

- Richiami di calcolo delle probabilità, I parte

Variabili aleatorie discrete: distribuzione di probabilità, funzione di ripartizione e di densità. Il modello bernoulliano e il modello binomiale. Valore atteso di v.a. e di somme di v.a. Varianza di v.a. e di somme di v.a. indipendenti.

- Richiami di calcolo delle probabilità, II parte

Approssimazione gaussiana e teorema limite centrale.

Teorema 3.6.1 (di De Moivre-Laplace) Sia S_n il numero di successi in n prove di Bernoulli, in ognuna delle quali il successo ha probabilità $p \in (0, 1)$. Allora, per ogni $a < b$,

$$P\left(a < \frac{S_n - np}{\sqrt{np(1-p)}} \leq b\right) \rightarrow \Phi(b) - \Phi(a), \quad \text{per } n \rightarrow +\infty,$$

dove Φ è la funzione di ripartizione di una gaussiana standard.

Teorema 4.11.6 Sia X_1, X_2, \dots una successione di variabili aleatorie i.i.d. con media μ e varianza σ^2 , con $0 < \sigma^2 < +\infty$. Allora per ogni $x \in \mathbb{R}$:

$$\lim_{n \rightarrow +\infty} P\left(\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \leq x\right) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du = \Phi(x). \quad (4.11.1)$$

V.a. assolutamente continue: densità e funzione di ripartizione.

V.a. gaussiane non standard, momenti e standardizzazione, uso delle tavole della gaussiana standard.

Definizione 3.7.1 Sia X una variabile aleatoria assolutamente continua o discreta tale che $|X|^k$ ammetta valore atteso. Allora il numero $E(X^k)$ è detto momento k -esimo o momento di ordine k della variabile aleatoria X .

Proposizione 3.7.2 Sia X una variabile aleatoria che ammette momento k -esimo, per qualche $k \geq 2$. Allora X ammette momento h -esimo per ogni $1 \leq h < k$.

Definizione 3.7.5 Sia X una variabile aleatoria per la quale esiste un intervallo aperto \mathcal{O} contenente lo 0 tale che e^{tX} ammette media per ogni t in \mathcal{O} . Allora la funzione

$$m_X(t) := E(e^{tX})$$

definita (almeno) per ogni $t \in \mathcal{O}$ è detta funzione generatrice dei momenti di X .

Proposizione 3.7.8 Sia X una variabile aleatoria che ammette funzione generatrice dei momenti m_X . Allora esistono tutti i momenti di X e

$$E(X) = m'_X(0), \quad E(X^2) = m''_X(0), \dots$$

- Richiami di calcolo delle probabilità, III parte

Quantili di v.a. discrete e assolutamente continue.

Quantili della gaussiana non standard con le tavole.

Funzione generatrice dei momenti. Metodi di calcolo della varianza. Trasformate di v.a. discrete e assolutamente continue con il metodo della funzione di ripartizione.

Distribuzione del minimo di 3 v.a. indipendenti. Esempi con i modelli di Poisson, normale, esponenziale. La v.a. $\chi^2(1)$. Il modello gamma come famiglia contenente i modelli esponenziale e $\chi^2(1)$.

Enunciati sulle distribuzioni di somme di v.a. indipendenti notevoli (somme di bernoulliane, di χ^2 , di esponenziali, di gamma, di gaussiane, di Poisson).

- Introduzione alla stima puntuale

Definizione 2.1 Sia x_1, \dots, x_n una realizzazione del campione casuale X_1, \dots, X_n estratto da $f(x, \theta)$ con $\theta \in \Theta$ e sia $\gamma \in (0, 1)$. Il sottoinsieme di Θ dato da $S(x_1, \dots, x_n)$ è detto *regione (o insieme) di confidenza per θ di livello $\gamma 100\%$* se

$$P_\theta(S(X_1, \dots, X_n) \ni \theta) \geq \gamma$$

cioè se l'insieme aleatorio $S(X_1, \dots, X_n)$ contiene il vero valore del parametro θ con probabilità almeno pari a γ .

Definizioni di famiglia parametrica, campione casuale,

Definizione 1.1 Siano X_1, \dots, X_n n variabili aleatorie indipendenti e identicamente distribuite con comune funzione di densità $f(x, \theta)$, $\theta \in \Theta$. Diremo che X_1, \dots, X_n è un *campione casuale* di dimensione n estratto dalla popolazione di densità $f(x, \theta)$.

statistica,

Definizione 1.2 Una *statistica* è una variabile aleatoria T funzione del campione: $T = g(X_1, \dots, X_n)$. La distribuzione (o legge) di una statistica T è detta *distribuzione (o legge) campionaria*.

caratteristica di una popolazione

Definizione 1.3 Una funzione $\kappa : \Theta \rightarrow \mathbb{R}$ è detta *caratteristica della popolazione*. Se κ è costante su Θ , κ è una *caratteristica banale (trivial characteristic)*.

stimatore

Definizione 1.4 Siano X_1, \dots, 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, \dots, X_n , è una statistica $T = g(X_1, \dots, X_n)$ usata per stimare $\kappa(\theta)$. Il valore assunto da uno stimatore T di $\kappa(\theta)$ è detto *stima* di $\kappa(\theta)$.

mean squared error

Definizione 1.10 Se T è uno stimatore di $\kappa(\theta)$ tale che $E_\theta[(T - \kappa(\theta))^2] < \infty$ per ogni $\theta \in \Theta$, allora $E_\theta[(T - \kappa(\theta))^2]$ è detto *errore quadratico medio* di T rispetto a $\kappa(\theta)$ ed è indicato con l'acronimo *MSE* che sta per Mean Squared Error (o Mean Square Error).

Il MSE di uno stimatore T esiste se e solo se T ha (media e) varianza finite, o, equivalentemente, se e solo se ha momento secondo finito. Infatti, usando la disuguaglianza

Osservazione 1.11 Per calcolare l'errore quadratico medio è utile decomporlo nel seguente modo:

$$E_\theta[(T - \kappa(\theta))^2] = E_\theta[((T - E_\theta(T)) + (E_\theta(T) - \kappa(\theta)))^2] = E_\theta[(T - E_\theta(T))^2 + (E_\theta(T) - \kappa(\theta))^2]$$

cioè

$$(3) \quad E_\theta[(T - \kappa(\theta))^2] = \text{Var}_\theta(T) + [E_\theta(T) - \kappa(\theta)]^2$$

La quantità $(E_\theta(T) - \kappa(\theta))$ nella decomposizione (3) è detta *distorsione (bias)* di T .

Sulla base dell'errore quadratico medio, *preferiremo* T_1 a T_2 se

- $E_\theta[(T_1 - \kappa(\theta))^2] \leq E_\theta[(T_2 - \kappa(\theta))^2] \quad \forall \theta \in \Theta$ e
- $E_\theta[(T_1 - \kappa(\theta))^2] < E_\theta[(T_2 - \kappa(\theta))^2]$ per qualche $\theta \in \Theta$

il problema dello stimatore ottimo.

In linea teorica, fra tutti gli stimatori di $\kappa(\theta)$ a varianza finita, ci piacerebbe scegliere quello che minimizza il MSE qualunque sia θ , cioè ci piacerebbe scegliere uno stimatore T_0 tale che

$$E_\theta[(T_0 - \kappa(\theta))^2] \leq E_\theta[(T - \kappa(\theta))^2] \quad \forall \theta \in \Theta \quad \forall T$$

Ma un tale T_0 non esiste: supponete infatti che $\kappa(\theta) = \theta \in \mathbb{R}$. Forse è insensato, comunque potreste decidere di stimare $\kappa(\theta)$ con la statistica costante $\tilde{T} = 5$, il cui MSE in $\theta = 5$ vale $E_5[(\tilde{T} - 5)^2] = E_5[(5 - 5)^2] = 0$; cioè sebbene la scelta di $\tilde{\theta} = 5$ sia assurda, per $\theta = 5$ la costante 5 si comporta meglio di qualunque altro stimatore.

La ragione di ciò è che la classe di tutti gli stimatori con MSE finito è troppo grande; per esempio contiene anche stimatori banali (per esempio $\tilde{T} = 5$) “buoni” solo in un singolo punto. Allora, per uscire dall’impaccio è sufficiente restringerci nella ricerca a una sottoclasse di stimatori. La sottoclasse ci è suggerita dalla decomposizione (3) del MSE, nella quale leggiamo che minimizzare MSE equivale a minimizzare contemporaneamente varianza e distorsione di T . Allora, aggiustiamo il tiro e restringiamo la nostra ricerca degli stimatori ottimali alla classe di quelli che hanno distorsione nulla, detti *non distorti* o anche *corretti*. L’errore quadratico medio di uno stimatore non distorto di una caratteristica $\kappa(\theta)$ coincide con la sua varianza e il programma di ricerca di uno stimatore ottimale fra quelli non distorti diventa la ricerca di uno stimatore che abbia varianza più piccola per ogni $\theta \in \Theta$.

Definizione 2.1 Una statistica T che ammette media per ogni θ in Θ è detta *stimatore non distorto* o corretto (*unbiased*) della caratteristica $\kappa(\theta)$ se

$$(4) \quad E_\theta(T) = \kappa(\theta) \quad \forall \theta \in \Theta$$

stimatori a varianza minima tra i non distorti: UMVUE .

Riprendiamo ora il filo della ricerca di stimatori ottimali. Riassumendo: abbiamo un campione casuale X_1, \dots, X_n estratto da $\{f(\cdot, \theta), \theta \in \Theta\}$ e stiamo cercando “lo” stimatore T^* “ottimo” per $\kappa(\theta)$, cioè stiamo cercando “lo” stimatore T^* che soddisfa le seguenti proprietà:

1. T^* è non distorto per $\kappa(\theta)$;
2. $\text{Var}_\theta(T^*) \leq \text{Var}_\theta(T)$ per ogni θ e per ogni stimatore T non distorto e a varianza finita.

Definizione 3.1 Uno stimatore T^* che gode delle proprietà 1. e 2. è detto *stimatore non distorto a varianza uniformemente minima* (*Uniform Minimum Variance Unbiased Estimator*).

Proposizione 3.2 (Unicità dell’UMVUE) Se uno stimatore UMVUE per $\kappa(\theta)$ esiste, allora esso è essenzialmente unico, cioè se T_1, T_2 sono entrambi UMVUE, allora $P_\theta(T_1 = T_2) = 1 \quad \forall \theta \in \Theta$.

Dimostrazione Innanzitutto osserviamo che $E(T_1 - T_2) = \kappa(\theta) - \kappa(\theta) = 0$. Rimane da dimostrare che $\text{Var}(T_1 - T_2) = 0$. Infatti, per le proprietà della varianza abbiamo che $E(T_1 - T_2) = 0$ e $\text{Var}(T_1 - T_2) = 0$ implicano $P_\theta(T_1 - T_2 = 0) = 1$. Chiamiamo v il comune valore di $\text{Var}(T_1), \text{Var}(T_2)$. Allora

$$\text{Var}(T_1 - T_2) = \text{Var}(T_1) + \text{Var}(T_2) - 2\text{Cov}(T_1, T_2) = 2v \text{Cov}(T_1, T_2) \geq 0$$

se e solo se $\text{Cov}(T_1, T_2) \leq v$. D’altro canto, siccome $(T_1 + T_2)/2$ è stimatore non distorto di $\kappa(\theta)$, allora la sua varianza è maggiore di quella dei due stimatori UMVUE, cioè abbiamo

$$\text{Var}((T_1 + T_2)/2) = (1/4)(\text{Var}(T_1) + \text{Var}(T_2) + 2\text{Cov}(T_1, T_2)) = v/2 + \text{Cov}(T_1, T_2)/2 \geq v .$$

Segue che $\text{Cov}(T_1, T_2) \geq v$. Ma, $\text{Cov}(T_1, T_2) \leq v$ e $\text{Cov}(T_1, T_2) \geq v$ se e solo se $\text{Cov}(T_1, T_2) = v$, da cui deduciamo $\text{Var}(T_1 - T_2) = 0$. ■

Proposizione 3.2 (Unicità dell’UMVUE) Se uno stimatore UMVUE per $\kappa(\theta)$ esiste, allora esso è essenzialmente unico, cioè se T_1, T_2 sono entrambi UMVUE, allora $P_\theta(T_1 = T_2) = 1 \quad \forall \theta \in \Theta$.

Non distorsione dello stimatore della media campionaria.

Esempio 2.2 Se X_1, \dots, X_n i.i.d. $\sim f(x, \theta)$, $\theta \in \Theta$, ed $E_\theta(X_1)$ esiste qualunque sia θ , allora $E_\theta(\bar{X}) = E_\theta(X_1)$, $\forall \theta$ e quindi:

La media campionaria \bar{X} è stimatore non distorto della media "teorica" $E_\theta(X_1)$.

Non distorsione dello stimatore della varianza campionaria.

Se inoltre esiste anche $\text{Var}_\theta(X_1)$, $\forall \theta \in \Theta$, allora $E_\theta(S^2) = \text{Var}_\theta(X_1)$, $\forall \theta \in \Theta$ e quindi

la varianza campionaria S^2 è stimatore non distorto della varianza "teorica" $\text{Var}_\theta(X_1)$.

Metodo dei momenti. (Questi ultimi due argomenti svolti nelle ore di esercitazione, il 26/3).

Sia X_1, \dots, X_n un campione casuale estratto dalla densità $f(x, \theta)$ con $\theta = (\theta_1, \dots, \theta_m) \in \Theta \subset \mathbb{R}^m$. Supponiamo che per ogni θ X_1 ammette i primi m momenti definiti da $\mu_1(\theta) := E_\theta(X_1), \dots, \mu_m(\theta) := E_\theta(X_1^m)$. Per esempio nel caso continuo abbiamo che se $\int_{\mathbb{R}} |x|^r |f(x, \theta)| dx < \infty$ allora il momento r -esimo $\mu_r(\theta) = E_\theta(X_1^r)$ esiste ed è dato da $\mu_r(\theta) = \int_{\mathbb{R}} x^r f(x, \theta) dx$. Introduciamo poi le m statistiche M_1, \dots, M_m date dai primi m momenti campionari

$$M_r = \frac{1}{n} \sum_{j=1}^n X_j^r \quad r = 1, \dots, m$$

In particolare, il primo momento campionario è la media campionaria: $M_1 = \bar{X}$. Consideriamo poi il sistema di m equazioni

$$(18) \quad \begin{cases} \mu_1(\theta) = M_1 \\ \dots \\ \mu_m(\theta) = M_m \end{cases}$$

nelle m incognite $\theta_1, \dots, \theta_m$. Se il sistema (18) ammette soluzione $\hat{\theta}_1, \dots, \hat{\theta}_m$, allora $\hat{\theta}_1, \dots, \hat{\theta}_m$ sono statistiche, perché dipendono soltanto dai momenti campionari M_1, \dots, M_m . Possiamo quindi usare $\hat{\theta}_1, \dots, \hat{\theta}_m$ come stimatori rispettivamente di $\theta_1, \dots, \theta_m$.

$\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_m)$ è lo stimatore di $\theta = (\theta_1, \dots, \theta_m)$ ottenuto con il metodo dei momenti.

Ovviamente,

se $\mu_r(\theta)$ esiste allora M_r è stimatore non distorto della caratteristica $\kappa(\theta) = \mu_r(\theta)$.

Infatti:

$$E_\theta(M_r) = E_\theta \left(\frac{1}{n} \sum_{j=1}^n X_j^r \right) = \frac{1}{n} \sum_{j=1}^n E_\theta(X_j^r) = \mu_r(\theta).$$

- Stima efficiente

Abbiamo già introdotto l'errore quadratico medio come misura della bontà di uno stimatore (quanto minore è l'MSE, tanto migliore è lo stimatore) e abbiamo visto come nella classe degli stimatori non distorti il MSE si riduce alla varianza cosicché il problema della minimizzazione del MSE diventa il problema della minimizzazione della varianza. Ci chiediamo ora se esiste un confine inferiore (lower bound) della varianza nella classe di tutti gli stimatori non distorti, che sia funzione soltanto della caratteristica da stimare $\kappa(\theta)$ e del modello statistico mediante la verosimiglianza L_θ . Inoltre, se questo confine inferiore per la varianza esiste, ci chiediamo se sia possibile costruire uno stimatore che abbia varianza coincidente con esso. Se ciò è possibile, allora quello stimatore realizzerà il programma di ricerca dell'UMVUE.

La stima efficiente e il teorema di Frechét-Cramér-Rao (con dimostrazione), preceduto dal richiamo della definizione di covarianza e dalla definizione di funzione di verosimiglianza.

Definizione 5.1 La *funzione di verosimiglianza* (*Likelihood function*) di n variabili aleatorie X_1, \dots, X_n è data dalla funzione di densità congiunta di X_1, \dots, X_n , considerata come funzione di θ . Se X_1, \dots, X_n è un campione casuale estratto dalla densità $f(x, \theta)$, $\theta \in \Theta$, la funzione di verosimiglianza è

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

Nota 6.3 La dimostrazione del Teorema 6.1 si basa su alcune proprietà della covarianza che qui riprendiamo. La *covarianza* di due variabili aleatorie X, Y a varianza finita è data da $\text{Cov}(X, Y) = E[(X - E(X))(Y - E(Y))]$ e si può calcolare come $\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$. La covarianza gode delle seguenti proprietà:

- (j) $|\text{Cov}(X, Y)| \leq \sqrt{\text{Var}(X) \text{Var}(Y)}$;
- (jj) $|\text{Cov}(X, Y)| = \sqrt{\text{Var}(X) \text{Var}(Y)}$ se e solo se esistono $a, b \in \mathbb{R}$ tali che $P(Y = aX + b) = 1$. Inoltre, se $\text{Var}(X) > 0$ allora $a = \text{Cov}(X, Y) / \text{Var}(X)$ e $b = E(Y) - aE(X)$
- (jjj) se $E(Y) = 0$ allora $b = -aE(X)$ e $P(Y = a(X - E(X))) = 1$

Teorema 6.1 Siano X_1, \dots, X_n variabili aleatorie i.i.d. con comune densità $f(x, \theta)$, $\theta \in \Theta \subset \mathbb{R}$ e sia $T = g(X_1, \dots, X_n)$ uno stimatore non distorto della caratteristica $\kappa(\theta)$ a varianza finita. Assumiamo che le seguenti condizioni di regolarità siano soddisfatte:

- (i) Θ è un intervallo aperto di \mathbb{R} ;
- (ii) $S = \{x : f(x, \theta) > 0\}$ è indipendente da θ ;
- (iii) $\theta \mapsto f(x, \theta)$ è derivabile su Θ , $\forall x \in S$;
- (iv) $E_\theta \left(\frac{\partial}{\partial \theta} \log f(X_1, \theta) \right) = 0 \quad \forall \theta \in \Theta$;
- (v) $0 < E_\theta \left[\left(\frac{\partial}{\partial \theta} \log f(X_1, \theta) \right)^2 \right] < \infty \quad \forall \theta \in \Theta$;
- (vi) $\kappa : \Theta \rightarrow \mathbb{R}$ è derivabile su Θ e

$$\kappa'(\theta) = E_\theta \left(T \frac{\partial}{\partial \theta} \log L_\theta(X_1, \dots, X_n) \right) \quad \forall \theta \in \Theta$$

Allora

$$(10) \quad \text{Var}_\theta(T) \geq \frac{(\kappa'(\theta))^2}{nI(\theta)} \quad \forall \theta \in \Theta$$

dove

$$I(\theta) = E_\theta \left[\left(\frac{\partial}{\partial \theta} \log f(X_1, \theta) \right)^2 \right]$$

Inoltre, l'uguaglianza in (10) vale se e solo se esiste una funzione $a(n, \theta)$ tale che

$$(11) \quad P_\theta \left(\frac{\partial}{\partial \theta} \log L_\theta(X_1, \dots, X_n) = a(n, \theta)(T - \kappa(\theta)) \right) = 1 \quad \forall \theta \in \Theta$$

Dimostrazione del Teorema 6.1 Per maggiore semplicità notazionale, introduciamo le variabili aleatorie Y_1, \dots, Y_n definite da

$$Y_j = \frac{\partial}{\partial \theta} \log f(X_j, \theta), \quad \forall j = 1, \dots, n$$

Y_1, \dots, Y_n sono variabili aleatorie i.i.d. a media nulla e varianza finita $I(\theta) \forall \theta$. Infatti

$$E_\theta(Y_j) = E_\theta \left(\frac{\partial}{\partial \theta} \log f(X_j, \theta) \right) = 0 \quad [\text{per l'ipotesi (iv)}]$$

e

$$\text{Var}_\theta(Y_j) = E_\theta(Y_j^2) = E_\theta \left[\left(\frac{\partial}{\partial \theta} \log f(X_j, \theta) \right)^2 \right] = I(\theta) \in (0, \infty) \quad [\text{per l'ipotesi (v)}]$$

Inoltre

$$\frac{\partial}{\partial \theta} \log L_\theta(X_1, \dots, X_n) = \frac{\partial}{\partial \theta} \log \prod_{j=1}^n f(X_j, \theta) = \sum_{j=1}^n \frac{\partial}{\partial \theta} \log f(X_j, \theta) = \sum_{j=1}^n Y_j$$

da cui deriviamo che

$$(14) \quad E_\theta \left(\frac{\partial}{\partial \theta} \log L_\theta(X_1, \dots, X_n) \right) = \sum_{j=1}^n E_\theta(Y_j) = 0$$

e

$$(15) \quad \text{Var}_\theta \left(\frac{\partial}{\partial \theta} \log L_\theta(X_1, \dots, X_n) \right) = \sum_{j=1}^n \text{Var}_\theta(Y_j) = nI(\theta)$$

L'ipotesi (vi) e l'equazione (14) insieme forniscono

$$\kappa'(\theta) = E_\theta \left(T \frac{\partial}{\partial \theta} \log L_\theta(X_1, \dots, X_n) \right) = \text{Cov} \left(T, \frac{\partial}{\partial \theta} \log L_\theta(X_1, \dots, X_n) \right)$$

cosicché

$$\begin{aligned} (\kappa'(\theta))^2 &= \left(\text{Cov} \left(T, \frac{\partial}{\partial \theta} \log L_\theta(X_1, \dots, X_n) \right) \right)^2 \\ &\leq \text{Var}_\theta(T) \text{Var}_\theta \left(\frac{\partial}{\partial \theta} \log L_\theta(X_1, \dots, X_n) \right) \quad [\text{per la proprietà (j) della covarianza}] \\ &= \text{Var}_\theta(T) nI(\theta) \quad [\text{per l'equazione (15)}] \end{aligned}$$

La disuguaglianza (10) segue in virtù dell'ipotesi (v).

Infine, per verificare che la condizione (11) è necessaria e sufficiente perché valga l'uguaglianza in (10), applichiamo la proprietà (jjj) della covarianza alle variabili $X = T$ e $Y = \frac{\partial}{\partial \theta} \log L_\theta(X_1, \dots, X_n)$. ■

Definizione 6.6 Uno stimatore T^* non distorto la cui varianza raggiunge il confine inferiore di Fréchet-Cramer-Rao è detto *efficiente*.

È importante ribadire che:

- a) Se uno stimatore T efficiente esiste, ovviamente esso è anche *UMVUE*.
- b) Con un campione casuale, la varianza dello stimatore efficiente è inversamente proporzionale al numero di osservazioni nel campione.

Lo stimatore di massima verosimiglianza: definizione.

Definizione 7.4 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 (che sta per *Maximum Likelihood*) o MLE (*Maximum Likelihood Estimator*).

- Ancora stima efficiente

Se esiste uno stimatore efficiente allora esso coincide essenzialmente con lo stimatore ML (con dimostrazione).

Proposizione 7.12 *Se le condizioni di regolarità necessarie perché la disuguaglianza di Fréchet-Cramer-Rao sussista sono soddisfatte e $\kappa(\theta)$ ammette uno stimatore T “efficiente” (cioè non distorto e la cui varianza raggiunge il confine inferiore di Fréchet-Cramer-Rao), allora esiste uno stimatore ML essenzialmente unico e coincide con T .*

Dimostrazione Sia T lo stimatore efficiente di $\kappa(\theta)$. Esso è essenzialmente unico e per il Teorema 6.1 abbiamo

$$(19) \quad P_\theta \left(\frac{\partial}{\partial \theta} \log L_\theta(x_1, \dots, x_n) = a(\theta, n)(T - \theta) \right) = 1 \quad \forall \theta \in \Theta$$

La funzione $a(\theta, n)$ è non nulla per ogni $\theta \in \Theta$ perché, in virtù di (19), abbiamo

$$0 < nI(\theta) = \text{Var}_\theta \left(\frac{\partial}{\partial \theta} \log L_\theta(x_1, \dots, x_n) \right) = \text{Var}_\theta(a(\theta, n)(T - \theta)) = a^2(\theta, n) \text{Var}_\theta(T)$$

Inoltre, se valgono le condizioni di regolarità (i) – (iii) del Teorema 6.1, per determinare uno stimatore $\hat{\theta}$ ML dobbiamo risolvere l'equazione

$$(20) \quad \frac{\partial}{\partial \theta} \log L_\theta(x_1, \dots, x_n) = 0$$

Tenuto conto che $a(\theta, n) \neq 0 \forall \theta$ e sostituendo $a(\theta, n)(T - \theta)$ a $\frac{\partial}{\partial \theta} \log L_\theta(x_1, \dots, x_n)$ in (20), otteniamo che necessariamente $P_\theta(\hat{\theta} = T) = 1 \forall \theta \in \Theta$. ■

La proprietà di invarianza della stima ML (con dimostrazione).

Se siamo interessati a stimare una caratteristica della popolazione $\kappa(\theta)$, possiamo partire dalla *funzione di verosimiglianza indotta da $\kappa(\theta)$* definita da

$$L_{\kappa}^*(x_1, \dots, x_n) := \sup_{\{\theta \in \Theta : \kappa(\theta) = \kappa\}} L_{\theta}(x_1, \dots, x_n)$$

In modo naturale definiamo *stimatore di massima verosimiglianza di $\kappa(\theta)$* una statistica che massimizza la funzione di verosimiglianza indotta da $\kappa(\theta)$ L_{κ}^* .

Se $\hat{\theta}(x_1, \dots, x_n)$ è uno stimatore ML di θ , allora

$$(a) \quad \sup_{\{\theta \in \Theta : \kappa(\theta) = \kappa\}} L_{\theta}(x_1, \dots, x_n) \leq \sup_{\theta \in \Theta} L_{\theta}(x_1, \dots, x_n), \text{ dal momento che } \{\theta \in \Theta : \kappa(\theta) = \kappa\} \text{ è un sottoinsieme di } \Theta,$$

$$(b) \quad L_{\hat{\theta}}(x_1, \dots, x_n) = \sup_{\{\theta \in \Theta : \kappa(\theta) = \kappa(\hat{\theta}(x_1, \dots, x_n))\}} L_{\theta}(x_1, \dots, x_n), \text{ dal momento che } \hat{\theta} \text{ massimizza } L_{\theta} \text{ e } \hat{\theta} \in \{\theta \in \Theta : \kappa(\theta) = \kappa(\hat{\theta}(x_1, \dots, x_n))\}$$

Pertanto, abbiamo

$$\begin{aligned} L_{\kappa}^*(x_1, \dots, x_n) &= \sup_{\{\theta \in \Theta : \kappa(\theta) = \kappa\}} L_{\theta}(x_1, \dots, x_n) \leq \sup_{\theta \in \Theta} L_{\theta}(x_1, \dots, x_n) = L_{\hat{\theta}}(x_1, \dots, x_n) = \\ &= \sup_{\{\theta \in \Theta : \kappa(\theta) = \kappa(\hat{\theta}(x_1, \dots, x_n))\}} L_{\theta}(x_1, \dots, x_n) = L_{\kappa(\hat{\theta})}^*(x_1, \dots, x_n) \end{aligned}$$

La disuguaglianza appena dimostrata ci dice che se $\hat{\theta}(x_1, \dots, x_n)$ è uno stimatore ML di θ , allora $\kappa(\hat{\theta})$ è uno stimatore ML di $\kappa(\theta)$. Questa proprietà va sotto il nome di *proprietà di invarianza degli stimatori di massima verosimiglianza*.

La stima efficiente nei modelli della famiglia esponenziale (escluso chi, pur essendo nello scaglione M-Z, ha seguito nello scaglione A-L).

Metateorema 6.11 *Gli stimatori efficienti esistono solo per le caratteristiche $\kappa(\theta)$ della famiglia esponenziale $f(x, \theta) = C(x) \exp\{A(\theta)g(x) + B(\theta)\}$ del tipo $\kappa(\theta) = -cB'(\theta)/A'(\theta) + d$. Inoltre questi stimatori hanno forma $(c/n) \sum_{j=1}^n g(X_j) + d$, $c, d \in \mathbb{R}$.*

La consistenza in media quadratica (spiegata durante l'esercitazione del 16/4).

Proposizione 7.13 *Sia X_1, \dots, X_n, \dots una successione di variabili aleatorie i.i.d. con comune funzione di densità $f(x, \theta)$, $\theta \in \Theta$ e sia $\{T_n\}_n$ la successione degli stimatori ML di $\kappa(\theta)$. Se $f(x, \theta)$ soddisfa le condizioni di regolarità della disuguaglianza di Fréchet-Cramer-Rao e altre ancora (di esistenza, continuità e limitatezza delle derivate seconda e terza di $f(x, \theta)$ rispetto a θ), allora la successione $\{T_n\}_n$ è*

1. *asintoticamente non distorta per $\kappa(\theta)$,*
2. *consistente in media quadratica per $\kappa(\theta)$,*
3. *asintoticamente gaussiana con media asintotica $\kappa(\theta)$ e varianza asintotica $\frac{[\kappa'(\theta)]^2}{nI(\theta)}$, cioè*

$$\lim_{n \rightarrow \infty} P \left(\frac{T_n - \kappa(\theta)}{\sqrt{\frac{[\kappa'(\theta)]^2}{nI(\theta)}}} \leq z \right) = \Phi(z) \quad \forall z \in \mathbb{R}$$

Definizione 4.2 *Sia X_1, \dots, X_n, \dots una successione di variabili aleatorie i.i.d. con comune funzione di densità $f(x, \theta)$, $\theta \in \Theta$ e sia T_n uno stimatore di $\kappa(\theta)$ che è funzione delle prime n osservazioni. La successione $\{T_n\}_n$ è *consistente in media quadratica per $\kappa(\theta)$* se*

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

- Distribuzioni campionarie degli stimatori

Casi notevoli di distribuzioni campionarie degli stimatori: modello gaussiano, modello esponenziale, modello di Poisson, modello binomiale.

Risultati asintotici: teorema limite centrale e distribuzione asintotica gaussiana per stimatori che sono trasformate lineari di una media campionaria; distribuzione asintotica dello stimatore ML.

- Intervalli di confidenza

Metodo della quantità pivotale con esempi: modello gaussiano, modello esponenziale.

Definizione 2.6 (Quantità pivotale) Sia X_1, \dots, X_n un campione casuale estratto dalla popolazione di densità $f(x, \theta)$ e sia Q_θ una funzione di X_1, \dots, X_n e del parametro θ , cioè $Q_\theta = q(X_1, \dots, X_n, \theta)$. Se la legge di Q_θ non dipende da θ , Q_θ è detta *quantità pivotale*.

Quantità asintoticamente pivotali con esempi: modello di Poisson e uso della distribuzione asintotica dello stimatore ML.

Intervallo approssimato per la media. Partiamo dal caso in cui la caratteristica per cui costruire un IC è la media teorica del modello $\mu = E(X_1)$. Sia X_1, \dots, X_n un campione casuale estratto da una popolazione che ha media μ e varianza σ^2 e σ^2 è non nulla. Sappiamo dal teorema centrale del limite che la media campionaria \bar{X} è asintoticamente gaussiana con media μ e varianza σ^2/n . Quindi, se abbiamo a disposizione un numero sufficientemente elevato di osservazioni, approssimativamente vale che $\frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}} \sim \mathcal{N}(0, 1)$. Se σ^2 non è nota, il risultato è preservato anche quando sostituiamo σ^2 con la varianza campionaria S^2 , cioè per n "grande", approssimativamente $\frac{\bar{X} - \mu}{\sqrt{S^2/n}} \sim \mathcal{N}(0, 1)$. Ripetendo quanto già fatto nella costruzione di un intervallo di confidenza per la media da popolazione gaussiana, otteniamo che

$$\left(\bar{x} - z_{\frac{1+\gamma}{2}} \frac{s}{\sqrt{n}}, \bar{x} + z_{\frac{1+\gamma}{2}} \frac{s}{\sqrt{n}} \right)$$

è un intervallo di confidenza per μ se la varianza è incognita, di livello approssimativamente pari a $\gamma 100\%$ (\bar{x} e s^2 rappresentano le realizzazioni campionarie di media e varianza campionarie).

Intervallo approssimato per una caratteristica $\kappa(\theta)$. Per costruire intervalli di confidenza approssimati per una generica caratteristica $\kappa(\theta)$, una strada si basa sull'uso dello stimatore di massima verosimiglianza per $\kappa(\theta)$, a patto che esista e la gaussianità asintotica dello stimatore valga.

Consideriamo un campione casuale numeroso X_1, \dots, X_n estratto da una popolazione di densità $f(x, \theta)$ e sia $\hat{\kappa}$ lo stimatore di massima verosimiglianza della caratteristica della popolazione $\kappa(\theta)$. Supponiamo che siano soddisfatte tutte le ipotesi di regolarità che garantiscono la gaussianità asintotica di $\hat{\kappa}$. Segue che, per n "grande", $\frac{\hat{\kappa} - \kappa(\theta)}{\sqrt{(\kappa'(\theta))^2/(nI(\theta))}} \sim \mathcal{N}(0, 1)$, dove $I(\theta)$ è l'informazione di Fisher. Allora, per ottenere un intervallo di confidenza $\gamma 100\%$ approssimato, dovremo a procedere a invertire la seguente

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

Se non riusciamo a invertire, potremo procedere a sostituire all'informazione di Fisher $I(\theta)$ e a $(\kappa'(\theta))^2$ le rispettive stime di massima verosimiglianza date da $I(\hat{\theta})$ e $\kappa'(\hat{\theta})^2$. Otterremo così il seguente intervallo di confidenza $\gamma 100\%$ approssimato:

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

- Verifica delle ipotesi

La falsificazione delle ipotesi scientifiche tramite la verifica di ipotesi statistiche.

Definizione 1.1 Un'ipotesi statistica è un'asserzione o congettura sulla f.d.r. incognita F . Se F , e quindi la corrispondente funzione di densità f , è nota a meno di un parametro $\theta = (\theta_1, \dots, \theta_m) \in \Theta \subset \mathbb{R}^m$, l'ipotesi statistica è un'asserzione su θ .

Ipotesi nulla e alternativa, regione critica e funzione test.

Ma in cosa consiste una regola di decisione? Procediamo a partizionare l'insieme \mathbb{R}^n di tutte le realizzazioni campionarie di X_1, \dots, X_n in due regioni \mathcal{G} e \mathcal{G}^c . Quindi effettuiamo il campionamento, registriamo i risultati x_1, \dots, x_n e se $(x_1, \dots, x_n) \in \mathcal{G}$ falsifichiamo H_0 , se, invece, $(x_1, \dots, x_n) \in \mathcal{G}^c$ non falsifichiamo H_0 . Riassumendo:

rifiutiamo H_0 se $(x_1, \dots, x_n) \in \mathcal{G}$ e accettiamo H_0 se $(x_1, \dots, x_n) \in \mathcal{G}^c$

Definizione 1.2 Se \mathcal{G} è un sottoinsieme di \mathbb{R}^n tale che se $(x_1, \dots, x_n) \in \mathcal{G}$ rifiutiamo H_0 , \mathcal{G} è detta *regione critica* o *di rifiuto* (di H_0) e \mathcal{G}^c è detta *regione di accettazione* (di H_0).

Ripetiamo: una sola decisione può e deve essere presa: o rifiutare H_0 , che equivale ad accettare H_1 , o accettare H_0 , che equivale a rifiutare H_1 .

Definizione 1.3 Un *test di ipotesi* (*hypothesis test*) è una tripletta costituita dal campione X_1, \dots, X_n , dalle ipotesi nulla e alternativa H_0, H_1 e da un sottoinsieme \mathcal{G} di \mathbb{R}^n che ha l'interpretazione di regione critica: $(X_1, \dots, X_n; H_0, H_1; \mathcal{G})$.

Errori di primo e secondo tipo.

Può succedere di prendere una decisione sbagliata. L'errore può essere di due tipi:

Errore di I tipo o prima specie: quando rifiutiamo H_0 ma H_0 è vera;

Errore di II tipo o seconda specie: quando accettiamo H_0 ma H_0 è falsa.

$$\begin{aligned} (1) \quad & \alpha(\theta) = P_\theta(\text{"Rifiutare } H_0\text{"}) = P_\theta((x_1, \dots, x_n) \in \mathcal{G}), \quad \theta \in \Theta_0 \\ (2) \quad & \beta(\theta) = P_\theta(\text{"Accettare } H_0\text{"}) = P_\theta((x_1, \dots, x_n) \in \mathcal{G}^c), \quad \theta \in \Theta_1 \end{aligned}$$

Livello del test.

La probabilità di errore di I tipo non può superare il valore

$$(3) \quad \alpha := \sup_{\theta \in \Theta_0} \alpha(\theta) = \sup_{\theta \in \Theta_0} P_\theta((x_1, \dots, x_n) \in \mathcal{G})$$

Definizione 1.4 L'estremo superiore della probabilità di errore di I tipo $\alpha := \sup_{\theta \in \Theta_0} \alpha(\theta)$ è detto *livello di significatività* del test (*level of significance*) o *ampiezza* o *dimensione* (*size*) della regione critica.

Funzione di potenza.

Definizione 1.5 La funzione $\pi(\theta)$ definita da $\pi(\theta) = 1 - \beta(\theta) = P_\theta((x_1, \dots, x_n) \in \mathcal{G}), \theta \in \Theta_1$ è la *funzione di potenza* del test (*power function*).

La funzione di potenza di un test ha quindi l'interpretazione di probabilità di prendere la corretta decisione di rifiutare l'ipotesi nulla quando effettivamente l'ipotesi nulla è falsa.

Esemplifichiamo di seguito le nozioni fin qui introdotte sul familiare modello gaussiano.

Test uniformemente più potente e lemma di Neyman-Pearson.

Noi costruiremo un test uniformemente più potente nel caso di ipotesi H_0, H_1 entrambe semplici. In particolare il risultato è contenuto nel *Lemma di Neyman-Pearson*.

Lemma 2.1 (Lemma di Neyman-Pearson) Sia X_1, \dots, X_n un campione casuale con verosimiglianza $L_\theta(x_1, \dots, x_n)$ e supponiamo di voler verificare $H_0 : \theta = \theta_0$ contro $H_1 : \theta = \theta_1$. Sia \mathcal{G} la regione critica definita da

$$\mathcal{G} = \left\{ (x_1, \dots, x_n) : \frac{L_{\theta_0}(x_1, \dots, x_n)}{L_{\theta_1}(x_1, \dots, x_n)} \leq \delta \right\}.$$

Allora \mathcal{G} è la regione critica che genera massima potenza fra tutte le regioni critiche di ampiezza minore o uguale all'ampiezza di \mathcal{G} .

Test UPP per il parametro della distribuzione esponenziale con ipotesi nulla semplice e ipotesi alternativa composta.

Esempio di applicazione del lemma per le famiglie esponenziali (escluso chi, pur essendo nello scaglione M-Z, ha seguito nello scaglione A-L).

Il test del rapporto di verosimiglianza generalizzato.

Osservazione 2.5 (Legame con gli stimatori di massima verosimiglianza) In alcuni problemi di verifica di ipotesi, nella costruzione del test del rapporto di verosimiglianza usiamo gli stimatori di massima verosimiglianza. Per esempio, consideriamo un campione casuale X_1, \dots, X_n i.i.d. $\sim f(x, \theta)$ con $\theta \in \mathbb{R}$ e sia $\hat{\theta}_{ML}$ lo stimatore di massima verosimiglianza di θ . Supponiamo poi di voler verificare $H_0 : \theta = \theta_0$ contro $H_1 : \theta \neq \theta_0$. In questo caso il rapporto di verosimiglianza si riduce a

$$\Lambda(x_1, \dots, x_n) = \frac{L_{\theta_0}(x_1, \dots, x_n)}{\sup_{\theta \in \mathbb{R}} L_\theta(x_1, \dots, x_n)} = \frac{L_{\theta_0}(x_1, \dots, x_n)}{L_{\hat{\theta}_{ML}}(x_1, \dots, x_n)}$$

- Ancora verifica delle ipotesi

Il p-value per test unilaterali e bilaterali.

Per descrivere il p-value, partiamo dall'unica procedura test generale che abbiamo elaborato, cioè il test del rapporto di verosimiglianza. Supponiamo di voler verificare l'ipotesi nulla $H_0 : \theta \in \Theta_0$ e sia λ il valore assunto dal rapporto di verosimiglianza in corrispondenza della realizzazione campionaria (x_1, \dots, x_n) : $\lambda = \Lambda((x_1, \dots, x_n))$. Sia p il livello del test del rapporto di verosimiglianza con frontiera $\delta = \lambda$, cioè $p = \sup_{\theta \in \Theta_0} P_\theta(\Lambda \leq \lambda)$ e sia $\{\Lambda \leq \delta(\alpha)\}$ la regione critica di livello α . Ho scritto $\delta(\alpha)$, perché il valore di frontiera δ cresce con α . Quindi la disuguaglianza $\lambda \leq \delta(\alpha)$ è soddisfatta se e solo se $p \leq \alpha$. Ma se $\lambda \leq \delta(\alpha)$ noi rifiutiamo H_0 a livello α . Concludiamo che

per ogni $\alpha \geq p$ rifiutiamo H_0 e per ogni $\alpha < p$ accettiamo H_0 .

Esempi: test sulla media della distribuzione esponenziale e della gaussiana.

Relazione tra intervalli di confidenza e verifica delle ipotesi.

Sia (x_1, \dots, x_n) la realizzazione di un campione casuale X_1, \dots, X_n estratto da $f(x, \theta)$ e sia $IC(x_1, \dots, x_n)$ un intervallo di confidenza per θ di livello γ , cioè tale che $P_\theta(\theta \in IC(X_1, \dots, X_n)) = \gamma$. Siamo interessati a verificare: $H_0 = \theta_0$ contro $H_1 : \theta \neq \theta_0$. Consideriamo il sottoinsieme di \mathbb{R}^n

$$(11) \quad A_0 = \{(x_1, \dots, x_n) \in \mathbb{R}^n : \theta_0 \in IC(x_1, \dots, x_n)\}$$

Allora $(x_1, \dots, x_n) \in A_0$ se e solo se $\theta_0 \in IC(x_1, \dots, x_n)$ e quindi

$$P_{\theta_0}(A_0) = P_{\theta_0}(\theta_0 \in IC(X_1, \dots, X_n)) = \gamma$$

Segue che $G := A_0^c$ è una regione critica per $H_0 = \theta_0$ contro $H_1 : \theta \neq \theta_0$ di ampiezza $\alpha = 1 - \gamma$.

Praticamente, per verificare $H_0 : \theta = \theta_0$ contro $H_1 : \theta \neq \theta_0$ a un livello di significatività α , determiniamo un intervallo di confidenza bilatero $(1 - \alpha)100\%$ per θ . Se l'intervallo contiene θ_0 accettiamo H_0 , altrimenti la rifiutiamo.

- Verifica delle ipotesi con due campioni.

Differenza tra medie di v.a. gaussiane indipendenti, con varianze uguali e incognite, varianze note, varianze diverse e incognite. Rapporto tra varianze di v.a. gaussiane indipendenti e la F di Fisher. Dati gaussiani accoppiati: differenza tra medie; indipendenza.

- Statistica non parametrica

Funzione di ripartizione empirica e sue proprietà.

Definizione 1.1 La *funzione di ripartizione empirica (o campionaria)* associata al campione \hat{F}_n è una funzione su \mathbb{R} a valori in $[0, 1]$ definita da

$$(1) \quad \hat{F}_n(x) = \frac{\#\{j : X_j \leq x\}}{n} \quad \forall x \in \mathbb{R}$$

Possiamo rappresentare \hat{F}_n in termini di $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ nel seguente modo:

$$(2) \quad \hat{F}_n(x) = \begin{cases} 0 & x < X_{(1)} \\ \frac{k}{n} & X_{(k)} \leq x < X_{(k+1)} \quad (k = 1, \dots, n-1) \\ 1 & x \geq X_{(n)} \end{cases}$$

Osservate che la funzione \hat{F} è aleatoria e dipende soltanto dal campione casuale, quindi è una statistica. Inoltre, qualunque sia la realizzazione campionaria, è una funzione a gradini, compresa fra 0 e 1, monotona crescente e continua da destra. Cioè ogni realizzazione di \hat{F}_n può essere pensata come una f.d.r. discreta.

- il momento r -esimo di \hat{F}_n è

$$M_r = \frac{1}{n} \sum_{j=1}^n X_j^r$$

cioè quello che avevamo chiamato momento campionario r -esimo. In particolare,

- la media di \hat{F}_n è $M_1 = \bar{X}$, cioè la media campionaria;
- la varianza di \hat{F}_n è

$$\frac{1}{n} \sum_{j=1}^n (X_j - \bar{X})^2 = \frac{(n-1)S^2}{n}$$

Analizziamo ora le proprietà probabilistiche della statistica $\hat{F}_n(x)$.

Fissato x , introduciamo le v.a. Y_1, \dots, Y_n definite da $Y_j = \mathbf{1}_{(-\infty, x]}(X_j)$ per $j = 1, \dots, n$. Le v.a. Y_1, \dots, Y_n costituiscono un campione casuale estratto dalla densità bernoulliana di parametro $F(x)$ e $\hat{F}_n(x)$ può essere interpretata come la media campionaria delle Y_j , cioè $\hat{F}_n(x) = \bar{Y}$. Seguono da questa rappresentazione probabilistica di \hat{F}_n le seguenti proprietà:

1. Per ogni $x \in \mathbb{R}$ fissato, $n\hat{F}_n(x)$ rappresenta il numero di osservazioni di valore al più pari a x e ha distribuzione binomiale di parametri $nF(x)$;
2. $\hat{F}_n(x)$ è stimatore non distorto e consistente in media quadratica di $F(x)$; infatti $E_F(\hat{F}_n(x)) = F(x)$ e $\text{Var}_F(\hat{F}_n(x)) = F(x)(1-F(x))/n \rightarrow 0$, per $n \rightarrow \infty \forall F, \forall x$.

In realtà vale un risultato più forte di convergenza di \hat{F}_n a F uniforme in x , fornito dal seguente teorema di Glivenko-Cantelli:

3. Sia X_1, X_2, \dots una sequenza di v.a. i.i.d. con comune f.d.r. F . Allora

$$P\left(\lim_{n \rightarrow \infty} \sup_{x \in \mathbb{R}} |\hat{F}_n(x) - F(x)| = 0\right) = 1$$

Quindi, la “funzione aleatoria” \hat{F}_n è uno stimatore consistente “globalmente” in senso forte per F .

6. La successione $\{\hat{F}_n(x)\}_n$ è asintoticamente gaussiana, cioè

$$\lim_{n \rightarrow \infty} P_F\left(\sqrt{n} \frac{\hat{F}_n(x) - F(x)}{\sqrt{F(x)[1-F(x)]}} \leq z\right) = \Phi(z), \quad \forall z \in \mathbb{R}, \quad \forall x \in \mathbb{R} \text{ t.c. } 0 < F(x) < 1$$

Per convincersi di ciò è sufficiente applicare il teorema centrale del limite alla media campionaria di v.a. bernoulliane i.i.d.

Test di adattamento: Kolmogorov-Smirnov, Lilliefors per dati gaussiani con media e varianza incognite;

³ Siano X_1, \dots, X_n i.i.d. $\sim F$. Per affrontare il problema ipotetico $H_0 : F = F_0$ contro $H_1 : F \neq F_0$, introduciamo la statistica test

$$D_n := \sup_{x \in \mathbb{R}} |\hat{F}_n(x) - F_0(x)|$$

Se H_0 è vera, sono verosimili valori piccoli di D_n . Inoltre, il Teorema di Glivenko-Cantelli

La statistica test D_n è detta *statistica di Kolmogorov-Smirnov* e il test basato su D_n è il *test di Kolmogorov-Smirnov*.

chi² per dati discreti,

⁵ A differenza del test di Kolmogorov-Smirnov che può essere usato solo se i dati sono continui e l'ipotesi nulla è semplice, il test di buon adattamento χ^2 permette di affrontare anche i problemi ipotetici: a) $H_0 : F = F_0$ contro $H_1 : F \neq F_0$ e b) $H_0 : F \in \mathcal{F}_0$ contro $H_1 : F \notin \mathcal{F}_0$ per qualunque tipo di dati discreti e continui.

$$H_0 : p_i = p_{0i} \quad \forall i = 1, \dots, k \quad \text{contro} \quad H_1 : p_i \neq p_{0i} \text{ per qualche } i.$$

Dato un campione casuale X_1, \dots, X_n estratto da F calcoliamo la *frequenza assoluta campionaria* di ogni modalità a_i , cioè quante osservazioni assumono valore a_i :

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

e misuriamo lo scostamento fra i dati e il modello specificato in H_0 mediante la *statistica di Pearson*

$$(4) \quad Q := \sum_{i=1}^k \frac{(N_i - np_{0i})^2}{np_{0i}}$$

$\text{rifiutiamo } H_0 \text{ se } Q > \chi_{k-1}^2(1 - \alpha)$

In altri termini, il test χ^2 calcola lo scostamento fra f.d.r. empirica \hat{F}_n e teorica F_0 in termini di scostamento fra frequenze relative campionarie $(\hat{p}_{n1}, \dots, \hat{p}_{nk})$ e densità teoriche p_{01}, \dots, p_{0k} .

χ^2 per dati qualunque (con dati grezzi o raggruppati in classi), con distribuzione interamente specificata o parzialmente (parametri incogniti).

Test χ^2 per H_0 semplice. Consideriamo prima di tutto il problema di verifica delle ipotesi $H_0 : F = F_0$ contro $H_1 : F \neq F_0$ con F_0 completamente specificata. Sia X_1, \dots, X_n un campione casuale e A_1, \dots, A_k k intervalli disgiunti di \mathbb{R} . Per ogni $i = 1, \dots, k$ calcoliamo

- a) il numero N_i di osservazioni che cadono in A_i ;
- b) la probabilità teorica sotto H_0 che X cada in A_i cioè $p_{0i} = P_{F_0}(X \in A_i)$ (osserviamo che se H_0 è vera, il numero medio delle osservazioni che cadono in A_i è np_{0i});
- c) lo scostamento fra \hat{F}_n e F_0 in termini di scostamento fra N_i e np_{0i} mediante la statistica

di Pearson $Q := \sum_{i=1}^k (N_i - np_{0i})^2 / (np_{0i})$. Ad un livello di significatività α e con un campione numeroso,

Test χ^2 per H_0 composta. Effettuiamo ora un test di buon adattamento χ^2 di un modello specificato a meno di qualche parametro incognito e abbiamo raggruppati i dati in k intervalli $(a_0, a_1], (a_1, a_2], \dots, (a_{k-1}, a_k]$.

Per esempio, vogliamo verificare se

$$H_0 : X \sim \mathcal{N}(\mu, \sigma^2), \mu \in \mathbb{R}, \sigma^2 > 0 \quad \text{contro} \quad H_1 : X \text{ non è gaussiana}$$

Per implementare il test calcoliamo per ogni $i = 1, \dots, k$, il numero N_i di osservazioni a valori in $(a_{i-1}, a_i]$ e la probabilità che X cada in $(a_{i-1}, a_i]$ cioè

$$(5) \quad p_i(\mu, \sigma^2) = \Phi\left(\frac{a_i - \mu}{\sigma}\right) - \Phi\left(\frac{a_{i-1} - \mu}{\sigma}\right)$$

Quindi stimiamo i parametri μ, σ^2 e calcoliamo le p_i in (5) usando degli stimatori $\hat{\mu}, \hat{\sigma}^2$. Infine calcoliamo la statistica di Pearson

$$(6) \quad Q^* = \sum_{i=1}^k \frac{[N_i - np_i(\hat{\mu}, \hat{\sigma}^2)]^2}{np_i(\hat{\mu}, \hat{\sigma}^2)}$$

La f.d.r. asintotica di Q^* in (6) sotto H_0 è ancora χ^2 ma con diversi gradi di libertà rispetto al caso di ipotesi nulla semplice: se μ e σ^2 sono stimati “in modo opportuno”, perdiamo un grado di libertà per ogni parametro stimato e quindi, asintoticamente $Q^* \sim \chi_{k-1-2}^2$.

Ma come possiamo stimare μ e σ^2 ?

Se abbiamo i dati grezzi usiamo gli stimatori ML: $\hat{\mu}_{ML} = \bar{X}$ e $\hat{\sigma}_{ML}^2 = \sum_{j=1}^n (X_j - \bar{X})^2 / n$. Se, invece, disponiamo solo dei dati raggruppati, una ricetta semplice per stimare μ e σ^2 è la seguente: calcoliamo il valore centrale di ogni intervallo $(a_{i-1}, a_i]$ (di lunghezza finita), cioè

$$c_i = \frac{a_{i-1} + a_i}{2}, \quad i = 1, \dots, k$$

e lo pesiamo con la numerosità campionaria N_i dell'intervallo. Poi, calcoliamo media campionaria e momento secondo campionario di questi dati, cioè

$$M_1 = \frac{\sum_{i=1}^k c_i N_i}{n}, \quad M_2 = \frac{\sum_{i=1}^k c_i^2 N_i}{n}$$

e applichiamo il metodo dei momenti. Per un campione casuale gaussiano otteniamo

$$\hat{\mu} = M_1, \quad \hat{\sigma}^2 = M_2 - M_1^2$$

Il test di buon adattamento χ^2 esemplificato per il modello gaussiano può essere in generale usato ogni qualvolta l'ipotesi nulla specifichi una f.d.r. dipendente da m parametri incogniti, $\theta_1, \dots, \theta_m$: si stimano $\theta_1, \dots, \theta_m$ usando il metodo dei momenti con i valori centrali delle classi, ciascuno pesato per la numerosità campionaria della classe, e si usano le stime ottenute $\hat{\theta}_1, \dots, \hat{\theta}_m$ per calcolare le probabilità p_1, \dots, p_k specificate da H_0 . Si può dimostrare che sotto H_0 la statistica

$$Q^* = \sum_{i=1}^k \frac{[N_i - np_i(\hat{\theta}_1, \dots, \hat{\theta}_m)]^2}{np_i(\hat{\theta}_1, \dots, \hat{\theta}_m)}$$

ha f.d.r. asintotica χ_{k-1-m}^2 : perdiamo un grado di libertà per ogni parametro stimato. A questo punto, la regola di rifiuto da adottare è la seguente:

Per n grande, a livello α ,

$\text{rifiutiamo } H_0 : X \sim F(x; \theta_1, \dots, \theta_m) \text{ a livello } \alpha \text{ se } Q^* > \chi_{k-1-m}^2(1 - \alpha)$

Il p -value di questo test asintotico è dato da $1 - F_{\chi_{k-1-m}^2}(Q^*)$.

Statistica non parametrica II

Bande di confidenza per la funzione di ripartizione.

Test chi² di indipendenza.

Nel caso di dati congiuntamente gaussiani i test di indipendenza e concordanza si traducono in test sul coefficiente di correlazione lineare ρ . Vale infatti che

tistica test opportuna. Nel caso di un campione accoppiato gaussiano $(X_1, Y_1), \dots, (X_n, Y_n)$ con parametri tutti incogniti, uno stimatore per ρ è dato dal *coefficiente di correlazione campionario o empirico*:

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

La statistica R è in modulo minore o uguale di 1: $-1 \leq R \leq 1$ e per quanto concerne la sua distribuzione vale il seguente risultato:

Teorema 8.2 Sia $(X_1, Y_1), \dots, (X_n, Y_n)$ i.i.d. $\sim \mathcal{N}$ e $\rho = 0$. Allora

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

$\mathcal{G}_1 = \left\{ \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \geq t_{n-2}(1-\alpha) \right\}$ è una regione critica per $H_0 : \rho = 0$ o $H_0 : \rho \leq 0$ contro $H_1 : \rho > 0$

$\mathcal{G}_2 = \left\{ \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} < -t_{n-2}(1-\alpha) \right\}$ è una regione critica per $H_0 : \rho = 0$ o $H_0 : \rho \geq 0$ contro $H_1 : \rho < 0$

$\mathcal{G}_3 = \left\{ \left| \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \right| > t_{n-2}(1 - \frac{\alpha}{2}) \right\}$ è una regione critica per $H_0 : \rho = 0$ contro $H_1 : \rho \neq 0$

Test di Wilcoxon-Mann-Whitney.

¹¹ Estraiamo due campioni casuali indipendenti X_1, \dots, X_m da F e Y_1, \dots, Y_n da G e riuniamo tutte le osservazioni in un unico campione di ampiezza $m+n$. Registriamo il *rango* ("rank" o *grado*) di ogni osservazione, cioè la posizione che essa occupa nella classifica di tutte le $m+n$ osservazioni dalla più piccola alla più grande, chiamiamo R_i il rango di X_i e sommiamo i ranghi delle X_i : $T_X = \sum_{i=1}^m R_i$.

Rifiuto $H_0 : F(x) = G(x) \forall x$ e accetto $H_1 : "F(x) \geq G(x) \forall x \in \mathbb{R} \text{ e } F(x) > G(x) \text{ per qualche } x"$ se $T_X < w_\alpha$

Rifiuto $H_0 : F(x) = G(x) \forall x$ e accetto $H_1 : "F(x) \leq G(x) \forall x \in \mathbb{R} \text{ e } F(x) < G(x) \text{ per qualche } x"$ 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 $T_X \notin [w_{\alpha/2}, w_{1-\alpha/2}]$

Il test descritto è noto come *test della somma dei ranghi di Wilcoxon-Mann-Whitney*.