Parte 1. Inferenza parametrica

1. Probabilità

Alcune formule e proprietà utili:
Funzione di ripartizione: $F_X = P(X \leq x)$
Densità: $f_X(x) = F'_X = P(X = x) \quad \forall x \in \mathbb{R}$
$P(X > x) = 1 - P(X \leq x)$
$P(x < X \leq y) = P(X \leq y) - P(X \leq x)$

1.1. Proprietà del valore atteso.

(1) Se $P(X = c) = 1$ allora $E(X) = c$
(2) $\mathbb{E}(aX) = a\mathbb{E}(X)$
(3) $\mathbb{E}(X + a) = \mathbb{E}(X) + a$
(4) $\mathbb{E}(g(X) + h(X)) = \mathbb{E}(g(X)) + \mathbb{E}(h(X))$ se h e g sono funzioni tali che $\mathbb{E}(g(X))$ e $\mathbb{E}(h(X))$ esistono
(5) $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ se X e Y sono indipendenti.

1.2. Proprietà della varianza.

(1) $Var(aX) = a^2 Var(X)$
(2) $Var(X + \beta) = Var(X)$ con $\beta \in \mathbb{R}$
(3) $Var(V) = \mathbb{E}(V^2) - \mathbb{E}^2(V) = \mathbb{E}[(V - \mu)^2]$ con $\mu := \mathbb{E}(V)$
(4) $Var(X + Y) = Var(X) + Var(Y) + 2\,Cov(X, Y)$ con $Cov(X, Y) = \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))]$.
Nel caso di variabili X e Y indipendenti, si riduce a $Var(X + Y) = Var(X) + Var(Y)$.
(5) Se $X = cost$ allora $Var(X) = 0$.

1.3. Proprietà della normale.

(1) Se $X \sim N(\mu_X, \sigma_X^2)$ e $Y \sim N(\mu_Y, \sigma_Y^2)$ e $X \perp Y$ allora $aX + bY \sim N(a\mu_X + b\mu_Y,\, a^2\sigma_X^2 + b^2\sigma_Y^2)$
(2) Se $X \sim N(\mu, \sigma^2)$ allora $aX + b \sim N(a\mu + b, a^2\sigma^2)$

1.4. Proprietà della normale standard $N(0, 1)$.

$\phi : \mathbb{R} \to [0, 1]$ è la funzione di ripartizione della normale standard
$z : [0, 1] \to \mathbb{R}$ è il quantile della normale standard, cioè la funzione opposta di ϕ.

(1) $\phi(-x) = 1 - \phi(x)$

2. Momenti

Definition 2.1. Data una variabile aleatoria X, il momento n-esimo di X è il numero reale

$$\mu_n := \mathbb{E}(X^n)$$

Remark. Il momento primo equivale al valor medio di X: $\mu_1 = \mathbb{E}(X)$.
Il momento secondo e il momento primo assieme definiscono la varianza: $\mathrm{Var}(X) = \mu_2 - (\mu_1)^2 = \mathbb{E}(X^2) - \mathbb{E}(X)^2$

Una distribuzione di probabilità è completamente determinata dai suoi momenti

Definition 2.2. Sia X una variabile aleatoria. La *funzione generatrice dei momenti* M_X di X è definita come

$$M_X(t) := \mathbb{E}(e^{tX}) = \int e^{tx} d\mathbb{P}_X(x)$$

per tutti i valori di t per cui l'espressione ha senso.
$d\mathbb{P}_X(x)$ è la densità di probabilità di X.

Proposition 2.3. La funzione generatrice dei momenti prende questo nome perchè a partire da essa è possibile ottenere (per differenziazione nel punto $t = 0$) tutti i momenti di X secondo la formula

$$\mu_n = \mathbb{E}(X^n) = \left[\left(\frac{d}{dt}\right)^n M_X(t)\right]_{t=0}$$

sotto la condizione che $\mathbb{E}(e^{\varepsilon|X|})$ esista per qualche $\varepsilon > 0$ (se questa condizione vale, esistono tutti i momenti di X).

Definition 2.4. Se X e Y sono variabili aleatorie indipendenti e $S = X + Y$ allora $M_S(t) = M_X(t)M_Y(t)$ per ogni t per cui il membro a destra ha senso.

3. Famiglia delle densità gamma

Definition 3.1. Si dice che una variabile aleatoria X ha **densità gamma di parametri** a, β (entrambi > 0) e si scrive $X \sim \Gamma(a, \beta)$ se la funzione di ripartizione della variabile è

$$f(x, a, \beta) = \frac{(1/\beta)^a}{\Gamma(a)} e^{-x/\beta} x^{a-1} \mathbf{1}_{(0,\infty)}(x)$$

In particolare, $\Gamma(a)$ assume la forma data dalla definizione seguente.

Definition 3.2. L'*integrale gamma* $\Gamma(a)$ è

$$\Gamma(a) = \int_0^\infty e^{-x} x^{a-1} dx, \quad a > 0$$

Note. La notazione $\Gamma(a)$ si riferisce all'integrale gamma, mentre $\Gamma(a, \beta)$ alla densità gamma.

Proprietà di $\Gamma(a)$:

(1) Valori particolari: $\Gamma(1) = 1$ e $\Gamma(1/2) = \sqrt{\pi}$
(2) Derivando per parti $\Gamma(a+1)$, si ottiene:$\Gamma(a+1) = a\Gamma(a)$
(3) Se a è un numero naturale $n \geq 1$, allora $\Gamma(n+1) = n! \quad \forall n \in \mathbb{N}$

Proposition 3.3. $X \sim \Gamma(a, \beta)$ ha **funzione generatrice dei momenti**

$$M(t) = E\left(e^{tX}\right) = \frac{1}{(1-\beta t)^a}$$

Si ha inoltre $\mathbb{E}(X) = a\beta$, $\mathbb{E}(X^2) = a(a+1)\beta^2$ e $\mathrm{Var}(X) = a\beta^2$

3.1. Proprietà di $\Gamma(a, \beta)$.

Sia $X \sim \Gamma(a, \beta)$.

(1) Se $c > 0$ e $Y = cX$ allora $Y \sim \Gamma(a, c\beta)$
(2) Se X e Y sono v.a. indipendenti e $Y \sim \Gamma(c, \beta)$ allora $X + Y \sim \Gamma(a + c, \beta)$
(3) Se $c > 0$ vale anche l'inverso della precendente, cioè da $Y \sim \Gamma(c, \beta)$ e $X+Y \sim \Gamma(a+c, \beta)$ si può ricavare $X \sim \Gamma(a, \beta)$.

3.2. La distribuzione esponenziale.

La densità *esponenziale* $Exp(\beta) = \Gamma(1, \beta)$ è un caso particolare della densità gamma. Assume la seguente forma:

$$Exp(\beta) = \frac{1}{\beta} \exp\left\{-\frac{x}{\beta}\right\} \mathbb{I}_{(0,+\infty)}(x)$$

e la sua FDR è

$$F_X(x) = P(X \leq x) = 1 - e^{-\frac{y}{\beta}}$$

dove β è il parametro che caratterizza l'esponenziale.

3.3. Distribuzione chi quadro.

La densità *chi quadro a n gradi di libertà* è un sottocaso della distribuzione gamma: $\chi^2_n = \Gamma(\frac{n}{2}, 2)$.
Già sappiamo che se $X \sim N(0, 1)$ allora $X^2 \sim \chi^2_1$. Si dimostra inoltre che, in tal caso, $\sum_{i=1}^n X_i^2 \sim \chi^2_n$

3.4. Distribuzione F di Fisher.

Si supponga di avere U e V variabili aleatorie con distribuzione χ^2 con rispettivamente m e n gradi di libertà. Se U e V sono statisticamente indipendenti, la statistica

$$\frac{U/m}{V/n}$$

ha distribuzione F con m gradi al numeratore e n gradi al denominatore, la cui densità è

$$f_F(x) = \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \left(\frac{m}{n}\right)^{m/2} x^{\frac{m}{2}-1} \left(1 + \frac{m}{n}x\right)^{-\frac{m+n}{2}}$$

Il suo grafico assomiglia a quello di una χ^2 ma con picco più alto e più schiacciato lungo l'asse x.
Se $X_1, \ldots, X_m \sim N(\mu_X, \sigma_X^2)$ e $Y_1, \ldots, Y_n \sim N(\mu_X, \sigma_X^2)$ sono indipendenti e S^2 è lo stimatore della varianza, allora

$$\frac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2}$$

ha distribuzione F con $m-1$ gradi di libertà al numeratore e $n-1$ al denominatore.

3.4.1. *Proprietà.* $V \sim F_{m,n} \quad \Rightarrow \quad \frac{1}{V} \sim F_{n,m}$

4. Funzione di verosimiglianza

Definition 4.1. La *funzione di verosimiglianza (Likelihood function)* di n variabili aleatorie $X_1, \ldots, X_n$ è data dalla funzione di densità congiunta di $X_1, \ldots, X_n$ considerata come funzione di θ. Se $X_1, \ldots, X_n$ è un campione casuale estratto dalla densità $f(x, \theta)$, $\theta \in \Theta$, la funzione di verosimiglianza è

$$\theta \mapsto L_\theta(x_1, \ldots, x_n) = \prod_{j=1}^{n} f(x_j, \theta)$$

5. Stimatori

Definition 5.1. Siano $X_1, \ldots, X_n$ *i.i.d.* $\sim f(x, \theta)$, $\theta \in \Theta$, e $\kappa(\theta)$ una caratteristica della popolazione. Uno *stimatore* di $\kappa(\theta)$, basato sul campione $X_1, \ldots, X_n$ è una statistica $T = g(X_1, \ldots, X_n)$ usata per stimare $\kappa(\theta)$. Il valore assunto da uno stimatore T di $\kappa(\theta)$ è detto stima di $\kappa(\theta)$.

Uno stimatore, quindi, è una statistica che permette di stimare una quantità a partire dalla sola conoscenza dei campioni.

Definition 5.2. Si dice *distorsione* (bias) di uno stimatore il valore atteso dell'errore commesso nella stima

$$bias(T) = \mathbb{E}(T - \theta) = \mathbb{E}(T) - \theta$$

Perciò uno stimatore con *bias* pari a zero si dice non distorto:

Definition 5.3. Una statistica T che ammette media per ogni θ in Θ è detta *stimatore non distorto* o corretto (unbiased) della caratteristica $\kappa(\theta)$ se

$$\mathbb{E}_\theta(T) = \kappa(\theta) \quad \forall \theta \in \Theta$$

La media campionaria è stimatore non distorto della media teorica. La varianza campionaria è stimatore non distorto della varianza teorica.

Note 5.4. Combinazioni lineari di stimatori non distorti danno origine a stimatori non distorti.

La qualità di uno stimatore è misurata tramite il suo errore quadratico medio:

Definition 5.5. Si definisce *errore quadratico medio* (Mean Square Error) il valore

$$MSE := \mathbb{E}\left[(T - \theta)^2\right]$$

che, tramite le proprietà del valore atteso e della varianza[1], si dimostra essere

$$MSE = Var(T) + bias^2(T)$$

In particolare, se T è uno stimatore non distorto, si ha

$$MSE(T) = Var(T)$$

Definition 5.6. Si dice *consistente in media quadratica* uno stimatore T_n di $\kappa(\theta)$ il cui MSE tende a 0 al crescere del numero di campioni, cioè tale che

$$\lim_{n \to \infty} \mathbb{E}[(T_n - \kappa(\theta))^2] = 0 \quad \forall \theta \in \Theta$$

dove n è il numero di campioni.

Dal punto di vista pratico, per verificare la consistenza in media quadratica è conveniente verificare le due seguenti condizioni:

$$\lim_{n \to \infty} \mathbb{E}_\theta(T_n) = \kappa(\theta)$$

$$\lim_{n \to \infty} Var_\theta(T_n) = 0$$

Definition 5.7. Sia $X_1, \ldots, X_n$ una successione di variabili aleatorie i.i.d. con comune densità $f(x, \theta)$ con $\theta \in \Theta$ e sia T_n una statistica funzione solo delle n osservazioni. La successione $\{T_n\}_n$ è *asintoticamente gaussiana* con media asintotica $\mu_n(\theta)$ e varianza asintotica $\sigma_n^2(\theta)$ se

$$\lim_{n \to \infty} P\left(\frac{T_n - \mu_n(\theta)}{\sigma_n} \leq z\right) = \phi(z) \quad \forall z \in \mathbb{R}$$

Questa proprietà è utile nel caso di grandi campioni, per poter approssimare lo stimatore T_n con una gaussiana.

---

[1] $\mathbb{E}[(T-\theta)^2] = \mathbb{E}[T^2 + \theta^2 - 2T\theta] = \mathbb{E}[T^2] + \theta^2 - 2\theta\mathbb{E}[T] + \mathbb{E}^2[T] - \mathbb{E}^2[T] = \underbrace{\mathbb{E}[T^2] - \mathbb{E}^2[T]}_{Var(T)} + \underbrace{\mathbb{E}^2[T] + \theta^2 - 2\theta\mathbb{E}[T]}_{(\mathbb{E}[T]-\theta)^2 = bias^2(T)} = Var(T) - bias^2(T)$

5.1. Stimatori a massima verosimiglianza.

Definition 5.8. Siano $X_1, \dots, X_n$ un campione casuale con funzione di verosimiglianza L_θ, $\theta \in \Theta$, $x_1, \dots, x_n$ una realizzazione campionaria e $g(x_1, \dots, x_n)$ un valore in Θ tale che

$$L_{g(x_1,\dots,x_n)}(x_1, \dots, x_n) = \max_{\theta\in\Theta} L_\theta(x_1, \dots, x_n)$$

La statistica $\hat{\theta} = g(X_1, \dots, X_n)$ è detta *stimatore di massima verosimiglianza di* θ. Per indicare $\hat{\theta}$ useremo l'acronimo ML (*Maximum Likelihood*) o MLE (*Maximum Likelihood Estimator*).

Generalmente, per semplificare alcuni conti, quando si calcola uno stimatore a massima verosimiglianza si preferisce introdurre il logaritmo di L.

Lo stimatore risulta quindi essere $\hat{\theta} : \frac{\partial \log L}{\partial \theta} = 0$

Lo stimatore di una caratteristica $\kappa(\theta)$ dipendente dalla quantità θ stimata da $\hat{\theta}$ è dato da $\kappa(\hat{\theta})$.

Lo stimatore di massima verosimiglianza di una distribuzione esponenziale è la media campionaria: $\hat{\theta}(X_1, \dots, X_n) = \bar{X}$.

Lo stimatore di massima verosimiglianza di una distribuzione di Poisson è la media campionaria.

5.2. Stimatori UMVUE.

Definition 5.9. Uno stimatore T^*che gode delle proprietà

(1) T^* è non distorto per $\kappa(\theta)$
(2) $Var_\theta(T^*) \leq Var_\theta(T)$ per ogni θ e per ogni stimatore T non distorto e a varianza finita

è detto *stimatore non distorto a varianza uniformemente minima* (*Uniform Miminum Variance Unbiased Estimator)*, o *stimatore* UMVUE.

Remark 5.10. Proprietà degli stimatori UMVUE

Unicità: se lo stimatore UMVUE esiste, è unico.
Simmetria: Sia $T^* = g(X_1, \dots, X_n)$ UMVUE, allora

$$P_\theta\left(g(X_1, \dots, X_n) = g(X_{\pi(1)}, \dots, X_{\pi(n)})\right) = 1 \quad \forall\theta \in \Theta$$

per ogni permutazione π di $\{1, \dots, n\}$.
Nosense: Lo stimatore UMVUE potrebbe esistere ma essere insensato.

5.2.1. *Disuguaglianza di Fréchet-Cramer-Rao.* È possibile trovare un confine inferiore (lower bound) della varianza nella classe di tutti gli stimatori non distorti che sia funzione solo della caratteristica da stimare $\kappa(\theta)$ e del modello statistico mediante la verosimiglianza L_θ. È anche possibile costruire uno stimatore che abbia varianza coincidente con esso. Tale stimatore sarà lo stimatore UMVUE. Il lower bound e lo stimatore possono essere trovati tramite la disuguaglianza di Fréchet-Cramer-Rao,

$$Var_\theta(T) \geq \frac{(\kappa'(\theta))^2}{nI(\theta)} \quad \forall\theta \in \Theta$$

definita dall'omonimo teorema:

Theorem 5.11. *Sia $(X_1, \dots, X_n)$ un campione aleatorio dalla famiglia di densità $f(x, \theta)$ a parametro reale $\theta \in \Theta \subset \mathbb{R}$ (perchè dovremo derivarlo).*
Sia $\kappa(\theta)$ la caratteristica da stimare e $T = g(X_1, \dots, X_n)$ lo stimatore non distorto *per $\kappa(\theta)$ (a varianza finita).*
Supponiamo che valgano le seguenti ipotesi (dette "di regolarità"):

(1) Θ *intervallo aperto di* $\mathbb{R}$
(2) $S = \{x : f(x, \theta) > 0\}$ *non dipende da* θ *(NB: S è il supporto)*
(3) $\theta \mapsto f(x, \theta)$ *è differenziabile in* $\Theta, \forall x$
(4) $\mathbb{E}_\theta\left[\frac{\partial}{\partial\theta} \log f(X_1, \theta)\right] = 0$, $\forall\theta$
(5) *Deve essere:* $0 < I(\theta) < +\infty$, $\forall\theta$, *con* $I(\theta) = \mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta} \log f(X_1, \theta)\right)^2\right]$ *che è detta* ***informazione di fisher***

Note 5.12. Se la (4) è verificata, allora $I(\theta) = Var\left[\frac{\partial}{\partial\theta} \log f(X_1, \theta)\right]$, perché $Var[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$, ma per la (4) $\mathbb{E}[X] = 0$.

(6) κ *è differenziabile in* Θ *e* $\kappa'(\theta) = \mathbb{E}\left[T \cdot \frac{\partial}{\partial\theta} \log L(\theta; X_1, \dots, X_n)\right]$ $\forall\theta \in \Theta$, *dove L è la funzione di verosimiglianza.*

Allora:

$$Var(T) \geq \frac{(\kappa'(\theta))^2}{n \cdot I(\theta)}, \quad \forall\theta \in \Theta$$

Note 5.13. I modelli *Esponenziale*, *Gaussiano* e *di Poisson* soddisfano le ipotesi di Fréchet-Cramer-Rao.

Definition 5.14. Uno stimatore T^* di $\kappa(\theta)$ **non distorto** la cui **varianza raggiunge il confine inferiore** di Fréchet-Cramer-Rao è detto *efficiente* e $Var(T^*) = \frac{(\kappa'(\theta))^2}{nI(\theta)}$.

Nel caso in cui $\kappa(\theta) = \theta$, allora $Var(T^*) = \frac{1}{nI(\theta)}$.
Uno stimatore **efficiente è anche UMVUE**.
Condizione necessaria e sufficiente perchè uno stimatore sia efficiente è che

$$\frac{\partial}{\partial\theta} \log L(\theta, X_1, \ldots, X_n) = a(n,\theta)(T - \kappa(\theta))$$

cioè che la derivata in θ del logaritmo della funzione di verosimiglianza sia una funzione lineare di $T - \kappa(\theta)$, con $\kappa(\theta)$ quantità da stimare e T stima di $\kappa(\theta)$.

6. Media e varianza campionarie

Definition 6.1. Data una serie di variabili aleatorie $X_1, \ldots, X_n$ la media campionaria è definita come $\overline{X} = \frac{\sum_{i=1}^n X_i}{n}$.

$\overline{X}$ è uno stimatore (puntuale) non distorto del valore atteso μ, in quanto $\mathbb{E}(\overline{X}) = \mu$

Definition 6.2. La varianza campionaria è definita come $S^2 := \frac{1}{n-1}\sum_{j=1}^n (X_j - \overline{X})^2 = \frac{\sum_{j=1}^n (X_j^2) - n\bar{X}^2}{n-1}$.

S^2 è uno stimatore (puntuale) della varianza σ^2. $\mathbb{E}(S^2) = \sigma^2$ (perchè $\mathbb{E}\left(\sum_{j=1}^n (X_j - \overline{X})^2\right) = (n-1)\sigma^2$) quindi, poichè in media assume il valore corretto, viene definito *stimatore non distorto* della varianza.

6.1. Distribuzioni di media e varianza campionarie di popolazione gaussiana.

Proposition 6.3. *Sia* $X_1, \ldots, X_n$ un campione casuale gaussiano dalla f.d.r. $N(\mu, \sigma^2)$. Per ogni $\mu \in \mathbb{R}$ e per ogni $\sigma^2 > 0$

(1) $\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$
(2) le statistiche S^2 e $\overline{X}$ sono *indipendenti*.
(3) $(n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$
Questo sarà utile come **statistica test** per calcolare gli intervalli di confidenza **per la varianza**.
(4) La statistica $\frac{\overline{X}-\mu}{S/\sqrt{n}}$ (dove $S = \sqrt{S^2}$) ha densità t *di student* con $n-1$ gradi di libertà.
Questo è utile in quanto $\frac{\overline{X}-\mu}{S/\sqrt{n}}$ è la media campionaria normalizzata da usare come **statistica test per la media** quando la varianza è incognita, e quindi stimata da S. Conoscendone la distribuzione, possiamo sfruttare le tavole per lavorare con questa statistica.

6.2. t di Student.

Definition 6.4. Siano Z e Y due v.a. indipendenti. Sia $Z \sim \mathcal{N}(0,1)$ e $Y \sim \chi^2_k$.
Si dice che $\frac{Z}{\sqrt{\frac{Y}{k}}}$ è distribuita secondo una t di Student con k gradi di libertà, cioè t_k. Tale distribuzione ha densità:

$$f_k(t) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{k\pi}\,\Gamma\left(\frac{k}{2}\right)} \frac{1}{\left(1 + \frac{t^2}{k}\right)^{\frac{k+1}{2}}} \quad t \in \mathbb{R}$$

che è simile ad una gaussiana, ma con code più alte.
Quando $k \to +\infty$, la distribuzione t si avvicina sempre più ad una normale.

7. Intervalli di confidenza

Gli stimatori puntuali non sono particolarmente interessanti in quanto è nulla la probabilità che assumano il vero valore (incognito) della variabile da stimare. Ad esempio, nel caso della media campionaria (stimatore della media)

$$P_{\mu,\sigma^2}(\overline{X} = c) = 0, \quad \forall c \in \mathbb{R},\ \mu \in \mathbb{R},\ \sigma^2 > 0$$

Possiamo tuttavia calcolare, *a priori* e indipendentemente dalla realizzazione campionaria, con un certo grado di fiducia un intervallo all'interno del quale andrà con buona approssimazione a cadere il valore cercato.
Per trovare intervalli di confidenza bilateri di livello $\gamma 100\%$ si usa la seguente formula:

$$\mathbb{P}(a < T < b) = \gamma$$

dove T è la statistica test opportuna e a e b sono quantili di tale statistica test.
Tale formula dovrà essere risolta in funzione della quantità per la quale si cerca l'intervallo.

7.1. Per la media.

Per la media si usa come statistica test $\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$ se la varianza σ^2 è nota, oppure $\frac{\bar{x}-\mu_0}{s/\sqrt{n}} \sim t_{n-1}$, con s^2 varianza campionaria, se la varianza è incognita.

7.2. **Per la varianza.**

- μ incognita

  Per trovare un intervallo di confidenza si parte dalla quantità aleatoria $\frac{S^2(n-1)}{\sigma^2} \sim \chi^2_{n-1}$. Ciò che si vuole determinare sono a e b tali che $P_{\mu,\sigma^2}\left(a < \frac{S^2(n-1)}{\sigma^2} < b\right) = \gamma$.

  a e b sono quantili di una f.d.r. χ^2_{n-1}.

  Si presentano diversi casi:

  (1) $[a = 0]$
  $b = \chi^2_{n-1}(\gamma)$

  Definition 7.1. Sia $\gamma \in (0,1)$ e sia s^2 il valore assunto da S^2 in corrispondenza della realizzazione campionaria $x_1, \ldots, x_n$ di un campione casuale estratto da una popolazione $N(\mu, \sigma^2)$. Allora

  $$\left(\frac{s^2(n-1)}{\chi^2_{n-1}(\gamma)}, +\infty\right)$$

  (dove $\chi^2_{n-1}(\gamma)$ è il quantile di ordine γ della f.d.r χ^2_{n-1}) è un *intervallo di confidenza a una coda superiore di livello* $\gamma 100\%$ *per la varianza* σ^2, quando μ è incognita. Inoltre, la statistica $\frac{S^2(n-1)}{\chi^2_{n-1}(\gamma)}$ è detta *limite inferiore di confidenza per la varianza.*

  (2) $[b = +\infty]$
  $a = \chi^2_{n-1}(1-\gamma)$

  Definition 7.2. Sia $\gamma \in (0,1)$ e sia s^2 il valore assunto da S^2 in corrispondenza della realizzazione campionaria $x_1, \ldots, x_n$di un campione casuale estratto da una popolazione $N(\mu, \sigma^2)$. Allora

  $$\left(0, \frac{s^2(n-1)}{\chi^2_{n-1}(1-\gamma)}\right)$$

  è un *intervallo di confidenza a una coda inferiore di livello* $\gamma 100\%$ *per la varianza* σ^2 quando μ è incognita. Inoltre, la statistica $\frac{S^2(n-1)}{\chi^2_{n-1}(1-\gamma)}$ è detta *limite superiore di confidenza per la varianza.*

  (3) $[0 < a < b < +\infty]$
  La massa rimanente deve essere distribuita uniformemente a destra e a sinistra dell'intervallo, quindi: $a = \chi^2_{n-1}(\frac{1-\gamma}{2})$ e $b = \chi^2_{n-1}(\frac{1-\gamma}{2} + \gamma) = \chi^2_{n-1}(\frac{1+\gamma}{2})$

  Definition 7.3. Sia $\gamma \in (0,1)$ e sia s^2 il valore assunto da S^2 in corrispondenza della realizzazione campionaria $x_1, \ldots, x_n$ di un campione casuale estratto dalla f.d.r. $N(\mu, \sigma^2)$. Allora

  $$\left(\frac{s^2(n-1)}{\chi^2_{n-1}\left(\frac{1+\gamma}{2}\right)}, \frac{s^2(n-1)}{\chi^2_{n-1}\left(\frac{1-\gamma}{2}\right)}\right)$$

  è un *intervallo di confidenza bilatero per* σ^2 *di livello* $\gamma 100\%$, quando μ è incognita.

- μ nota

  Essendo μ nota, possiamo stimare σ^2 con la statistica $S_0^2 := \frac{\sum_{j=1}^n (X_j - \mu)^2}{n}$ (che è lo stimatore di massima verosimiglianza di μ).

  $\frac{S_0^2 n}{\sigma^2}$ ha densità χ^2_n, quindi si ottiengono i seguenti *intervalli di confidenza per* σ^2 *di livello* $\gamma 100\%$ *quando* μ *è nota*:

  - $\left(\frac{\sum_{j=1}^n (x_j-\mu)^2}{\chi^2_n(\gamma)}, +\infty\right)$ (intervallo di confidenza a una coda superiore)
  - $\left(0, \frac{\sum_{j=1}^n (x_j-\mu)^2}{\chi^2_n(1-\gamma)}\right)$ (intervallo di confidenza a una coda inferiore)
  - $\left(\frac{\sum_{j=1}^n (x_j-\mu)^2}{\chi^2_n\left(\frac{1+\gamma}{2}\right)}, \frac{\sum_{j=1}^n (x_j-\mu)^2}{\chi^2_n\left(\frac{1-\gamma}{2}\right)}\right)$ (intervallo di confidenza bilatero)

8. Intervalli di confidenza per grandi campioni

8.1. **Per la media** μ**.** Sia $X_1, \ldots, X_n$ un campione con n grande da una popolazione con media μ e varianza σ^2.
Essendo n grande, il campione può essere trattato come una normale $N(\mu, \sigma^2)$.
La statistica $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$, dove $\bar{X}$ è la media campionaria, è distribuita come una normale standard $N(0,1)$.
È quindi possibile definire un intervallo di confidenza per la media μ di dimensione γ calcolando

$$P\left(-z_{\frac{1+\gamma}{2}} \frac{\sigma}{\sqrt{n}} < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < z_{\frac{1+\gamma}{2}}\right) \simeq \gamma$$

L'intervallo di confidenza è quindi

$$IC = \left(\bar{X} - z_{\frac{1+\gamma}{2}} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\frac{1+\gamma}{2}} \frac{\sigma}{\sqrt{n}}\right)$$

8.2. **Per una generica caratteristica $\kappa(\theta)$.** Supponiamo di dover stimare una caratteristica $\kappa(\theta)$ di cui abbiamo lo stimatore di massima verosimiglianza $\hat{\kappa} \sim \mathcal{N}\left(\kappa(\theta), \frac{(\kappa'(\theta))^2}{nI(\theta)}\right)$, con media pari alla caratteristica da stimare e varianza (calcolabile come $Var(\hat{\kappa})$) che raggiunge il limite inferiore di Fréchet-Cramer-Rao. Per n grande, $\frac{\hat{\kappa}-\kappa(\theta)}{\sqrt{\frac{(\kappa'(\theta))^2}{nI(\theta)}}} \sim N(0,1)$, dove $I(\theta) = \mathbb{E}\left[(\frac{\partial}{\partial\theta}\log f(X_1,\theta))^2\right]$ è l'informazione di Fisher.

Un intervallo di confidenza di ampiezza γ si definisce quindi a partire da

$$P_\theta\left(-z_{\frac{1+\gamma}{2}} < \frac{\hat{\kappa}-\kappa(\theta)}{\sqrt{\frac{(\kappa'(\theta))^2}{nI(\theta)}}} < z_{\frac{1+\gamma}{2}}\right) = \gamma$$

ed è

$$\left(\hat{\kappa} - z_{\frac{1+\gamma}{2}}\sqrt{\frac{(\kappa'(\hat{\theta}))^2}{nI(\hat{\theta})}},\ \hat{\kappa} + z_{\frac{1+\gamma}{2}}\sqrt{\frac{(\kappa'(\hat{\theta}))^2}{nI(\hat{\theta})}}\right)$$

dove tutti i θ presenti al denominatore del pivot della equazione precedente possono essere approssimati con l'MLE $\hat{\theta}$ perché è dimostrabile che questa sostituzione mantiene l'asintoticità a $\mathcal{N}(0,1)$.

9. Test di ipotesi

Definition 9.1. Una ipotesi H è una affermazione sulla distribuzione F della popolazione. Un'ipotesi si definisce

semplice: se l'ipotesi specifica completamente (determina) un'unica distribuzione
composta: altrimenti

Ciò che ci interessa è una procedura statistica (test) che stabilisca se i dati campionari sono compatibili con l'ipotesi H. In tal caso si dice che *accetto* H.

Se i dati non sono compatibili con H, allora *rifiuto* H.

Definition 9.2. Una *verifica di ipotesi* è una terna ordinata

$$(\underbrace{X_1,\dots,X_n}_{\text{campione}};\underbrace{H_0,H_1}_{\text{ipotesi}};\underbrace{G}_{\text{regione critica}})$$

con $G \subseteq \mathbb{R}^n$.

Se $(x_1,\dots,x_n) \in G \Rightarrow$ rifiuto l'ipotesi H_0 e accetto H_1
Se $(x_1,\dots,x_n) \in G^c \Rightarrow$ non rifiuto H_0 e rifiuto H_1.

		VERO	
		H_0	H_1
ACCETTO	H_0	OK	Errore di II specie
	H_1	Errore di I specie	OK

Definition 9.3. (Taglia del test) $\alpha := \sup P_\theta(\underline{x} \in G)$ con $\theta \in \Theta_0$

α è anche detto *livello di significatività* del test ed è la probabilità di commettere un errore di I specie, cioè $\alpha = P_{H_0}(\text{Accetto } H_1) = P_{H_0}(\text{Rifiuto } H_0)$.

Definition 9.4. Il più piccolo valore di α per cui, in presenza di $\underline{x}$, rifiuto H_0 è detto *p-value*. Per calcolare il p-value, si calcola la probabilità $P_{H_0}(\text{Rifiuto } H_0)$ sotto l'ipotesi che la regione di rifiuto cominci nel punto indicato dall'attuale realizzazione campionaria della statistica test.

Quindi:
Se p-value $\leq \alpha$ rifiuto H_0 di livello di significatività α.
Se p-value $\geq \alpha$ non rifiuto H_0 di livello di significatività α.

Analogamente ad α, è possibile definire una funzione β che rappresenta la probabilità di commettere un errore di II specie:

Definition 9.5. Sia $\beta := \sup P_\theta(\underline{x} \in G^c)$ con $\theta \in \Theta_1$.
Cioè, $\beta = P_{H_1}(\text{Rifiuto } H_1)$
Allora $\pi = 1-\beta(\theta)$ con $\theta \in \Theta_1$ è la funzione di **potenza** del test.

Calcolare la potenza di un test sotto l'ipotesi H_1, equivale a calcolare la probabilità dell'appartenenza di T alla regione critica, con θ determinato dall'ipotesi scelta, riconducendo la scrittura della regione critica a quella di una distribuzione nota, se ciò è necessario per il calcolo di $\mathbb{P}$: $\pi(\theta \in \Theta_{H_1}) = 1-\beta(\theta) = 1-P_{H_1}(\text{Rifiuto } H_1) = 1-P_{H_1}(\text{Accetto } H_0) = P_{H_1}(\text{Rifiuto } H_0)$.

Data una dimensione prefissata di α, è possibile costruire una regione critica tale da massimizzare la potenza del test. Ciò può essere fatto tramite il Lemma di Neyman-Pearson.

Definition 9.6. (Lemma di Neyman-Pearson) Dato un campione $(X_1, \dots, X_n)$ da $f(x, \theta)$ con $\theta \in \Theta = \{\theta_0, \theta_1\}$, $H_0 : \theta = \theta_0$, $H_1 : \theta = \theta_1$; $L_0(\underline{x}) = L(\theta_0; x_1, \dots, x_n)$, $L_1(\underline{x}) = L(\theta_1; x_1, \dots, x_n)$.

Sia $G = G(\delta) = \left\{(x_1, \dots, x_n) \in \mathbb{R}^n : \frac{L_0(\underline{x})}{L_1(\underline{x})} \leq \delta\right\}$ la regione critica e sia α la sua taglia.

Allora, tra tutte le regioni critiche per verificare H_0 contro H_1 di taglia α, H è quella con potenza massima.

Una volta impostata la regione critica, per definirla completamente bisogna calcolare δ in modo tale che effettivamente $\mathbb{P}_{H_0}\left(\frac{L_0(\underline{x})}{L_1(\underline{x})} \leq \delta\right) = \alpha$.

NB: nel fare ciò, tutto ciò che è costante può essere incorporato direttamente dentro a δ, rendendo così più semplice la definizione della regione critica.

Remark 9.7. È importante scegliere correttamente cosa va in H_0 e cosa in H_1.

In H_0: ciò che ci viene chiesto di verificare. "Verificare l'ipotesi che..."

In H_1 : ciò che vogliamo dimostrare. "C'è evidenza sperimentale che...?", "Possiamo concludere che...?".

10. Test per campioni gaussiani accoppiati indipendenti

Siano $X_1, \dots, X_m$ *i.i.d.* $\sim N(\mu_X, \sigma_X^2)$ e $Y_1, \dots, Y_n$ *i.i.d.* $\sim N(\mu_X, \sigma_X^2)$.

10.1. Test F.

Applicabile quando σ_X^2, σ_Y^2 sono incognite e m, n sono grandi.

Mira a verificare l'ipotesi $H_0 : \sigma_X^2 = \sigma_Y^2$ contro l'ipotesi $H_1 : \sigma_X^2 \neq \sigma_Y^2$.

Le varianze, in quanto incognite, devono essere approssimate a partire dai campioni nel seguente modo:

$$S_X^2 = \frac{\sum_{j=1}^m x_j^2 - m * \bar{x}^2}{m-1}$$

$$S_Y^2 = \frac{\sum_{j=1}^n y_j^2 - n * \bar{y}^2}{n-1}$$

dove $\bar{x}, \bar{y}$ sono le medie dei due campioni.

L'ipotesi H_0 è rifiutata quando $T = \frac{S_X^2}{S_Y^2} \sim F_{m-1,n-1}$ cade nella regione di rifiuto:

$$G = \left\{T \leq F_{m-1,n-1}\left(\frac{\alpha}{2}\right) \text{ oppure } T \geq F_{m-1,n-1}\left(1 - \frac{\alpha}{2}\right)\right\}$$

cioè nell'itervallo di confidenza di ampiezza α:

$$\left(F_{m-1,n-1}\left(\frac{\alpha}{2}\right), F_{m-1,n-1}\left(1 - \frac{\alpha}{2}\right)\right)$$

dove $F_{m,n}$ è la funzione F di Fisher come presente nelle tabelle (attenzione all'ordine dei pedici, alcune tabelle li riportano al contrario).

Parte 2. Inferenza non parametrica

Definition 10.1. La funzione di ripartizione empirica associata al campione $\hat{F}_n$ è una funzione su $\mathbb{R}$ a valori in $[0, 1]$ definita da

$$\hat{F}_n(x) = \frac{\#\{j : X_j \leq x\}}{n} \quad \forall x \in \mathbb{R}$$

11. Media e varianza campionarie

Definition 11.1. Il momento r-esimo *campionario* di $\hat{F}_n$ è $M_r = \frac{1}{n}\sum_{j=1}^n X_j^r$

La media di $\hat{F}_n$ è uguale alla media campionaria , e, come nel caso parametrico, si ha:

Definition 11.2. $\mathbb{E}[\hat{F}_n] = M_1 = \frac{1}{n}\sum_{j=1}^n X_j$

Quindi $\overline{X} = M_1$.

Nel caso si abbia la distribuzione campionaria è comodo calcolare la media come "media pesata":

$$\mathbb{E}[\hat{F}_n] = \sum_{j=1}^n x_j d_j$$

dove x_j è il valore del campione e d_j la sua densità, eventualmente ricavabile dalla funzione di ripartizione empirica $\hat{F}_n$ come $d_j = F_n(x_j) - F(x_j - 1)$.

Al contrario della media, la varianza di $\hat{F}_n$ è diversa dalla varianza campionaria, infatti si ha:

Definition 11.3. (Varianza) $Var[\hat{F}_n] = \frac{1}{n}\sum_{j=1}^n (X_j - \overline{X})^2$

e

Definition 11.4. (Varianza campionaria) $S^2 = \frac{1}{n-1}\sum_{j=1}^n (X_j - \overline{X})^2 = Var[\hat{F}_n]\frac{n}{n-1}$

La varianza è calcolabile anche tramite le normali proprietà di media e varianza, quindi $Var(\hat{F}_n) = \mathbb{E}[\hat{F}_n^2] - \mathbb{E}^2[\hat{F}_n]$.
In tal caso, si può calcolare $\mathbb{E}[\hat{F}_n^2]$ come il momento secondo, cioè $\mathbb{E}[\hat{F}_n^2] = \sum_{j=1}^{n} x^2 d_j$, con d_j definito come sopra.

12. Test di Kolmogorof-Smirnov

Serve a **verificare se una funzione F è distribuita secondo una determinata funzione F_0 completamente specificata** (cioè con tutti i parametri fissati: ad esempio $\mathcal{N}(0,1)$, $Exp(5)$, ecc) e **continua**.
Vale per un numero n qualsiasi di campioni.
In particolare si controlla l'**ipotesi** nulla

$$H_0 : F = F_0$$

contro

$$H_1 : F \neq F_0$$

Statistica test: $D_n := \sup_{x \in \{n \text{ campioni}\}} |\hat{F}(x) - F_0(x)|$
Rifiuto H_0 se $D_n > q_{D_n}(1-\alpha)$ dove α è il livello di significatività e q_{D_n} è il quantile della statistica test di Kolmogorof-Smirnov, cioè rifiuto se la funzione di ripartizione empirica si discosta più di un massimo sopportabile dalla FDR teorica F_0 (per quanto osservabile con i campioni a nostra disposizione).
NB: essendo $\hat{F}_n(x)$ discontinua, la differenza da $F_0(x)$ va calcolata sia nell'intorno sinistro, sia nell'intorno destro di ogni campione, dove essa assumerà valori diversi. Bisognerà quindi calcolare sia $|\hat{F}_n(x_i) - F_0(x_i)|$ sia $|\hat{F}_n(x_{i-1}) - F_0(x_i)|$.

13. Test χ^2 di adattamento

Serve a verificare se una serie di dati si adatta ad un determinato modello teorico.
Può essere usato solo per **campioni di dimensione n grande**, perchè è un test **asintotico**. Bisogna infatti verificarne le seguenti **regole empiriche** di applicabilità: $n \geq 50$ e $n \cdot p_{0_i} > 5, \ \forall i \in 1, \ldots, k$
Funziona su **dati discreti**, oppure su **dati continui** purchè essi vengano **discretizzati** tramite suddivisione in classi (il test non distingue come la massa si distribuisce all'interno delle classi, ma solo tra le diverse classi).
Si imposta sulle seguenti ipotesi:
$H_0 : F = F_0$ contro $H_1 : F \neq F_0$
che possono essere riscritte come:
$H_0 : p_i = p_{0i} \quad \forall i = 1, \ldots, k$ contro $H_1 : p_i \neq p_{0i}$ per qualche i.
dove $p_i = P_F(X_i = a_i)$ e $p_{0i} = P_{F_0}(X_i = a_i)$, ($a_i$ sono le diverse classi), cioè p_i è la densità osservata in corrispondenza di una certa classe, e p_{0_i} è la densità teorica che tale classe dovrebbe avere.
Per eseguire il test, bisogna calcolare la *frequenza assoluta campionaria* di ogni a_i, cioè il numero di osservazioni del campione che assumono valore a_i:

$$N_i = \#\{j : X_j = a_i\} \quad \forall i = 1, \ldots, k$$

e misurare lo scostamento fra tali osservazioni e i valori teorici che esse dovrebbero avere ($n \cdot p_{0_i}$).
Tale misura viene effettuata mediante la statistica di Pearson:

$$Q_n := \sum_{i=1}^{k} \frac{(N_i - n\,p_{0i})^2}{n\,p_{0i}} = \sum_{i=1}^{k} \frac{N_i^2}{np_{0i}} - n$$

Rifiutiamo H_0 a livello α sse $Q_n > \chi^2_{k-1}(1-\alpha)$, cioè se lo scostamento è troppo grande.
Il p-value di questo test è calcolabile come: $p = 1 - F_{\chi^2_{k-1}}(q_n)$ dove q_n è la realizzazione di Q_n.
NB: nel caso la distribuzione teorica non sia completamente specificata e che si debbano stimare a partire dal campione m parametri, la regione di rifiuto sarà: $Q_n > \chi^2_{k-m-1}(1-\alpha)$. Analogamente, varierà il p-value.

Remark 13.1. Se bisogna decidere il numero di classi in cui suddividere $\mathbb{R}$ per l'uso con questo test, il valore ideale è $k = \lfloor n^{2/5} \rfloor$.
Si sceglieranno poi gli estremi di tali classi in modo che ognuna di esse sia equiprobabile sotto F_0: $\mathbb{P}(X_i \in A_i) = \frac{1}{k} \ \forall i$

14. Test di indipendenza

14.1. Test χ^2 di indipendenza.

Serve a verificare se due serie di campioni X e Y sono tra loro indipendenti.
Il test χ^2 di indipendenza può essere impostato a partire dalle seguenti ipotesi:
H_0: $H(x,y) = F(x) \cdot G(y) \quad \forall x \in \mathbb{R} \quad \forall y \in \mathbb{R}$ cioè X e Y sono indipendenti.
H_1: $H(x,y) \neq F(x) \cdot G(y)$ per almeno un $(x,y) \in \mathbb{R}^2$.
Il test lavora discretizzando F in r classi e G in c classi, e contando il numero di coppie tali che il primo elemento è nella classe A_i e il secondo in B_j, ossia $N_{ij} = \#(X_k, Y_k) \in A_i \cdot B_j$, cioè le densità congiunte.
Si calcola poi la probabilità teorica che una coppia ha di appartenere a ogni classe: $p_{i,j} = P_H(X_1 \in A_i, Y_1 \in B_j), \ i = 1, \ldots, r \ \ j = 1, \ldots, c$
Possiamo anche calcolare, al suo posto, direttamente il numero teorico di elementi di ogni classe, pari a $E_{i,j} = \frac{f_X(x) \cdot f_Y(y)}{n}$, dove f_x e f_Y sono le distribuzioni marginali.

Il test è **asintotico** ed è applicabile solo se valgono le seguenti regole empiriche: $n \geq 50$, $\frac{n}{r} > 5$, $\frac{n}{c} > 5$.

Usiamo come **statistica test**: $U := \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(N_{ij}-E_{ij})^2}{E_{ij}} = \left(\sum_i\sum_j\frac{(N_{ij})^2}{E_{ij}}\right) - n$ cioè lo scostamento dei valori registrati dai valori teorici.

Rifiuto H_0 a livello α se U_n è grande, cioè se $U_n \geq \chi^2_{(r-1)(c-1)}(1-\alpha)$.

14.2. **Test di indipendenza per dati gaussiani.** Se (X,Y) è congiuntamente gaussiana, X,Y sono indipendenti se e solo se il coefficiente di correlazione lineare ρ è nullo.

$$\rho = \frac{Cov(X,Y)}{\sqrt{Var(X)\cdot Var(Y)}}$$

Un valore positivo di ρ indica concordanza di tipo lineare tra i due campioni, mentre un valore negativo indica discordanza (quando un campione cresce, l'altro tende a diminuire).

H_0:$\rho = 0$ (indipendenza) contro H_1:$\rho \neq 0$.

Nel caso di campione accoppiato gaussiano con parametri tutti incogniti, uno stimatore di ρ è il *coefficiente di correlazione campionario* (o empirico)

$$R = \frac{\sum_{j=1}^{n}(X_j-\bar{X})(Y_j-\bar{Y})}{\sqrt{\sum_{j=1}^{n}(X_j-\bar{X})^2\sum_{j=1}^{n}(Y_j-\bar{Y})^2}}$$

per la quale vale sempre $-1 \leq R \leq 1$ e il seguente

Theorem 14.1. *Sia* $(X_1,Y_1),\ldots,(X_n,Y_n)$ *i.i.d.* $\sim N$ *e* $\rho = 0$. *Allora*

$$T := \frac{\sqrt{n-2}\,R}{\sqrt{1-R^2}} \sim t_{n-2} \quad n \geq 3$$

Tale grandezza è la statistica test che utilizziamo per verificare l'indipendenza dei campioni.

Quindi: si rifiuta H_0 nel caso in cui $T \geq t_{n-2}(1-\frac{\alpha}{2})$. Per n grandi, t è approssimabile da una gaussiana standard.

È anche possibile impostare i seguenti test, con le relative regioni di rifiuto:

$H_0 : \rho \leq 0$ contro $H_1 : \rho > 0$ con $G = \{\text{campioni} : T \geq t_{n-2}(1-a)\}$

e

$H_0 : \rho \geq 0$ contro $H_1 : \rho < 0$ con $G = \{\text{campioni} : T \leq -t_{n-2}(1-a)\}$

15. Test di omogeneità di Wilcoxon-Mann-Whitney

Un test di omogeneità serve a **verificare se due campioni aleatori sono regolati dallo stesso modello**, cioè se hanno la stessa funzione di ripartizione.

Tramite un'opportuna ipotesi alternativa H_1, può essere utilizzato anche per **determinare se una variabile *domina stocasticamente* l'altra** (cioè se "è più grande").

Siano X e Y le due variabili aleatorie dei cui campioni si vuole verificare l'omogeneità e F e G le loro funzioni di ripartizione e $X_1,\ldots,X_m$ *i.i.d.* $\sim F$ e $Y_1,\ldots,Y_n$ *i.i.d.* $\sim G$ i due campioni di dati raccolti.

L'ipotesi nulla indica omogeneità ed è:

$$H_0 : F(x) = G(x) \quad \forall x \in \mathbb{R}$$

L'alternativa può indicare non omogeneità:

$$H_1 : F(x) \neq G(x) \quad \text{per qualche } x \in \mathbb{R}$$

oppure può indicare che X domina stocasticamente Y:

$$H_1 : F(x) \leq G(x) \ \ \forall x \in \mathbb{R} \text{ e } F(x) < G(x) \text{ per qualche } x$$

oppure può indicare che Y domina stocasticamente X:

$$H_1 : F(x) \geq G(x) \quad \forall x \in \mathbb{R} \text{ e } F(x) > G(x) \text{ per qualche } x$$

Per eseguire il test si riuniscono tutte le osservazioni di X e Y in un unico campione di lunghezza $m+n$, le si dispongono in ordine crescente e si assegna loro un rango r crescente dalla minore ($r=1$) alla maggiore ($r=m+n$). Si assume per semplicità che non ci siano ripetizioni nel campione.

Chiamiamo T_X la somma dei ranghi delle osservazioni presenti da X: $T_X = \sum_{i=1}^{m} R_i$ con $R_i = rango(X_i)$

Chiamiamo w_a il quantile della f.d.r. di T_X (tabulato per $m,n \leq 20$).

Se $X \overset{st}{\leq} Y$ mi aspetto che tante x_i siano più piccole delle y_j, quindi T_X assumerà valori piccoli.

Se $X \overset{st}{\geq} Y$ mi aspetto che tante x_i siano più grandi delle y_j, quindi T_X assumerà valori grandi.

Valgono le seguenti regole di significatività per α:

Rifiuto $H_0 : F(x) = G(x)\ \forall x$ e accetto $H_1 : F(x) \geq G(x)\ \forall x$ e $F(x) > G(x)$ per qualche x (ossia $X \overset{st}{\leq} Y$) se $T_x < w_\alpha$.

Rifiuto $H_0 : F(x) = G(x)\ \forall x$ e accetto $H_1 : F(x) \leq G(x)\ \forall x$ e $F(x) < G(x)$ per qualche x (ossia $X \overset{st}{\geq} Y$)se $T_x > w_{1-\alpha}$.

Rifiuto $H_0 : F(x) = G(x)\ \forall x$ e accetto $H_1 : F(x) \neq G(x)$ se $T_x < w_{\alpha/2}$ oppure $T_X > w_{1-\alpha/2}$.

NB: la statistica T_X è distribuita (più o meno come una gaussiana) attorno alla propria media c. Quindi $w_{m,n}(1-\alpha) = c + (c - w_{m,n}(\alpha)) = 2c - w_{m,n}(\alpha) = m(m+n+1) - w_{m,n}(\alpha)$