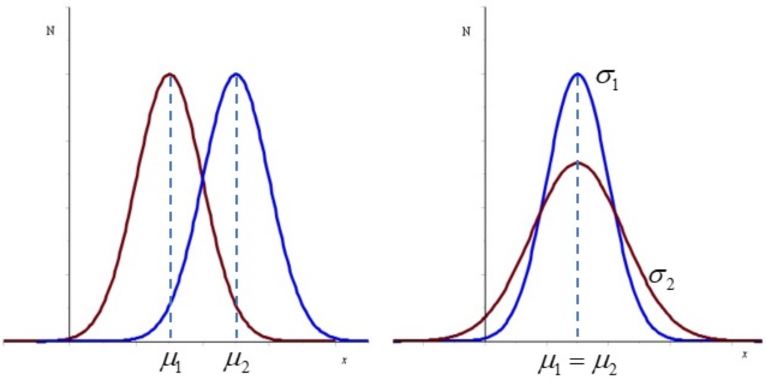

Figura 3: Distribuzioni normali

Poichè è impossibile riportare in tabella i valori di porzioni di area al di sotto della curva per ogni valore di μ e di σ si trasforma ogni variabile casuale normale X in una variabile casuale normale Z avente media nulla e varianza unitaria.

Definizione: Una variabile casuale normale con $\mu = 0$ e $\sigma = 1$ è detta normale standard o ridotta.
Se $X \sim N(\mu, \sigma^2) \Rightarrow Z = \frac{X-\mu}{\sigma} \sim N(0, 1)$
La funzione di densità di una variabile casuale normale standard Z è:

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}z^2}, \quad z \in \mathbb{R}$$

e risulta indipendente da μ e σ.
La sua funzione di ripartizione è:

$$\Phi(z) = \int_{-\infty}^{z} \phi(t)dt$$

Proprietà

- $P[a \leq X \leq b] = P\left[\frac{a-\mu}{\sigma} \leq Z \leq \frac{b-\mu}{\sigma}\right] = = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$
- $P[X \geq a] = 1 - P[X < a] = 1 - P\left[Z < \frac{a-\mu}{\sigma}\right] = 1 - \Phi\left(\frac{a-\mu}{\sigma}\right)$
  (ricorda che per le proprietà della funzione di ripartizione $\Phi(-\infty) = 0$ e $\Phi(+\infty) = 1$)
- $\Phi(-z_0) = 1 - \Phi(z_0)$

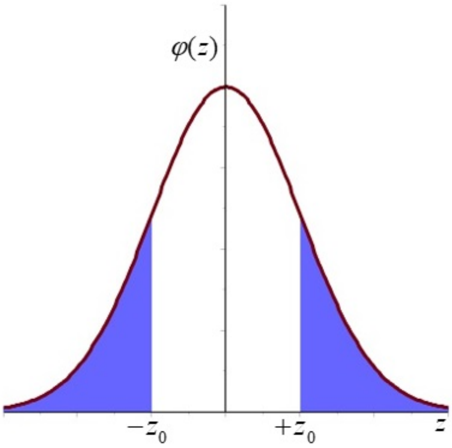

Figura 4: Distribuzione normale standard

Infatti
$P[Z < -z_0] = \Phi(-z_0)$
ma
$P[Z < -z_0] = P[Z > z_0] = 1 - P[Z \leq z_0] = 1 - \Phi(z_0)$

<u>Esempio</u>
Il peso di 2 confezioni di un prodotto è una variabile casuale normale X con media $\mu = 250$ g e deviazione standard $\sigma = 3$ g. Calcolare la probabilità che il peso delle 2 confezioni sia minore di 245 g.

$X \sim N(250, 3) \rightarrow Z = \frac{X-250}{3}$

$$P[X < 245] = P\left[Z < \frac{240-250}{3}\right] = P[Z < -\frac{5}{3}] =$$

$$= 1 - P[Z < \frac{5}{3}] = 1 - \Phi(1.67) = \quad \text{dalla tabella}$$

$$= 1 - 0.9525 = 0.0475$$

Problema inverso

Fino ad ora abbiamo visto come, conoscendo il valore in ascissa z (o x), sia possibile calcolare l'area sottesa dalla curva $\phi(z)$ (o $N(x)$), cioè valutare la funzione di ripartizione.

In alcuni casi invece il dato a disposizione è una determinata probabilità (= area) di cui si vuole conoscere il corrispondente valore di percentile.

Se $X \sim N(\mu, \sigma^2)$

Dato $\alpha \in [0, 1]$, determinare il valore x_α tale che $P[X > x_\alpha] = \alpha$

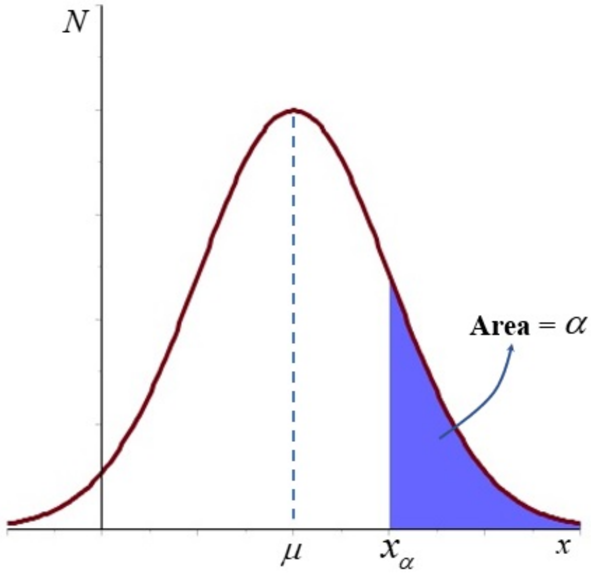

Figura 5

Dato $\alpha \in [0,1]$, determinare il valore z_α tale che $P[Z > z_\alpha] = \alpha$

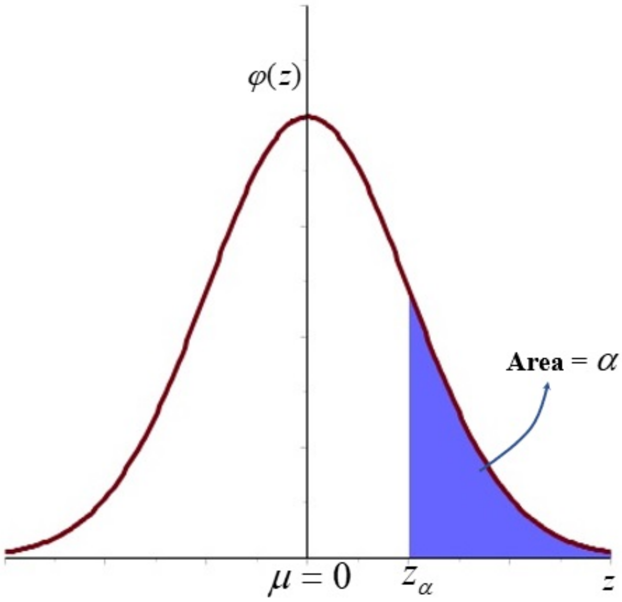

Figura 6

Esempio

Sia X una variabile casuale normale $N(\mu = 19, \sigma^2 = 49)$.

Determinare il valore x_α tale che:

$$P[X > x_\alpha] = 20\%$$

$0.20 = P[X > x_\alpha] = P\left[Z > \frac{x_\alpha - 19}{7}\right] = P[Z > z_\alpha]$
$= 1 - P[Z \leq z_\alpha]$

quindi

$P[Z \leq z_\alpha] = 0.8$

Dalla tabella della $\Phi(z)$ non troviamo esattamente 0.80 ma:

$0.7995 \rightarrow z = 0.84$

$0.8023 \rightarrow z = 0.85$

Si può mostrare che una buona approssimazione si ottiene se si approssima la curva $\Phi(z)$ tramite una retta nell'intorno considerato. Scriviamo l'equazione della retta passante per due punti:

$$\frac{y - y_1}{y_2 - y_1} = \frac{x - x_1}{x_2 - x_1}$$

e cerchiamo il valore di $\overline{x}$ che corrisponde ad $\overline{y} = 0.80$:

$$\frac{0.8 - 0.7995}{0.8023 - 0.7995} = \frac{\overline{x} - 0.84}{0.85 - 0.84} \quad \Rightarrow \quad \overline{x} = 0.8418$$

Allora il valore cercato è $z_\alpha = 0.8418$
Poichè $z_\alpha = \frac{x_\alpha - 19}{7}$ si ricava $x_\alpha = 24.8925$

Un'approssimazione più rozza si ottiene calcolando la media tra 0.84 e 0.85, cioè $z_\alpha = 0.845$, da cui si ricava $x_\alpha = 24.915$.

Valori tabulati

- $P[Z \geq z_\alpha] = 1\% = 0.01 \Rightarrow z_\alpha = 2.326$
- $P[Z \geq z_\alpha] = 5\% = 0.05 \Rightarrow z_\alpha = 1.645$
- $P[Z \geq z_\alpha] = 2.5\% = 0.025 \Rightarrow z_\alpha = 1.96$

Esempi

1) $P[Z < z_\alpha] = 0.9953$
Dalle tavole $z_\alpha = 2.6$

2) $P[Z > z_\alpha] = 0.2743$
$P[Z > z_\alpha] = 1 - P[Z \leq z_\alpha]$
$\Rightarrow P[Z \leq z_\alpha] = 1 - 0.2743 = 0.7257$
Dalle tavole $z_\alpha = 0.6$

3) $P[0 \leq Z \leq z_\alpha] = 0.377$
$P[0 \leq Z \leq z_\alpha] = P[Z \leq z_\alpha] - 0.5$
$(0.5 = P[-\infty < Z < 0])$
$\Rightarrow P[Z \leq z_\alpha] = 0.5 + 0.377 = 0.877$
Dalle tavole $z_\alpha = 1.16$

4) $P[|Z| \leq z_\alpha] = 0.5762$
$P[|Z| \leq z_\alpha] = P[-z_\alpha \leq Z \leq z_\alpha] = 2P[0 \leq Z \leq z_\alpha] = 2\left(P[Z \leq z_\alpha] - 0.5\right) = 2P[Z \leq z_\alpha] - 1$
$\Rightarrow P[Z \leq z_\alpha] = \frac{0.5762+1}{2} = 0.7881$
Dalle tavole $z_\alpha = 0.8$

5) $P[z_\alpha < Z < 1.6] = 0.7865$
$P[z_\alpha < Z < 1.6] = P[Z < 1.6] - P[Z < z_\alpha]$
Dalle tavole $P[z < 1.6] = 0.9452$
$P[Z < z_\alpha] = 0.9452 - 0.7865 = 0.1587 < 0.5$!!!
z_α è negativo
Cerco il valore z_α^* positivo, simmetrico (a destra dell'origine) tale che

$$P[Z > z_\alpha^*] = 0.1587$$

$P[Z > z_\alpha^*] = 1 - P[Z \leq z_\alpha^*]$
$\Rightarrow P[Z \leq z_\alpha^*] = 1 - 0.1587 = 0.8413$
Dalle tavole $z_\alpha^* = 1 \rightarrow z_\alpha = -1$

Esercizi

1) La potenza W dissipata da una resistenza è proporzionale al quadrato della differenza di potenziale V ($V > 0$) ai suoi capi, cioè $W = kV^2$, con k una costante.
Calcolare $E[W]$ e $P[W > 120]$ se $k = 3$ e la variabile casuale $V \sim N(6, 1)$ (in opportune unità di misura).

$E[W] = E[3V^2] = 3E[V^2]$
ma per ogni variabile casuale X
$\text{var}[X] = E[X^2] - E[X]^2$
quindi
$E[V^2] = \text{var}[V] + E[V]^2 = 1 + 36 = 37 \Rightarrow$
$\Rightarrow E[W] = 111$

$P[W > 120] = P[V^2 > 40] = P[V > 2\sqrt{10}]$
$Z = \frac{V - \mu_V}{\sigma_V} = V - 6$
$P[V > 2\sqrt{10}] = P[Z > 2\sqrt{10} - 6] \sim$
$\sim P[Z > 0.3246]$
0.3246 non c'è sulle tavole, ma:
$0.32 \to \text{area} = 0.62552$
$0.33 \to \text{area} = 0.62930$
$0.3246 \sim 0.325 \to$ faccio la media tra le aree otte-

nendo 0.6274.

$\Rightarrow P[Z > 0.3246] = 1 - P[Z < 0.3246] \sim 1 - 0.6274 = 0.3726$

2) Una macchina produce resistori elettrici che devono avere una resistenza media di 40 ohm e una deviazione standard di 2 ohm. Se la resistenza si distribuisce come una normale, qual è la percentuale di resistori che avranno una resistenza superiore ai 43 ohm?
Per trovare la percentuale si moltiplica per 100% la frequenza relativa. La frequenza relativa, nel caso di un intervallo è pari alla probabilità di osservare un valore qualunque all'interno dell'intervallo.

$X \sim N(40, 2)$

$$Z = \frac{X - 40}{2}$$

$$P[X > 43] = P\left[Z > \frac{43-40}{2}\right] = P[Z > 1.5] =$$
$$= 1 - P[Z < 1.5] = 1 - 0.9332 = 0.0668$$

$\Rightarrow$ la percentuale è pari al 6.68%.

Determinare la percentuale di resistenze che superano 43 ohm se la resistenza misurata è all'ohm più vicino.
Questo problema è diverso dal precedente perchè si assegna una misura di 43 ohm ai resistori le cui resistenze sono > 42.5 e < 43.5.

$$P[X > 43.5] = P\left[Z > \frac{43.5-40}{2}\right] = P[Z > 2.75] =$$
$$= 1 - P[Z < 1.75] = 1 - 0.9599 = 0.0401$$

$\Rightarrow$ la percentuale è pari al 4.01%, cioè il 4.01% delle resistenze supera i 43 ohm quando misurate all'ohm più vicino.
La differenza $6.68\% - 4.01\% = 2.67\%$ rappresenta tutti i valori di resistenza > 43 ohm e < 43.5 ohm che vengono misurati come 43 ohm.

Approssimazioni

La distribuzione normale risulta spesso una buona approssimazione di una distribuzione discreta quando quest'ultima dimostra di avere una forma simmetrica a campana.

$$X \sim B(x; \mu = np, \sigma^2 = npq) \xrightarrow[\substack{n\to\infty \\ p\to 0}]{} N(x; \mu, \sigma^2) \xrightarrow[Z=\frac{X-np}{\sqrt{npq}}]{} N(0,1)$$

$E[X_N] = E[X_B] \Rightarrow \mu = np$
$\mathrm{var}[X_N] = \mathrm{var}[X_B] \Rightarrow \sigma^2 = npq$
È sufficiente che $npq \geq 10$ (altri testi indicano $np \geq 5$ e $nq \geq 5$) perchè l'approssimazione sia buona.

$$X \sim P(x; \lambda) \xrightarrow[\lambda\to\infty]{} N(x; \mu, \sigma^2) \xrightarrow[Z=\frac{X-\lambda}{\sqrt{\lambda}}]{} N(0,1)$$

$E[X_N] = E[X_P] \Rightarrow \mu = \lambda$
$\mathrm{var}[X_N] = \mathrm{var}[X_P] \Rightarrow \sigma^2 = \lambda$
È sufficiente che $\lambda \geq 10$.

Osservazione

Quando si approssima una distribuzione discreta con una continua di solito si applica la correzione di continuità, cioè si modifica l'intervallo di integrazione ampliando di 0.5 gli estremi dell'intervallo su cui si integra la continua per approssimare la discreta, ottenendo così una approssimazione migliore.

Esempi

1) Un test a scelta multipla è composto da 200 quesiti, ognuno dei quali ha 4 possibili risposte di cui

solo una è corretta. Sapendo che lo studente non conosce le risposte, qual è la probabilità che le risposte corrette siano tra 25 e 30, se risponde a 80 quesiti su 200.

$X \sim B(n = 80, p = \frac{1}{4})$

$$P[25 \leq X \leq 30] = P[X = 25] + \ldots + P[X = 30] =$$
$$= \sum_{x=25}^{30} \binom{80}{x} \left(\frac{1}{4}\right)^x \left(\frac{3}{4}\right)^{80-x} = 0.1193$$

$npq = 80 \cdot \frac{1}{4} \cdot \frac{3}{4} = 15 (> 10)$

Posso approssimare con la normale

$\mu = np = 20$

$\sigma^2 = npq = 15 \Rightarrow \sigma = 3.873$

$X \sim N(20, 15)$

Applicando la correzione di continuità l'intervallo diventa (24.5, 30.5).

$$P[24.5 \leq X \leq 30.5] = P\left[\frac{24.5 - 20}{3.873} < Z < \frac{30.5 - 20}{3.873}\right]$$
$$= P[1.16 < Z < 2.71] = \Phi(2.71) - \Phi(1.16) =$$
$$= 0.9966 - 0.8770 = 0.1196$$

2) La probabilità che un individuo guarisca da una malattia rara è 0.4. Se 100 persone hanno contratto

la malattia, qual è la probabilità che meno di 30 sopravvivano?
$n = 100 \quad p = 0.4 \quad q = 0.6 \quad npq = 24 > 10$
$\mu = 40 \quad \sigma^2 = 24 \quad \sigma = 4.899$
con la binomiale
$P[0 \leq X < 30] = P[0 \leq X \leq 29] =$
$= \sum_{x=0}^{29} \binom{100}{x} (0.4)^x (0.6)^{100-x} = 0.0148$
con la normale, applicando la correzione di continuità:

$$P[-0.5 \leq X \leq 29.5] = P\left[\frac{-0.5 - 40}{4.899} < Z < \frac{29.5 - 40}{4.899}\right]$$
$$= P[-8.26 < Z < -2.14] = P[2.14 < Z < 8.26] =$$
$$\Phi(8.26) - \Phi(2.14) =$$
$$= 1 - 0.9838 = 0.0162$$

<u>Nota bene</u>: per qualunque variabile casuale discreta X vale:
$P[X \leq x] = P[X \leq x + \frac{1}{2}]$
$P[x \leq X \leq y] = P[X \leq y + \frac{1}{2}] - P[X \leq x - \frac{1}{2}]$

5.2.3 Distribuzione gamma

Il suo nome deriva dalla funzione *gamma di Eulero*, così definita:

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} \mathrm{e}^{-x} dx \,, \qquad \alpha > 0.$$

La variabile casuale continua X ha una distribuzione gamma di parametri r e λ se la sua funzione di densità è data da:

$$f(x; r, \lambda) = \begin{cases} \dfrac{\lambda^r}{\Gamma(r)} x^{r-1} \mathrm{e}^{-\lambda x} & x \geq 0 \,, \\ 0 & \text{altrove} \,, \end{cases}$$

con $r > 0$ e $\lambda > 0$.

- $E[X] = \frac{r}{\lambda}$
- $\mathrm{var}[X] = \frac{r}{\lambda^2}$

Si noti che se $r = 1$ la densità gamma è la densità esponenziale.

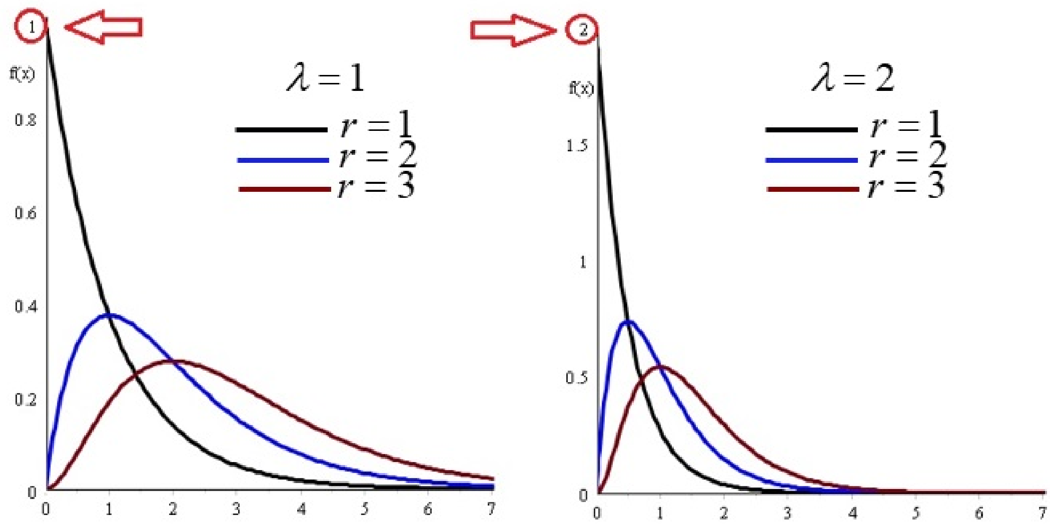

Figura 7: Distribuzioni gamma

5.2.4 Distribuzione esponenziale

Viene utilizzata nei problemi di affidabilità, nel modellare la durata della vita di un componente, il tempo al verificarsi di un guasto di sistemi elettronici. Una variabile casuale continua X è detta esponenziale se la sua funzione di densità è:

$$\exp(x;\lambda) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0\,, \\ 0 & \text{altrove}\,, \end{cases}$$

con $\lambda > 0$.

- $E[X] = \frac{1}{\lambda}$
- $\mathrm{var}[X] = \frac{1}{\lambda^2}$

- $$F(x) = \begin{cases} 1 - e^{-\lambda x} & x \geq 0\,, \\ 0 & \text{altrove}\,. \end{cases}$$

Assenza di memoria

Il tempo di vita di un componente è una variabile casuale esponenziale. Finchè il componente funziona, si comporta come se fosse nuovo, cioè l'usura dovuta al funzionamento nelle prime t ore iniziali non influenza la durata successiva di $(t + s)$ ore, cioè:

$$P[X > s + t | X > t] = P[X > s]$$

Esempi

1) La variabile casuale T, che descrive il tempo che deve trascorre, in anni, perchè un componente si guasti, è una variabile casuale esponenziale con media pari a 5.
Se 5 componenti vengono installati contemporaneamente, qual è la probabilità che almeno 2 componenti siano funzionanti alla fine dell'ottavo anno?

$E[T] = 5 \Rightarrow \lambda = \frac{1}{5}$

$$P[T > 8] = \int_8^\infty \frac{1}{5} e^{-\frac{1}{5}x} dx = e^{\frac{8}{5}} \sim 0.2$$

rappresenta la probabilità che il componente funzioni dopo 8 anni.
Adesso bisogna applicare la binomiale:

$$P[X \geq 2] = 1 - P[X < 2] = 1 - P[X = 0] - P[X = 1]$$
$$1 - \binom{5}{0}(0.2)^0(0.8)^5 - \binom{5}{1}(0.2)^1(0.8)^4 =$$
$$1 - (0.8)^5 - (0.8)^4 = 1 - 0.73728 = 0.2527$$

2) Il tempo (in anni) che trascorre prima che una lavatrice necessiti di una riparazione è una variabile casuale esponenziale di parametro $\lambda = \frac{1}{4}$.
Calcolare la probabilità che sia necessaria una riparazione nel 1° anno e la probabilità che sia necessaria una riparazione prima del 6° anno.
$P[X \leq 1] = 1 - e^{-\lambda x}\Big|_{x=1, \lambda=\frac{1}{4}} = 1 - e^{-\frac{1}{4}} = 0.221$
$P[X \leq 6] = 1 - e^{-\lambda x}\Big|_{x=6, \lambda=\frac{1}{4}} = 1 - e^{-\frac{3}{2}} = 0.777$
la seconda probabilità è alta: si può affermare che il prodotto non è vantaggioso.

5.2.5 Distribuzione χ^2 (chi-quadro)

È un caso speciale della distribuzione gamma, per $\lambda = \frac{1}{2}$ ed $r = \frac{\nu}{2}$, dove $\nu \in \mathbb{N}$ è chiamato *grado di libertà*. Ha un ruolo fondamentale nell'inferenza statistica. La variabile casuale χ^2 ha come funzione di densità:

$$\chi^2(x,\nu) = \begin{cases} \dfrac{1}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})} x^{\frac{\nu}{2}-1} \mathrm{e}^{-\frac{x}{2}} & x \geq 0\,, \\ 0 & \text{altrove}\,. \end{cases}$$

- $E[X] = \nu$
- $\mathrm{var}[X] = 2\nu$

(grafico $\rightarrow$ vedi la densità gamma).

5.2.6 Distribuzione di Student t

È simile alla normale standard, ma ha le code più pesanti. Interviene nella stima della media di una popolazione caratterizzata da una distribuzione normale. La variabile casuale t di Student ha la seguente funzione di densità:

$$t(x,v) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} \cdot \frac{1}{\sqrt{\pi\nu}} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \quad x \in \mathbb{R},$$

con $\nu \in \mathbb{N}$ chiamato grado di libertà.

- $E[X] = 0$ se $\nu > 1$
- $\mathrm{var}[X] = \frac{\nu}{\nu-2}$ se $\nu > 2$

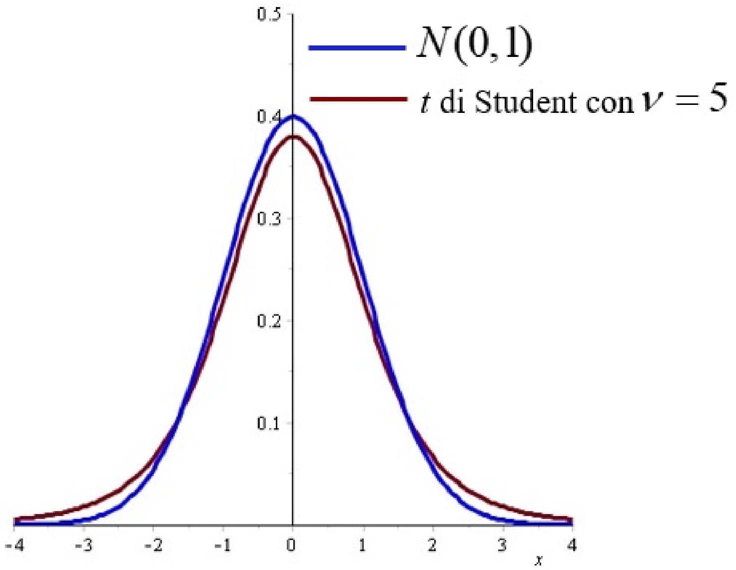

Figura 8: Distribuzione t di Student

<u>Esercizi</u>

1) Calcolare il percentile della χ^2 per $\nu = 5$ tale che:

$$P[\chi^2 > \chi^2_{5;0.025}] = 0.025$$

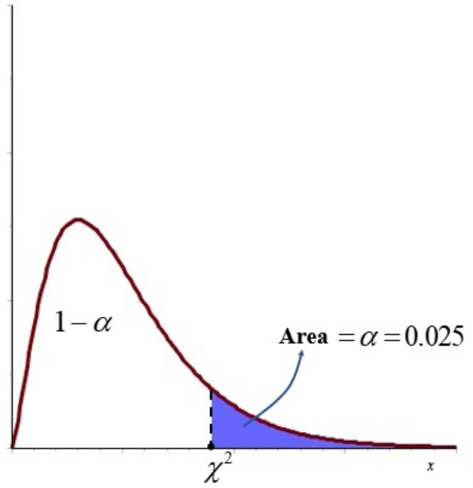

Figura 9: Distribuzione chi-quadro

$\chi^2_{5,0.025} = 12.83249$
$1 - \alpha = 0.975$
2) Calcolare il percentile della t di Student tale che per $\nu = 24$

$$P[T > t_{24,0.05}] = 0.05$$

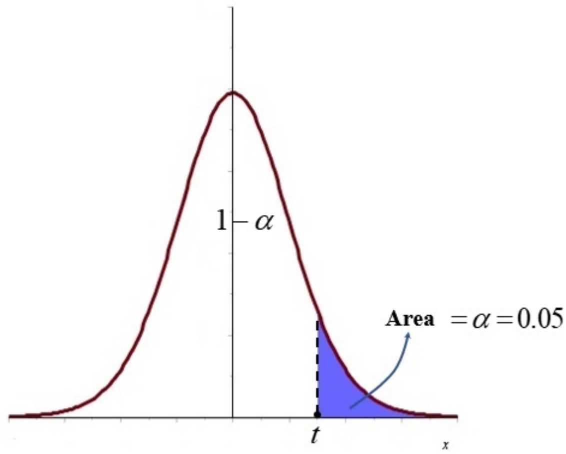

Figura 10: Distribuzione t

$1 - \alpha = 0.95$
$t_{24,0.05} = 1.71088$
3) Calcolare il percentile della t di Student tale che per $\nu = 24$

$$P[|T| > t_{24,0.05}] = 0.05$$

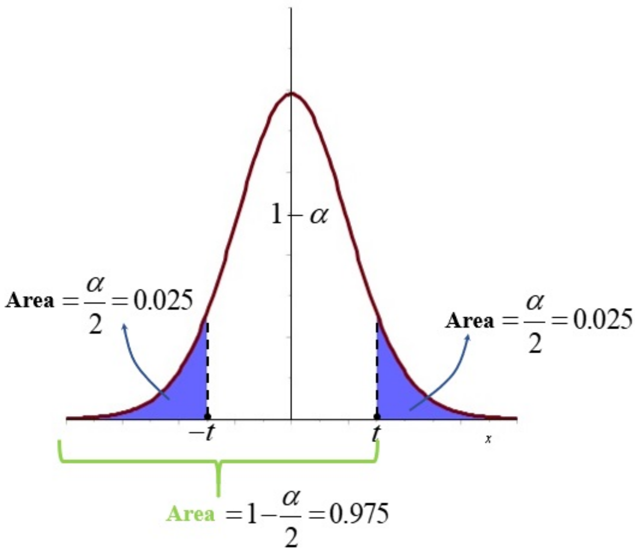

Figura 11: Distribuzione t

$P[|T| < t_{24,0.05}] = 0.95$
ma anche $P[T < t_{24,0.05}] = 0.975$
quindi $t_{24,0.05} = 2.06390$