

# Assiomi di Kolmogorov

Sia  $\Omega$  uno spazio campionario. Una funzione  $P$  che associa ad ogni sottoinsieme di  $\Omega$  un numero reale è detta probabilità se soddisfa le seguenti proprietà:

1.  $P(A) \geq 0$  per ogni sottoinsieme  $A$  di  $\Omega$
2.  $P(\Omega) = 1$
3. Sia  $A, B \in \mathcal{W}$  *disgiunti* t.c  $A \cap B = \emptyset$  allora:  
 $P(A \cup B) = P(A) + P(B)$

## Calcolo dell'intersezione di probabilità

### Probabilità Congiunta

La ragione per questa formula è che la probabilità di due eventi indipendenti che si verificano insieme si ottiene restringendo l'universo degli eventi possibili dopo che il primo evento è avvenuto. Se la probabilità del primo evento è  $P(A)$ , allora la probabilità che anche il secondo evento  $B$  si verifichi è ancora  $P(B)$  perché i due eventi sono indipendenti. Così, la probabilità che entrambi si verifichino è il prodotto delle singole probabilità. Per calcolare la probabilità dell'intersezione di due eventi possiamo usare la formula:

$$P(A \cap B) = \begin{cases} P(A) \cdot P(B) & \text{A,B Indipendenti} \\ P(A|B) \cdot P(B) = P(B|A) \cdot P(A) & \text{Caso GENERALE} \end{cases}$$

## Calcolo per l'unione di probabilità

$$P(A \cup B) = \begin{cases} P(A) + P(B) & \text{A,B indipendenti} \\ P(A) + P(B) - P(A \cap B) & \text{caso GENERALE} \end{cases}$$

## Indipendenza di eventi

Due eventi si dicono indipendenti se la probabilità che si verifichi uno non influenza la probabilità che si verifichi l'altro, inoltre l'intersezione dei due è nulla poichè sono indipendenti. Quindi se due eventi sono indipendenti allora:

$$0 = P(A \cap B) = P(A) \cdot P(B)$$

Quindi se volessimo calcolare  $P(A|B)$  sapendo che sono indipendenti tra di loro avremmo:

**importante**

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = 0$$

Questo perchè come abbiamo visto sopra sono disgiunti, è logico anche perchè essendo separati non possono avere elementi comuni nella probabilità.  
*es se esce 6 in un dado, mi annulla la probabilità che possa uscire il 4*

## Teorema di Bayes

$$P(B) = \sum_{i=1}^n P(B|E_i) \cdot P(E_i)$$

$$P(B) = P(B|A) \cdot P(A) + P(B|A^c) \cdot P(A^c)$$

Quindi ad esempio se stiamo esaminando un pezzo che è prodotto da cinque macchine diverse e vogliamo calcolare la probabilità che il pezzo sia prodotto da una macchina in particolare, possiamo usare il teorema di Bayes sapendo le probabilità che il pezzo sia prodotto da ciascuna macchina.

## Variabile aleatorie

### Discrete

La funzione di probabilità di una variabile aleatoria discreta è definita come:

$$p_x(x_i) = P(X = x_i)$$

La funzione cumulativa di una variabile aleatoria discreta è definita come:

$$F_x(x_i) = P(X \leq x_i) = \sum_{x_j \leq x_i} p_x(x_j)$$

Quindi somma le probabilità di tutti gli eventi minori o uguali a  $x_i$ .

Il valore atteso di una variabile aleatoria discreta è definito come:

$$E[X] = \sum_{i=1}^n x_i \cdot p_x(x_i)$$

La varianza di una variabile aleatoria discreta è definita come:

$$Var[X] = E[(X - E[X])^2] = \sum_{i=1}^n (x_i - E[X])^2 \cdot p_x(x_i)$$

## Continue

$$P(a \leq X \leq b) = \int_a^b f_x(x) dx$$

La funzione cumulativa di una variabile aleatoria continua è definita come:

$$F_x(x) = P(X \leq x) = \int_{-\infty}^x f_x(t) dt$$

Quindi somma le probabilità di tutti gli eventi minori o uguali a  $x$ .

la probabilità di  $X = x$  è sempre 0 (integrale da a-a)

Il valore atteso di una variabile aleatoria continua è definito come:

$$E[X] = \int_{-\infty}^{\infty} x \cdot f_x(x) dx$$

La varianza di una variabile aleatoria continua è definita come:

$$Var[X] = E[(X - E[X])^2] = \int_{-\infty}^{\infty} (x - E[X])^2 \cdot f_x(x) dx$$

**NB:** per semplificare il calcolo della varianza possiamo usare la formula:

$$Var[X] = E[X^2] - E[X]^2$$

## Distribuzioni

# Formulario distribuzioni

		$\mathbb{E}$	$Var$	Funzione di probabilità		Note
Bernulli	$B(p)$	$p$	$p(1 - p)$	$p^k(1 - p)^{1-k}$		-
Binomiale	$B(n, p)$	$np$	$np(1 - p)$	$\binom{n}{k} p^k (1 - p)^{n-k}$		-
Geometrica	$G(p)$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	$(1 - p)^{k-1} p$		parte da 1

		$\mathbb{E}$	$Var$	Funzione di probabilità		Note
Poisson - Binom	$P(\lambda)$	$\lambda$	$\lambda$	$\frac{e^{-\lambda} \lambda^k}{k!}$	$\lambda = np$	$n > 50, np < 5$
Poisson - temp	$P(\lambda t)$	$\lambda$	$\lambda$	$\frac{e^{-\lambda} \lambda^k}{k!}$	$\lambda$ =successi su unità di tempo, $t$ : tempo esercizio (prob 2 chiamate nei prox 4min(t=4))	es 3 chiam ogni 5 min: $\lambda = \frac{3}{5}$
Esponenziale	$Exp(\lambda)$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$P(X \leq x) = 1 - e^{-\lambda x}$	$S(X > x) = e^{-\lambda x}$	Soppravv - guasto dopo t
Normale	$N(\mu, \sigma^2)$	$\mu$	$\sigma^2$	$Z = \frac{X-\mu}{\sigma}$		Normalizzazione
Normale per Bernulli	$N(np, np(1 - p))$	$\mu$	$\sigma^2$	$Z = \frac{X-np}{\sqrt{np(1-p)}}$		$>, \leq \rightarrow x + 0.5$ $\geq, < \rightarrow x - 0.5$ <i>esattamente</i> $\rightarrow x \pm 0.5$
---	---	---	---	---	---	---
Varianza	$S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$	Deviazione standard: $S = \sqrt{S^2}$				

**Nota:**  
Deviazione standard  $\rightarrow \sigma$   
Varianza  $\rightarrow \sigma^2$

## Variabile esponenziale

$$X \sim Exp(\lambda)$$

La distribuzione esponenziale è una distribuzione di probabilità continua che descrive il tempo tra gli eventi in un processo di Poisson.

La **Funzione di sopravvivenza** che descrive la probabilità che un oggetto si guasti in un intervallo  $[0, t]$  è data da:

$$P(X \leq x) = 1 - e^{-\lambda x}$$

La probabilità che si **guasti dopo un tempo  $t$**  è data da:

$$P(X > t) = e^{-\lambda t} = S(x)$$

## Proprietà assenza di memoria

La distribuzione esponenziale gode della proprietà di assenza di memoria, cioè la probabilità che un evento si verifichi nei prossimi  $t$  minuti è indipendente dal tempo trascorso finora.

$$P(X > x_0 + x | X > x_0) = P(X > x)$$

Questa equazione ci dice quindi che anche se sappiamo che è passato un tempo  $x_o$  non condiziona la probabilità, ma rimane sempre la stessa.

Esempio:  
Quale è la probabilità che la riparazione richieda almeno 3 ore sapendo che ne richiede più di 2?

$$P(X > 3 | X > 2) = P(X > 1) = e^{-\lambda}$$

Quindi la probabilità è la stessa.  
**Importante:** quindi quando calcoliamo la probabilità  $x = |x - x_0|$ .

Media campionaria

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

Valore atteso della media campionaria:

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

Varianza della media campionaria:

$$Var[\bar{X}] = \frac{\sigma^2}{n}$$

Stima intervallare

Descrizione		Unilaterale	Bilaterale dx	Bilaterale sx
Stimare la media con varianza nota	$\mu$	$(\bar{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}})$	$(\bar{X} - z_{\alpha} \frac{\sigma}{\sqrt{n}}, \infty)$	$(-\infty, \bar{X} + z_{\alpha} \frac{\sigma}{\sqrt{n}})$
Stimare la media con varianza sconosciuta	$\mu$	$(\hat{x} - t_{\frac{\alpha}{2}, n-1} \frac{\hat{s}}{\sqrt{n}} < \mu < \hat{x} + t_{\frac{\alpha}{2}, n-1} \frac{\hat{s}}{\sqrt{n}})$	$(\hat{x} - t_{\alpha, n-1} \frac{\hat{s}}{\sqrt{n}}, \infty)$	$(-\infty, \hat{x} + t_{\alpha, n-1} \frac{\hat{s}}{\sqrt{n}})$
Stimare la varianza	$\sigma^2$	$(\frac{(n-1)\hat{s}^2}{\chi_{\frac{\alpha}{2}, n-1}^2} < \sigma^2 < \frac{(n-1)\hat{s}^2}{\chi_{1-\frac{\alpha}{2}, n-1}^2})$	$(\frac{(n-1)\hat{s}^2}{\chi_{\alpha, n-1}^2}, \infty)$	$(0, \frac{(n-1)\hat{s}^2}{\chi_{1-\alpha, n-1}^2})$
Stimare la media di una Bernulli	$p$	$(\hat{p} - z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}})$		

Test d'ipotesi

Tabella riassuntiva

Casistica		$H_0$	$H_1$	Statistica test(st)	Rifiuto $H_0$ se
Test per la media di una popolazione normale con varianza nota		$\mu = \mu_0$	$\mu \neq \mu_0$		$ st  > z_{\frac{\alpha}{2}}$
		$\mu \leq \mu_0$	$\mu > \mu_0$	$st = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$	$st > z_{\alpha}$
		$\mu \geq \mu_0$	$\mu < \mu_0$		$st < -z_{\alpha}$
Test per la media di una popolazione normale con varianza incognita		$\mu = \mu_0$	$\mu \neq \mu_0$		$ st  > t_{\frac{\alpha}{2}, n-1}$
		$\mu \leq \mu_0$	$\mu > \mu_0$	$st = \frac{\bar{X} - \mu_0}{\frac{\hat{s}}{\sqrt{n}}} \sim t_{n-1}$	$st > t_{\alpha, n-1}$
		$\mu \geq \mu_0$	$\mu < \mu_0$	con $\hat{s}$ deviazione standard(sd)	$st < -t_{\alpha, n-1}$
Test per la varianza di una popolazione normale con media e varianza incognite		$\sigma^2 = \sigma_0^2$	$\sigma^2 \neq \sigma_0^2$		$st > \chi_{\frac{\alpha}{2}, n-1}^2$ o $st < \chi_{1-\frac{\alpha}{2}, n-1}^2$
		$\sigma^2 \leq \sigma_0^2$	$\sigma^2 > \sigma_0^2$	$st = \frac{(n-1)\bar{S}^2}{\sigma_0^2} \sim \chi_{n-1}^2$	$st > \chi_{\alpha, n-1}^2$
		$\sigma^2 \geq \sigma_0^2$	$\sigma^2 < \sigma_0^2$	con $\bar{S}^2$ varianza campionaria(var)	$st < \chi_{1-\alpha, n-1}^2$

Casistica		$H_0$	$H_1$	Statistica test( $st$ )	Rifiuto $H_0$ se
Test asintotici per la media di una popolazione di Bernoulli		$p = p_0$	$p \neq p_0$		$ st  > z_{\frac{\alpha}{2}}$
		$p \leq p_0$	$p > p_0$	$st = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \sim N(0, 1)$	$st > z_\alpha$
		$p \geq p_0$	$p < p_0$	con $\hat{p}$ calcolato come $\frac{1}{n} \sum_{i=1}^n (X_i)$	$st < -z_\alpha$

Occhio ai segni dei vari quantili nei casi unilaterali

## Calcolo p-value

Bilaterale	Laterale sinistro( <i>minore di</i> )	Laterale destro( <i>maggiore di</i> )
$2 \cdot \mathbb{P}(Z >  st )$	$\mathbb{P}(Z < st)$	$\mathbb{P}(Z > st)$

## Test due popolazioni

# Tabella di tutti i test

Test	$H_0$	$H_1$	$ST$	Rifiuto $H_0$ se
<b>Confronto delle medie di due popolazioni normali, caso di varianze note</b>	$\mu_x = \mu_y$	$\mu_x \neq \mu_y$		$ st  > z_{\frac{\alpha}{2}}$
	$\mu_x \leq \mu_y$	$\mu_x > \mu_y$	$st = \frac{\overline{X}-\overline{Y}}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}} \sim N(0, 1)$	$st > z_\alpha$
	$\mu_x \geq \mu_y$	$\mu_x < \mu_y$		$st < -z_\alpha$
<b>Confronto delle medie di due popolazioni normali, caso di varianze non note, MA UGUALI</b>	$\mu_x = \mu_y$	$\mu_x \neq \mu_y$		$ st  > t_{\frac{\alpha}{2}, n+m-2}$
	$\mu_x \leq \mu_y$	$\mu_x > \mu_y$	$st = \frac{\overline{X}-\overline{Y}}{\sqrt{S_p^2(\frac{1}{n} + \frac{1}{m})}} \sim t_{n+m-2}$	$st > t_{\alpha, n+m-2}$
	$\mu_x \geq \mu_y$	$\mu_x < \mu_y$		$st < -t_{\alpha, n+m-2}$
<b>Confronto delle medie di due popolazioni normali, caso di varianze non note e DIVERSE</b>	$\mu_x = \mu_y$	$\mu_x \neq \mu_y$		$ st  > z_{\frac{\alpha}{2}}$
	$\mu_x \leq \mu_y$	$\mu_x > \mu_y$	$st = \frac{\overline{X}-\overline{Y}}{\sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{m}}} \sim N(0, 1)$	$st > z_\alpha$
	$\mu_x \geq \mu_y$	$\mu_x < \mu_y$		$st < -z_\alpha$
<b>Confronto delle medie di due popolazioni normali nel caso di campioni accoppiati</b>	$\mu_w = \mu_x - \mu_y$		i campioni diventano $w = x_i - y_i$ quindi si procede come una distribuzione sopra.	$st > t_{\alpha, n-1}$

Quando costruiamo i test per le popolazioni accoppiate **ricordiamo che**:

- $\mu_w = 0 \iff \mu_x = \mu_y$
- $\mu_w < 0 \iff \mu_x < \mu_y$
- $\mu_w > 0 \iff \mu_x > \mu_y$

$$\bar{w} = \frac{1}{n} \sum_{i=1}^n w_i$$

La deviazione standard campionaria è:

$$\bar{s}_w = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (w_i - \bar{w})^2}$$

La varianza combinata campionaria si calcola come:

$$S_p^2 = \frac{(n-1)S_x^2 + (m-1)S_y^2}{n+m-2}$$

Nel caso le ampiezze dei campioni siano uguali si può usare la varianza campionaria comune:

$$S_p^2 = \frac{(S_x^2 + S_y^2)(n-1)}{2(n-1)}$$

## Regressione lineare

## Tabella formulario

$$Yy = \beta_0 + \beta_1x + \epsilon$$

Dove:

- $y$  è la variabile dipendente, detta di RISPOSTA
- $x$  è la variabile indipendente, detta di INPUT
- Due **coefficienti costanti di regressione**:
  - $\beta_0$  è l'intercetta, ovvero il valore di  $Y$  quando  $X = 0$
  - $\beta_1$  è il coefficiente angolare, ovvero la variazione di  $Y$  per unità di variazione di  $X$
- $\epsilon$  è l'errore casuale, con media 0

nota: da qui in poi assumiamo che l'errore causale abbia distribuzione normale con media zero e varianza  $\sigma^2$ .

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sum_{i=1}^n (x_i)^2 - n \bar{x}^2}$$
$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$H_0$	$H_1$	Statistica di test	Rifiuto $H_0$ se
$\beta_1 = 0$	$\beta_1 \neq 0$	$st = \sqrt{\frac{(n-2)S_{XX}}{SS_R}} \cdot \beta_1$ con $S_{XX} = \sum_{i=1}^n (x_i)^2 - n \bar{x}^2$ e $SS_R = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2$	$ st  > t_{\frac{\alpha}{2}, n-2}$

$$S_{YY} = \sum_{i=1}^n (y_i - \bar{y})^2$$

## Variabilità

$$S_{YY} = \underbrace{SS_R}_{\text{Varianza Residua}} + \underbrace{(S_{YY} - SS_R)}_{\text{Varianza Spiegata}}$$

## Coefficiente di determinazione

$$R^2 = \frac{S_{YY} - SS_R}{S_{YY}} \in [0, 1]$$

Se  $R^2$  è vicino a 1 allora il modello è buono, altrimenti se è vicino a 0 allora il modello non è buono.

## Formule varie

Varianza:

$$S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

Deviazione standard:

$$S = \sqrt{S^2}$$

Stimatore di massima verosimiglianza per la varianza di una popolazione normale:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$