

Appunti per l'Orale di Statistica

Matteo Gianello

23 settembre 2013

Indice

| | | |
|----------|--|-----------|
| 1 | Media e varianza campionaria | 3 |
| 1.1 | Media campionaria | 3 |
| 1.2 | Varianza campionaria | 3 |
| 1.3 | Legge dei grandi numeri | 4 |
| 1.4 | Distribuzioni campionarie | 4 |
| 2 | Intervalli di confidenza | 6 |
| 2.1 | Intervalli di confidenza per la media | 6 |
| 2.2 | Intervalli di confidenza per la varianza | 6 |
| 3 | Teoria della stima puntuale | 8 |
| 3.1 | Definizioni | 8 |
| 3.1.1 | Errore quadratico medio | 8 |
| 3.2 | Stimatori non distorti | 8 |
| 3.3 | Proprietà asintotiche degli stimatori | 9 |
| 3.4 | Diseguaglianza di Fréchet-Cramer-Rao | 9 |
| 4 | Verifica di ipotesi | 11 |
| 4.1 | Definizioni | 11 |
| 4.2 | Lemma di Neyuman-Pearson | 11 |

1 Media e varianza campionaria

1.1 Media campionaria

Definiamo *campione casuale* lungo n un insieme di n variabili aleatorie i.i.d. estratte da una popolazione di densità f . La media $\mu = E(X_1)$ uguale per tutte le X_i del campione prende il nome di *media della popolazione* e la comune varianza σ^2 è detta *varianza della popolazione*.

Si chiama *media campionaria* di un campione casuale $X_1 \dots X_n$ la quantità:

$$\frac{X_1 + \dots + X_n}{n}$$

e si indica con \bar{X}_n . La media campionaria dipende dal campione è perciò una variabile aleatoria.

Valore atteso e varianza della media campionaria possono essere facilmente calcolate tramite le proprietà di valore atteso e varianza:

$$E(\bar{X}_n) = E\left(\frac{\sum_{j=1}^n X_j}{n}\right) = \frac{E\left(\sum_{j=1}^n X_j\right)}{n} = \frac{n \times \mu}{n} = \mu$$

$$Var(\bar{X}_n) = Var\left(\frac{\sum_{j=1}^n X_j}{n}\right) = \frac{n Var(X_1)}{n^2} = \frac{\sigma^2}{n}$$

1.2 Varianza campionaria

Dato un campione casuale $X_1 \dots X_n$ si definisce *varianza campionaria* la quantità:

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

per dimostrare la non distorsione di questo stimatore dobbiamo innanzitutto dimostrare le seguenti espressioni:

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

$$E\left(\sum_{j=1}^n (X_j - \bar{X})^2\right) = (n-1)\sigma^2 \quad (2)$$

Dimostrazione 1

$$\sum_{j=1}^n (X_j - \bar{X})^2 = \sum_{j=1}^n [(X_j - \mu) - (\bar{X} - \mu)]^2$$

$$\begin{aligned}
&= \sum_{j=1}^n [(X_j - \mu)^2 + (\bar{X} - \mu)^2 - 2(X_j - \mu)(\bar{X} - \mu)] \\
&= \sum_{j=1}^n (X_j - \mu)^2 + n(\bar{X} - \mu)^2 - 2(X_j - \mu) \sum_{j=1}^n (\bar{X} - \mu) \\
&= \sum_{j=1}^n (X_j - \mu)^2 + n(\bar{X} - \mu)^2 - 2n(X_j - \mu)^2 \\
&= \sum_{j=1}^n (X_j - \mu)^2 - n(\bar{X} - \mu)^2
\end{aligned}$$

Dimostrazione 2

$$\begin{aligned}
E\left(\sum_{j=1}^n (X_j - \bar{X})^2\right) &= E\sum_{j=1}^n (X_j - \mu)^2 - nE(\bar{X} - \mu)^2 \\
&= \sum_{j=1}^n E(X_j - \mu)^2 - n\text{Var}(\bar{X}) = n\sigma^2 - \sigma^2
\end{aligned}$$

Da queste due formule ricaviamo che $E(S^2) = \sigma^2$ condizione necessaria e sufficiente per la non distorsione dello stimatore.

1.3 Legge dei grandi numeri

Sia X_1, X_2, \dots una successione di variabili aleatorie i.i.d. con media μ e varianza σ^2 finite e sia \bar{X}_n la media campionaria, allora per ogni $\epsilon > 0$:

$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| > \epsilon) = 0$$

Dalla disuguaglianza di Chebychev che afferma:

$$P(|X - E(X)| > \epsilon) \leq \frac{\text{Var}(X)}{\epsilon^2}$$

1.4 Distribuzioni campionarie

La prima densità che introduciamo è la densità $\Gamma(\alpha, \beta)$ questa densità è utile in quanto permette di ricavare altre densità favose da essa. La sua funzione di densità ha la forma seguente:

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

Mentre la sua funzione generatrice dei momenti è:

$$M(t) = E(e^{tX}) = \frac{1}{(1 - \beta t)^\alpha} \quad \forall t < 1/\beta$$

Nel caso in cui $\alpha = 1$ abbiamo che la densità *Gamma* diventa un'esponenziale di parametro beta $\xi(\beta)$.

Nel caso invece in cui sia $\alpha = \frac{n}{2}$ la densità che troviamo è una χ_n^2 .

Se $X \sim N(0, 1)$ allora $X^2 \sim \chi_1^2$; se prendiamo un campione casuale $X_1, \dots, X_n \sim N(0, 1)$ allora $\sum_{j=1}^n X_j^2 \sim \chi_n^2$

Vediamo ora alcune proprietà che portano alle affermazioni appena fatte

1. Se $X \sim \Gamma(\alpha, \beta)$, $c > 0$ e $Y = cX$ allora $Y \sim \Gamma(\alpha, c\beta)$;
2. Se X, Y sono v.a. indipendenti con $X \sim \Gamma(\alpha, \beta)$ e $Y \sim \Gamma(c, \beta)$ allora $X + Y \sim \Gamma(\alpha + c, \beta)$ e viceversa.

Dimostrazione

1. $M_Y(t) = E(e^{tcX}) = M_X(ct) = \frac{1}{(1-\beta ct)^\alpha} \quad t < \frac{1}{c\beta}$

2.

$$\begin{aligned} M_{X+Y}(t) &= E(e^{t(X+Y)}) = E(e^{tX} e^{tY}) = E(e^{tX}) E(e^{tY}) \\ &= M_X(t) M_Y(t) = \frac{1}{(1-\beta t)^\alpha} \times \frac{1}{(1-\beta ct)^c} = \frac{1}{(1-\beta ct)^\alpha + c} \end{aligned}$$

2 Intervalli di confidenza

2.1 Intervalli di confidenza per la media

Abbiamo visto prima come nel caso di un campione casuale X_1, \dots, X_n estratto da una popolazione di densità gaussiana di parametri μ, σ si possa stimare la media con la *media campionaria* \bar{X} e la varianza con la *varianza campionaria* S^2 .

Questo tipo di stima però non ha molto senso in quanto sappiamo che:

$$P_{\mu, \sigma^2}(\bar{X} = c) = 0$$

Ovvero è nulla la probabilità che \bar{X} assuma il vero valore di μ .

Partiamo dal caso in cui solo la media μ è incognita mentre la varianza è nota e pari ad un valore fissato; noi possiamo allora stabilire una regione di probabilità in cui possiamo stabilire con una certa precisione la probabilità che il vero valore di μ sia in quella regione.

$$P_{\mu, \sigma^2} \left(-\epsilon < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < \epsilon \right) = \gamma$$

Dove γ è la confidenza con la quale siamo certi che il parametro stimato cada nell'intervallo. Svolgendo i calcoli sopra per isolare μ troviamo che gli estremi dell'intervallo di confidenza sono:

$$\mu \in \left(\bar{X} - z_{\frac{1+\gamma}{2}} \frac{\sigma}{\sqrt{n}} ; \bar{X} + z_{\frac{1+\gamma}{2}} \frac{\sigma}{\sqrt{n}} \right)$$

Nel caso in cui la media non sia nota si può utilizzare la varianza campionaria per stimare un IC per la media in questo caso le formule diventano:

$$P_{\mu, \sigma^2} \left(-\epsilon < \frac{\bar{X} - \mu}{S/\sqrt{n}} < \epsilon \right) = \gamma$$

Questa volta la quantità $\sqrt{n}(\bar{X} - \mu)/S$ è una t-student con n-1 gradi di libertà t_{n-1} perciò l'intervallo di confidenza diventa:

$$\mu \in \left(\bar{X} - t_{n-1} \left(\frac{1+\gamma}{2} \right) \frac{\sigma}{\sqrt{n}} ; \bar{X} + t_{n-1} \left(\frac{1+\gamma}{2} \right) \frac{\sigma}{\sqrt{n}} \right)$$

2.2 Intervalli di confidenza per la varianza

Nel caso in cui vogliamo trovare un intervallo di confidenza per la varianza quando la media è incognita, partiamo dalla quantità aleatoria $S^2(n-1)/\sigma^2 \sim \chi_{n-1}^2$ fissato un certo valore γ dobbiamo fissare gli estremi dell'intervallo che come prima:

$$P_{\mu, \sigma^2} \left(a < \frac{S^2(n-1)}{\sigma^2} < b \right) = \gamma$$

Essendo la distribuzione χ^2 una distribuzione asimmetrica possiamo ritrovarci in tre casi specifici:

1.

$$\sigma^2 \in \left(\frac{S^2(n-1)}{\chi_{n-1}^2(\gamma)} ; +\infty \right)$$

2.

$$\sigma^2 \in \left(0 ; \frac{S^2(n-1)}{\chi_{n-1}^2(1-\gamma)} \right)$$

3.

$$\sigma^2 \in \left(\frac{S^2(n-1)}{\chi_{n-1}^2(\frac{1+\gamma}{2})} ; \frac{S^2(n-1)}{\chi_{n-1}^2(\frac{1-\gamma}{2})} \right)$$

Nel caso di media nota i ragionamenti appena fatti sono ancora validi ma stimiamo σ^2 con la quantità:

$$S_0^2 = \frac{\sum_{j=1}^n (X_j - \mu)^2}{n}$$

3 Teoria della stima puntuale

3.1 Definizioni

Sia X una v.a. con funzione di ripartizione F e densità di probabilità f non completamente specificata, ovvero con un parametro θ m -dimensionale a valori in Θ sottoinsieme di \mathbb{R}^m . Definiamo la statistica T come una variabile aleatoria funzione del campione $T = g(X_1, \dots, X_n)$. La distribuzione di una statistica T è detta *distribuzione (o legge) campionaria*. Una statistica T non dipende mai dal parametro incognito mentre la distribuzione campionaria in generale dipenderà da θ .

Una funzione $\kappa : \Theta \rightarrow \mathbb{R}$ è detta *caratteristica della popolazione*.

Siano $X_1, \dots, X_n \sim f(x, \theta)$, $\theta \in \Theta$ 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 dallo stimatore è detta stima di $\kappa(\theta)$.

3.1.1 Errore quadratico medio

Potremmo trovarci, in un problema di stima, nella situazione di dover decidere fra stimatori diversi della stessa caratteristica $\kappa(\theta)$. Utilizzeremo come criterio di scelta la media della prossimità di T a $\kappa(\theta)$ che esprimiamo in termini di $E_\theta[(T - \kappa(\theta))^2]$.

Se T è stimatore di $\kappa(\theta)$ tale che $E_\theta[(T - \kappa(\theta))^2] < \infty \forall \theta \in \Theta$, allora $E_\theta[(T - \kappa(\theta))^2]$ è detto *errore quadratico medio* di T rispetto a $\kappa(\theta)$.

Il MSE di uno stimatore T esiste solo se T ha media e varianza finite, o, equivalentemente, se e solo se ha momento secondo finito.

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

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

in cui la quantità $[E_\theta(T) - \kappa(\theta)]$ è detta *distorsione di bias*.

Tra tutti gli stimatori di $\kappa(\theta)$ ci piacerebbe scegliere quello con MSE minore. Questa cosa equivale a minimizzare contemporaneamente varianza e distorsione; ma questo è impossibile per ogni θ perciò ci accontentiamo di utilizzare la sottoclasse di stimatori che hanno distorsione nulla; questa classe è detta di stimatori *non distorti o corretti*. L'errore quadratico medio di uno stimatore non distorto coincide con la sua varianza.

3.2 Stimatori non distorti

Una statistica T che ammette media per ogni θ in Θ è detta *stiamtore non distorto* della caratteristica $\kappa(\theta)$ se

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

Se $X_1, \dots, X_n \text{ iid} \sim f(x, \theta)$, $\theta \in \Theta$ 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.

La varianza campionaria S^2 è stimatore non distorto della varianza teorica.

3.3 Proprietà asintotiche degli stimatori

Definizione 3.1. Sia X_1, \dots, X_n 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 *asintoticamente non distorta* per $\kappa(\theta)$ se

$$\lim_{n \rightarrow \infty} E_\theta(T_n) = \kappa(\theta) \quad \forall \theta \in \Theta$$

Definizione 3.2. Sia X_1, \dots, X_n 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

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

Definizione 3.3. Sia X_1, \dots, X_n una successione di variabili aleatorie i.i.d. con comune funzione di densità $f(x, \theta)$, $\theta \in \Theta$ e sia T_n una statistica funzione soltanto delle prime n osservazioni. La successione $\{T_n\}_n$ è *asintoticamente gaussiana* con *media asintotica* $\mu_n(\theta)$ e *varianza asintotica* $\sigma_n^2(\theta)$ se

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

3.4 Diseguaglianza di Fréchet-Cramer-Rao

Abbiamo visto nei capitoli precedenti come per uno stimatore non distorto ridurre il MSE significhi ridurre la varianza dello stimatore. Ora ci chiediamo qual'è la minima varianza che lo stimatore può avere e quale sia lo stimatore con tale varianza.

Siano X_1, \dots, X_n variabili aleatorie iid con comune densità $f(x, \theta)$, $\theta \in \Theta$ 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:

1. Θ è un intervallo aperto in \mathbb{R} ;
2. $S = \{x : f(x, \theta) > 0\}$ è indipendente da θ ;
3. $\theta \mapsto f(x, \theta)$ è derivabile su Θ , $\forall x \in S$;
4. $E_\theta\left(\frac{\partial}{\partial \theta} \log f(X_1, \theta)\right) = 0 \quad \forall \theta \in \Theta$;
5. $0 < E_\theta\left[\left(\frac{\partial}{\partial \theta} \log f(X_1, \theta)\right)^2\right] < \infty \quad \forall \theta \in \Theta$;
6. $\kappa : \Theta \rightarrow \mathbb{R}$ è derivabile su Θ e:

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

Allora

$$Var_{\theta}(T) \geq \frac{(\kappa'(\theta))^2}{nI(\theta)} \quad \forall \theta \in \Theta \quad (3)$$

Dove

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

è nota come informazione di Fisher.

Dimostrazione Per maggiore semplicità 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 iid a media nulla e varianza finita $I(\theta) \forall \theta$. Infatti

$$E_{\theta}(Y_j) = E_{\theta} \left(\frac{\partial}{\partial \theta} \log f(X_j, \theta) \right) = 0 \quad [\text{per l'ipotesi (4)}]$$

e

$$Var_{\theta}(Y_j) = E_{\theta}(Y_j^2) = E_{\theta} \left[\left(\frac{\partial}{\partial \theta} \log f(X_j, \theta) \right)^2 \right] = I(\theta) \in (0, \infty) \quad [\text{per l'ipotesi (5)}]$$

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 ricaviamo che

$$E_{\theta} \left(\frac{\partial}{\partial \theta} \log L_{\theta}(X_1, \dots, X_n) \right) = \sum_{j=1}^n E_{\theta}(Y_j) \quad (4)$$

e

$$Var_{\theta} \left(\frac{\partial}{\partial \theta} \log L_{\theta}(X_1, \dots, X_n) \right) = \sum_{j=1}^n Var_{\theta}(Y_j) = nI(\theta) \quad (5)$$

Dall'ipotesi (6) e dall'equazione 4 ricaviamo che

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

cosicché

$$(\kappa'(\theta))^2 = \left[Cov \left(T, \frac{\partial}{\partial \theta} \log L_{\theta}(X_1, \dots, X_n) \right) \right]^2$$

da cui per le proprietà della covarianza

$$\begin{aligned} &\leq Var_{\theta}(T) Var_{\theta} \left(\frac{\partial}{\partial \theta} \log L_{\theta}(X_1, \dots, X_n) \right) \\ &= Var_{\theta}(T) nI(\theta) \quad [\text{per l'equazione 5}] \end{aligned}$$

4 Verifica di ipotesi

4.1 Definizioni

Un'ipotesi statistica è un'asserzione o una congettura sulla f.d.r. F . Se F , e quindi la corrispondente 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 . Un'ipotesi è *semplice* se specifica completamente la f.d.r. F della popolazione; è *composta* se non è semplice.

In questo capitolo ci occuperemo di ipotesi che riguardano solo il parametro incognito. Nei problemi che affronteremo sono presenti due ipotesi chiamate H_0 e H_1 una è detta *ipotesi nulla* mentre l'altra è detta *ipotesi alternativa*. Il vero valore che il parametro θ assume in natura è compatibile solo o con H_0 o con H_1 ma non con entrambi.

Dobbiamo ora stabilire una regola che ci permetta di decidere tra le due ipotesi. Decidiamo di partizionare l'insieme \mathbb{R}^n di tutte le realizzazioni campionarie X_1, \dots, X_n in due regioni \mathcal{G} e \mathcal{G}^c ed effettuiamo il campionamento; se $(x_1, \dots, x_n) \in \mathcal{G}$ rifiutiamo H_0 se invece $(x_1, \dots, x_n) \in \mathcal{G}^c$ allora accettiamo H_0 . La \mathcal{G} è detta regione di *rifiuto* o regione *critica* mentre \mathcal{G}^c è detta regione di *accettazione*.

Può capitare di prendere delle decisioni sbagliate; gli errori che si possono commettere sono 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.

Questi due errori non possono essere calcolati con precisione ma possiamo calcolare la loro probabilità. Sia $\alpha(\theta)$ la probabilità di errore del I tipo e sia $\beta(\theta)$ la probabilità di errore di II tipo allora:

$$\alpha(\theta) = P_\theta(\text{"Rifiutare } H_0\text{"}) = P_\theta((x_1, \dots, x_n) \in \mathcal{G}), \quad \theta \in \Theta_0 \quad (6)$$

$$\beta(\theta) = P_\theta(\text{"Accettare } H_0\text{"}) = P_\theta((x_1, \dots, x_n) \in \mathcal{G}^c), \quad \theta \in \Theta_1 \quad (7)$$

4.2 Lemma di Neyman-Pearson

Un buon test è tale per cui α e β sono trascurabili; ma in realtà esiste un trade-off cosicché è impossibile minimizzare entrambi. Perciò si procede a fissare un valore per l'errore che si ritiene più grave (quello di I tipo) e poi si minimizza il secondo. Un test creato secondo questo criterio è detto *test uniformemente più potente*. Nel caso di ipotesi entrambi semplice il test creato è specificato dal *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} .

Dimostrazione Dobbiamo dimostrare che per qualunque regione critica \mathcal{F} di ampiezza al più pari a quella di \mathcal{G} cioè tale che $P_{\theta_0}(\mathcal{F}) \leq P_{\theta_0}(\mathcal{G})$ vale $P_{\theta_1}(\mathcal{F}) \leq P_{\theta_1}(\mathcal{G})$.

Osserviamo che possiamo rappresentare \mathcal{F} e \mathcal{G} come:

$$\mathcal{F} = (\mathcal{F} \cap \mathcal{G}) \cup (\mathcal{F} \cap \mathcal{G}^c), \quad \mathcal{G} = (\mathcal{F} \cap \mathcal{G}) \cup (\mathcal{F}^c \cap \mathcal{G})$$

allora

$$P_{\theta_i}(\mathcal{F}) \leq P_{\theta_i}(\mathcal{G}) \text{ se solo se } P_{\theta_i}(\mathcal{F} \cap \mathcal{G}^c) \leq P_{\theta_i}(\mathcal{F}^c \cap \mathcal{G}) \text{ per } i = 0, 1. \quad (8)$$

Inoltre

$$L_{\theta_0}(x_1, \dots, x_n) \leq \delta L_{\theta_1}(x_1, \dots, x_n) \quad \forall (x_1, \dots, x_n) \in \mathcal{G}$$

$$L_{\theta_1}(x_1, \dots, x_n) < L_{\theta_0}(x_1, \dots, x_n)/\delta \quad \forall (x_1, \dots, x_n) \in \mathcal{G}^c$$

da cui otteniamo che:

$$P_{\theta_0}(A) \leq \delta P_{\theta_1}(A) \quad \forall A \subset \mathcal{G} \quad (9)$$

$$P_{\theta_1}(B) \leq \delta P_{\theta_0}(B) \quad \forall B \subset \mathcal{G}^c \quad (10)$$

Infatti

$$P_{\theta_1}(B) = \int_B L_{\theta_1}(x_1, \dots, x_n) dx_1 \dots dx_n = \frac{1}{\delta} \int_B L_{\theta_0}(x_1, \dots, x_n) dx_1 \dots dx_n = \frac{P_{\theta_0}(B)}{\delta}$$

Pertanto

$$\begin{aligned} P_{\theta_1}(\mathcal{F} \cap \mathcal{G}^c) &\leq P_{\theta_0}(\mathcal{F} \cap \mathcal{G}^c)/\delta \quad [\text{per la 10}] \\ &\leq P_{\theta_0}(\mathcal{F}^c \cap \mathcal{G})/\delta \quad [\text{per la 8 con } i = 0] \\ &\leq \delta P_{\theta_1}(\mathcal{F} \cap \mathcal{G}^c)/\delta \quad [\text{per la 9}] \\ &= P_{\theta_1}(\mathcal{F} \cap \mathcal{G}^c) \end{aligned}$$