Alma Mater Studiorum · Università di Bologna

SCUOLA DI SCIENZE
Corso di Laurea in Matematica

CONVERGENZA NORMALE DELLA MASSIMA VEROSIMIGLIANZA E IL TEST DEL CHI-QUADRO

Tesi di Laurea in Statistica Matematica

Relatore:
Chiar.mo Prof.
MASSIMO CAMPANINO

Presentata da:
ALESSANDRO MANCINI

VI Sessione
Anno Accademico 2020-2021

A mia MADRE, *mio* PADRE
e mio fratello GIACOMO,
per il loro sostegno incondizionato
anche nei momenti più difficili.

Introduzione

Sviluppato per la prima volta da Karl Pearson nel 1900, il test del Chi-Quadro è uno dei più importanti test statistici, utilizzato principalmente per verificare ipotesi sulla bontà di adattamento e sull'indipendenza di variabili aleatorie discrete finite. Il test ha numerose applicazioni. In un'indagine statistica, il test del Chi-Quadro per la bontà di adattamento permette di verificare se un sottoinsieme della popolazione può essere considerato rappresentativo della popolazione generale. In medicina, il test del Chi-Quadro per l'indipendenza è utilizzato ad esempio per controllare se vi sono legami tra l'utilizzo di un farmaco ed i possibili effetti collaterali.

L'obiettivo di questo elaborato è illustrare le basi del test del Chi-Quadro da un punto di vista matematico attraverso la convergenza normale dello stimatore di massima verosimiglianza per distribuzioni parametrizzate.

Nel capitolo 1 sono esposti i concetti base della statistica matematica necessari alla trattazione successiva, come la verosimiglianza e il test d'ipotesi.

Nel capitolo 2 viene dimostrata la convergenza normale dello stimatore di massima verosimiglianza nel caso di parametri unidimensionali. È anche esposta la generalizzazione di questi risultati al caso multidimensionale.

Il capitolo 3 definisce il rapporto di verosimiglianza di un test d'ipotesi e mostra la convergenza normale del logaritmo del rapporto nel casi unidimensionale e multidimensionale per ipotesi semplici o composite.

Infine nel capitolo 4 è illustrato il legame tra la convergenza normale del rapporto di verosimiglianza e il test del Chi-Quadro attraverso la distribuzione multinomiale.

La tesi è una rielaborazione delle pubblicazioni di *Wiebe R. Pestman* [1] e *Samuel S. Wilks* [2]. In particolare i capitoli 1 e 4 sono tratti da [1] mentre i capitoli 2 e 3 provengono da [2]. È stato consultato anche il libro di testo di *Dennis V. Lindley* [3] per una maggiore comprensione dell'argomento, ma non sono presenti riferimenti diretti all'interno dell'elaborato.

Indice

Capitolo 1

Conoscenze preliminari

In questo capitolo sono riassunti alcuni dei concetti e definizioni fondamentali della statistica, necessari per la trattazione dei capitoli successivi. Sono inoltre enunciati il teorema centrale del limite e la legge forte dei grandi numeri.

Definizione 1.1. Un *campione* è una sequenza di variabili aleatorie $X_1, \ldots, X_n$ statisticamente indipendenti ed identicamente distribuite. La distribuzione di probabilità degli X_i è detta *distribuzione della popolazione*. Una sequenza $x_1, \ldots, x_n$ ottenuta aleatoriamente a partire da $X_1, \ldots, X_n$ è un *esito* del campione.

Definizione 1.2. Sia $X_1, \ldots, X_n$ un campione e $g : \mathbb{R}^n \to \mathbb{R}$ Borel-misurabile. La variabile aleatoria $T := g(X_1, \ldots, X_n)$ è una *statistica* del campione.

Definizione 1.3. Siano un insieme Θ ed una funzione $f : \mathbb{R} \times \Theta \to [0, +\infty)$. Per ogni θ fissato consideriamo la funzione $f(\bullet, \theta)$ che associa $x \mapsto f(x, \theta)$. Se per ogni $\theta \in \Theta$ $f(\bullet, \theta)$ è una densità di probabilità, allora $\{f(\bullet, \theta)\}_{\theta \in \Theta}$ è una *famiglia di densità di probabilità*. Θ è detto *spazio dei parametri*.

Definizione 1.4. Sia $\{f(\bullet, \theta)\}_{\theta \in \Theta}$ una famiglia di densità di probabilità. Consideriamo una popolazione statistica con densità $f(\bullet, \theta)$. Una funzione $\kappa : \Theta \to \mathbb{R}$ è detta *caratteristica* della popolazione.

Definizione 1.5. Siano $X_1, \dots, X_n$ un campione da una popolazione con densità $f(\bullet, \theta)$ con $\theta \in \Theta$ e κ una caratteristica della popolazione. Una statistica $T = g(X_1, \dots, X_n)$ il cui valore è utilizzato come stima di $\kappa(\theta)$ è detta *stimatore di* κ basato sul campione $X_1, \dots, X_n$. Si dice che T è uno stimatore *non deviato* se $\mathbb{E}[T] = \kappa(\theta)$. Se ciò non vale si dice che T è *deviato.*

1.1 Verosimiglianza

Definizione 1.6. Sia $X_1, \dots, X_n$ un campione con densità di probabilità $f(x, \theta)$, $\theta \in \Theta$. La *verosimiglianza* del campione è definita come la densità di probabilità L_θ del vettore $(X_1, \dots, X_n)$.

$$L_\theta : \mathbb{R}^n \to [0, +\infty) \qquad L_\theta(x_1, \dots, x_n) := f(x_1, \theta) \cdots f(x_n, \theta)$$

Supponiamo ora che $\hat{\theta}$ sia l'unico θ che massimizza la funzione $\theta \mapsto L_\theta(x_1, \dots, x_n)$.

Definizione 1.7. Sotto le ipotesi di cui sopra, $\hat{\theta} = \hat{\theta}(x_1, \dots, x_n)$ è detta *stima di massima verosimiglianza* di θ. Sia κ una caratteristica della popolazione. Se $\kappa \circ \hat{\theta}$ è Borel-misurabile, la statistica

$$T := \kappa(\hat{\theta}(X_1, \dots, X_n))$$

è detta *stimatore di massima verosimiglianza* della caratteristica κ.

1.2 Test di ipotesi

Definizione 1.8. Una *ipotesi statistica* è una congettura H riguardante la distribuzione di una popolazione. H è detta *semplice* se descrive esaustivamente la distribuzione. Se ciò non accade, si dice che l'ipotesi è *composita.*

Definizione 1.9. Sia $X_1, \dots, X_n$ un campione con densità $f(\bullet, \theta)$, $\theta \in \Theta$. Siano H_0, H_1 due ipotesi statistiche tali che:

$$H_0 : \theta \in \Theta_0 \quad ; \quad H_1 : \theta \in \Theta_1 \qquad \Theta_1 = \Theta \setminus \Theta_0$$

Sia $G \subset \mathbb{R}^n$ un insieme Borel-misurabile. La sequenza ordinata $(X_1, \ldots, X_n, H_0, H_1, G)$ è detta *test di ipotesi.* H_0 è l'*ipotesi nulla*, H_1 è l'*ipotesi alternativa*, G è la *zona critica* del test.

Per svolgere un test di ipotesi prendiamo in considerazione l'esito $(x_1, \ldots, x_n)$ del campione. Se $(x_1, \ldots, x_n) \in G$ accetteremo l'ipotesi H_1. In caso contrario, H_0 sarà accettata.

Per via della natura aleatoria del campione i test di ipotesi non sono infallibili ed è possibile compiere due tipi di errore:

Definizione 1.10. Si ha un *errore del I tipo* se l'ipotesi H_1 è accettata ma vale H_0. La probabilità α di compiere un errore del I tipo è detta *significatività* del test. Si ha invece un *errore del II tipo* se H_0 è accettata ma vale H_1. La probabilità di compiere un errore del II è indicata con β. Il valore $1 - \beta$ è detto *potenza* del test.

1.3 Teorema centrale del limite

Definizione 1.11. Sia $(X_n)_{n \in \mathbb{N}}$ una successione di variabili aleatorie. $(X_n)_{n \in \mathbb{N}}$ di dice *convergente in distribuzione* alla variabile X se

$$\lim_{n \to +\infty} F_{X_n}(x) = F_X(x)$$

per ogni punto x in cui F_X è continua.

Teorema 1.12 (Teorema centrale del limite)**.** *Siano $X_1, \ldots, X_n$ variabili aleatorie statisticamente indipendenti ed identicamente distribuite con media μ e varianza $\sigma^2 \neq 0$. Sia $\overline{X}$ la media campionaria. La variabile aleatoria*

$$\frac{\overline{X} - \mu}{\sigma / \sqrt{n}}$$

converge in distribuzione alla normale standard.

Dimostrazione. Vedi [1] **I.9.3** □

Teorema 1.13 (Legge forte dei grandi numeri)**.** *Siano* $X_1, \ldots, X_n$ *variabili aleatorie statisticamente indipendenti ed identicamente distribuite con media* μ*. La media campionaria* $\overline{X}$ *converge quasi certamente a* μ *per* $n \to \infty$*.*

Dimostrazione. Vedi [1] **VII.2.14** $\square$

Capitolo 2

Convergenza normale della massima verosimiglianza

I risultati di questo capitolo e del seguente sono validi per qualsiasi tipo di funzione di distribuzione. Adottiamo per semplicità la notazione del caso assolutamente continuo, ma tutti i teoremi e le dimostrazioni hanno valenza generale. Tutte le definizioni e dimostrazioni presuppongono un parametro θ unidimensionale. Nell'ultima sezione è fornita una generalizzazione dei risultati al caso r-dimensionale. Si fa riferimento a [2], capitoli **12.1**, **12.3** e **12.7**.

Definizione 2.1. Sia $F(x, \theta) : \mathbb{R} \to [0, 1]$ una funzione di distribuzione, con $\theta \in \Theta$ parametro unidimensionale. Indichiamo con $f(x, \theta)$ quella funzione tale che

$$\int_{-\infty}^{x} f(y, \theta) dy = F(x, \theta)$$

Sia X una variabile aleatoria con distribuzione data da $F(x, \theta)$, definiamo le quantità:

$$\begin{aligned}
S(x,\theta) =& \frac{\partial}{\partial\theta}\log f(x,\theta) \\
S'(x,\theta) =& \frac{\partial}{\partial\theta}S(x,\theta) \\
A(\theta,\theta') =& \mathbb{E}_\theta(S(X,\theta')) \\
B^2(\theta,\theta') =& \mathbb{E}_\theta(S(X,\theta')^2)
\end{aligned}$$

Definizione 2.2. Una funzione di distribuzione $F(x,\theta)$, $\theta \in \Theta$ si dice *regolare rispetto alla derivata prima* in Θ se

$$\forall\theta\in\Theta, \quad \mathbb{E}(S(X,\theta)) = 0$$

con X variabile aleatoria distribuita secondo $F(x,\theta)$.

Definizione 2.3. Una funzione di distribuzione $F(x,\theta)$, $\theta \in \Theta$ si dice *regolare rispetto alla derivata seconda* in Θ se

$$\forall\theta\in\Theta, \quad \mathbb{E}(S'(X,\theta)) + \mathbb{E}(S(X,\theta)^2) = 0$$

con X variabile aleatoria distribuita secondo $F(x,\theta)$.

2.1 Convergenza normale del punteggio

Definizione 2.4. Sia $X_1,\dots,X_n$ un campione di variabili aleatorie indipendenti ed identicamente distribuite a valori reali con funzione di distribuzione $F(x,\theta_0)$, $\theta_0 \in \Theta$.
Definisco il *punteggio* del campione

$$S_n(X_1,\dots,X_n,\theta) = \frac{\partial}{\partial\theta}\log L_\theta(X_1,\dots,X_n)$$

Da ciò segue banalmente la relazione

$$S_n(X_1,\dots,X_n,\theta) = \sum_{i=1}^{n} S(X_i,\theta)$$

Teorema 2.5. *Sia $X_1, \dots, X_n$ un campione di variabili aleatorie indipendenti ed identicamente distribuite a valori reali con funzione di distribuzione $F(x, \theta_0)$, $\theta_0 \in \Theta$. Supponiamo esista e sia finita la varianza di $S(X, \theta_0)$. Se $F(x, \theta)$ è regolare rispetto alla derivata prima in Θ, allora la variabile aleatoria*

$$\frac{S_n(X_1, \dots, X_n, \theta_0)}{\sqrt{n} B(\theta_0, \theta_0)}$$

converge in distribuzione alla normale standard.

Dimostrazione.

$$\begin{aligned} S_n(x_1, \dots, x_n, \theta_0) =& \frac{\partial}{\partial \theta} \log f_n(x_1, \dots, x_n, \theta_0) = \frac{\partial}{\partial \theta} \log \prod_{i=1}^{n} f(x_i, \theta_0) = \\ =& \sum_{i=1}^{n} \frac{\partial}{\partial \theta} \log f(x_i, \theta_0) = \sum_{i=1}^{n} S(x_i, \theta_0) \end{aligned}$$

$S_n(X_1, \dots, X_n)$ è dunque somma di n variabili aleatorie indipendenti ed identicamente distribuite $S(X_i, \theta_0)$. Queste hanno aspettativa $\mathbb{E}(S(X_i, \theta_0)) = 0$ per regolarità, vediamo ora la varianza

$$\mathrm{var}(S(X_i, \theta_0)) = \mathbb{E}(S(X_i, \theta_0)^2) - \mathbb{E}(S(X_i, \theta_0))^2 = B^2(\theta_0, \theta_0)$$

Il teorema centrale del limite (1.12) conclude. □

Teorema 2.6. *Sia $X_1, \dots, X_n$ un campione di variabili aleatorie indipendenti ed identicamente distribuite a valori reali con funzione di distribuzione $F(x, \theta_0)$, $\theta_0 \in \Theta$ con $F(x, \theta)$ regolare rispetto alla derivata prima in Θ. Supponiamo $S(x, \bullet)$ continua in Θ $\forall x \in \Omega$ eccetto per un insieme di probabilità nulla. Allora esiste una successione $(\hat{\theta}_n)_{n \in \mathbb{N}}$ con $\hat{\theta}_n = \hat{\theta}_n(X_1, \dots, X_n)$ soluzione di $S_n(X_1, \dots, X_n, \theta) = 0$, $(\hat{\theta}_n)$ convergente quasi certamente a θ_0.*

Dimostrazione. Il campione ha distribuzione data da $F(x, \theta_0)$. Al variare di $\theta \in \Theta$ la quantità

$$\frac{1}{n} S_n(X_1, \dots, X_n, \theta)$$

converge quasi certamente a $A(\theta_0, \theta) = \mathbb{E}_{\theta_0}(S(X_i, \theta))$ per la legge forte dei grandi numeri (1.13). Per regolarità, si ha $A(\theta_0, \theta_0) = 0$.

Siccome θ_0 rappresenta un punto di massimo locale per il logaritmo della verosimiglianza, $\exists \delta > 0$ tale che $A(\theta_0, \bullet)$ è decrescente in $[\theta_0 - \delta, \theta_0 + \delta]$. Ne segue che $\forall \epsilon > 0$ $\exists n(\delta, \epsilon)$ tale che $\forall n > n(\delta, \epsilon)$ le seguenti disuguaglianze si realizzano con probabilità superiore a $1 - \epsilon$

$$\frac{1}{n} S_n(X_1, \ldots, X_n, \theta_0 - \delta) > 0$$

$$\frac{1}{n} S_n(X_1, \ldots, X_n, \theta_0 + \delta) < 0$$

Siccome $S(x, \bullet)$ è continua si ha che la proprietà

$$\forall n > n(\delta, \epsilon)\ \exists \theta \in]\theta_0 - \delta, \theta_0 + \delta[\text{ tale che } \frac{1}{n} S_n(X_1, \ldots, X_n, \theta) = 0$$

si verifica con probabiltà superiore a $1-\epsilon$ se θ_0 è il valore reale di θ. Ciò è equivalente all'esistenza di una successione di soluzioni di $S_n(X_1, \ldots, X_n, \theta) = 0$ convergenti a θ_0. $\square$

È importante osservare come se $\forall n > n_0$, $n_0 \in \mathbb{N}$ la soluzione $\hat{\theta}$ di $S_n(X_1, \ldots, X_n, \theta) = 0$ è unica, allora il massimo è assoluto e $\hat{\theta}$ è lo stimatore di massima verosimiglianza di θ_0.

2.2 Convergenza normale della massima verosimiglianza

Teorema 2.7. *Sia $X_1, \ldots, X_n$ campione di variabili aleatorie indipendenti ed identicamente distribuite a valori reali con funzione di distribuzione $F(x, \theta_0)$, $\theta_0 \in \Theta$ unidimensionale, $F(x, \theta_0)$ regolare rispetto alle prime due derivate in Θ. Supposiamo esista unico $\hat{\theta} = \hat{\theta}(X_1, \ldots, X_n)$ stimatore di massima verosimiglianza di θ per $n > n_0$. La variabile aleatoria*

$$\sqrt{n} B(\theta_0, \theta_0)(\hat{\theta} - \theta_0)$$

converge in distribuzione alla normale standard.

Dimostrazione. $\hat{\theta}$ converge quasi certamente a θ_0, dunque $\forall(\delta, \epsilon)\ \exists n(\delta, \epsilon) > n_0$ tale che

$$\mathbb{P}_{\theta_0}((x_1, \ldots, x_n) \in E_n\ \forall n > n(\delta, \epsilon)) > 1 - \epsilon$$

con $E_n = \{(x_1, \ldots, x_n) \| \ |\theta_0 - \hat{\theta}| < \delta\}$. All'interno di E_n posso sviluppare:

$$\frac{1}{\sqrt{n}} S_n(x_1, \ldots, x_n, \theta_0) = \frac{1}{\sqrt{n}} S_n(x_1, \ldots, x_n, \hat{\theta}) + \frac{1}{\sqrt{n}} S'_n(x_1, \ldots, x_n, \overline{\theta})(\theta_0 - \hat{\theta})$$

con $\overline{\theta}(X_1, \ldots, X_n)$ tale che $|\overline{\theta} - \theta_0| < |\hat{\theta} - \theta_0|$.
Siccome $S_n(x_1, \ldots, x_n, \hat{\theta}) = 0$, l'uguaglianza diventa

$$\frac{1}{\sqrt{n}} S_n(x_1, \ldots, x_n, \theta_0) = \frac{1}{\sqrt{n}} S'_n(x_1, \ldots, x_n, \overline{\theta})(\theta_0 - \hat{\theta})$$

La parte sinistra dell'equazione converge in distribuzione a una normale con media 0 e varianza $B^2(\theta_0, \theta_0)$. Ciò ci permette di affermare che

$$\sqrt{n} \frac{1}{n} \frac{S'_n(x_1, \ldots, x_n, \overline{\theta})}{B(\theta_0, \theta_0)} (\theta_0 - \hat{\theta})$$

converge in distribuzione a una normale standard. Inoltre per regolarità rispetto alla derivata seconda si ha $\hat{\theta} \to \theta_0 \Rightarrow \overline{\theta} \to \theta_0$ da cui

$$\frac{1}{n} S'_n(x_1, \ldots, x_n, \overline{\theta}) \to \mathbb{E}(S'(X_i, \theta)) = -\mathbb{E}(S(X_i, \theta)^2) = -B^2(\theta_0, \theta_0)$$

Possiamo dunque concludere che anche $\sqrt{n} B(\theta_0, \theta_0)(\hat{\theta} - \theta_0)$ converge in distribuzione alla normale standard. □

2.3 Generalizzazione per parametri multidimensionali

In questa sezione riportiamo senza dimostrazione le generalizzazioni dei risultati precedenti nel caso in cui il parametro θ sia r-dimensionale.

Definizione 2.8. Sia $F(x, \theta) : \mathbb{R} \to [0, 1]$ una funzione di distribuzione, con $\theta = (\theta_1, \dots, \theta_r) \in \Theta$ parametro r-dimensionale. Indichiamo con $f(x, \theta)$ quella funzione tale che

$$\int_{-\infty}^{x} f(y, \theta) dy = F(x, \theta)$$

Sia X una variabile aleatoria con distribuzione data da $F(x, \theta)$, definiamo le quantità

$$\begin{aligned} S_p(x, \theta) =& \frac{\partial}{\partial \theta_p} \log f(x, \theta) \\ S_{p,q}(x, \theta) =& \frac{\partial^2}{\partial \theta_p \partial \theta_q} \log f(x, \theta) \\ B^2_{p,q}(\theta, \theta') =& \mathbb{E}_\theta(S_p(X, \theta') S_q(X, \theta')) \end{aligned}$$

Definizione 2.9. Una funzione di distribuzione $F(x, \theta)$, $\theta \in \Theta$ si dice *regolare rispetto alla derivata prima* in Θ se

$$\forall p \; \forall \theta \in \Theta, \quad \mathbb{E}(S_p(X, \theta)) = 0$$

con X variabile aleatoria distribuita secondo $F(x, \theta)$.

Definizione 2.10. Una funzione di distribuzione $F(x, \theta)$, $\theta \in \Theta$ si dice *regolare rispetto alla derivata seconda* in Θ se

$$\forall p, q \; \forall \theta \in \Theta, \quad \mathbb{E}(S_{p,q}(X, \theta)) + \mathbb{E}(S_p(X, \theta)(S_q(X, \theta)) = 0$$

con X variabile aleatoria distribuita secondo $F(x, \theta)$.

Definizione 2.11. Sia $X_1, \dots, X_n$ un campione di variabili aleatorie indipendenti ed identicamente distribuite a valori reali con funzione di distribuzione $F(x, \theta)$, $\theta \in \Theta$. Definisco il *punteggio* del campione secondo θ_p come

$$S_p^{(n)}(X_1, \dots, X_n, \theta) = \frac{\partial}{\partial \theta_p} \log L_\theta(X_1, \dots, X_n)$$

Da ciò segue banalmente la relazione

$$S_p^{(n)}(X_1, \dots, X_n, \theta) = \sum_{i=1}^{n} S_p(X_i, \theta)$$

Definisco il *punteggio* del campione come il vettore

$$S^{(n)}(X_1, \dots, X_n, \theta) = (S_1^{(n)}, \dots, S_r^{(n)})$$

Teorema 2.12. *Sia $X_1, \ldots, X_n$ un campione di variabili aleatorie indipendenti ed identicamente distribuite a valori reali con funzione di distribuzione $F(x, \theta_0)$, $\theta_0 = (\theta_{0,1}, \ldots, \theta_{0,r}) \in \Theta$. Se $F(x, \theta)$ è regolare rispetto alla derivata prima in Θ, allora la variabile aleatoria*

$$\frac{1}{\sqrt{n}} S^{(n)}(X_1, \ldots, X_n, \theta_0)$$

converge in distribuzione a una normale r-dimensionale con media 0 e matrice di covarianza

$$B(\theta_0, \theta_0) = \begin{bmatrix} B^2_{1,1}(\theta_0, \theta_0) & \ldots & B^2_{1,r}(\theta_0, \theta_0) \\ \vdots & \ddots & \vdots \\ B^2_{r,1}(\theta_0, \theta_0) & \ldots & B^2_{r,r}(\theta_0, \theta_0) \end{bmatrix}$$

Teorema 2.13. *Sia $X_1, \ldots, X_n$ un campione di variabili aleatorie indipendenti ed identicamente distribuite a valori reali con funzione di distribuzione $F(x, \theta_0)$, $\theta_0 = (\theta_{0,1}, \ldots, \theta_{0,r}) \in \Theta$ con $F(x, \theta)$ regolare rispetto alla derivata prima in Θ. Supponiamo che $\forall p$ $S_p(x, \bullet)$ continua in Θ $\forall x \in \Omega$ eccetto per un insieme di probabilità nulla. Allora esiste una successione $(\hat{\theta}^{(n)})_{n \in \mathbb{N}}$ con $\hat{\theta}^{(n)} = \hat{\theta}^{(n)}(X_1, \ldots, X_n)$ soluzione del sistema di equazioni $S^{(n)}(X_1, \ldots, X_n, \theta) = 0$, $(\hat{\theta}^{(n)})$ convergente quasi certamente a θ_0.*

Si osservi anche in questo caso come se $\forall n > n_0$ $\hat{\theta}^{(n)}$ è unico, allora è lo stimatore di massima verosimiglianza.

Teorema 2.14. *Sia $X_1, \ldots, X_n$ campione di variabili aleatorie indipendenti ed identicamente distribuite a valori reali con funzione di distribuzione $F(x, \theta_0)$, $\theta_0 \in \Theta$ r-dimensionale, $F(x, \theta_0)$ regolare rispetto alle prime due derivate in Θ. Supposiamo esista unico $\hat{\theta} = \hat{\theta}(X_1, \ldots, X_n)$ stimatore di massima verosimiglianza di θ_0 per $n > n_0$. La variabile aleatoria*

$$\sqrt{n}(\hat{\theta} - \theta_0)$$

converge in distribuzione a una normale r-dimensionale con media 0 e matrice di covarianza B^{-1}, con $B = B(\theta_0, \theta_0)$ definita dal teorema 2.12.

Capitolo 3

Convergenza normale del rapporto di verosimiglianza

In questo capitolo si affronta il comportamento asintotico del logaritmo del rapporto di verosimiglianza quando n tende a infinito. I risultati sono tratti da [2], capitoli **13.4**, **13.7** e **13.8**. Sia $X_1, \ldots, X_n$ campione di variabili aleatorie indipendenti ed identicamente distribuite a valori reali con funzione di distribuzione $F(x, \theta)$, $\theta \in \Theta$ e consideriamo le ipotesi $H_0 : \theta \in \Theta_0$ ed $H_1 : \theta \in \Theta \setminus \Theta_0$. Definiamo il *rapporto di verosimiglianza* come la statistica Λ del campione:

$$\Lambda(X_1, \ldots, X_n) = \frac{\sup_{\theta \in \Theta_0} L_\theta(X_1, \ldots, X_n)}{\sup_{\theta \in \Theta} L_\theta(X_1, \ldots, X_n)}$$

Se $\hat{\theta}$ e $\hat{\theta}'$ sono gli stimatori di massima verosimiglianza di θ in Θ e Θ_0 rispettivamente, possiamo esprimere il rapporto di verosimiglianza come

$$\Lambda(X_1, \ldots, X_n) = \frac{L_{\hat{\theta}'}(X_1, \ldots, X_n)}{L_{\hat{\theta}}(X_1, \ldots, X_n)} = \prod_{i=1}^{n} \frac{f(X_i, \hat{\theta}')}{f(X_i, \hat{\theta})}$$

Di particolare interesse è soprattutto il logaritmo del rapporto

$$\log \Lambda(X_1, \ldots, X_n) = \sum_{i=1}^{n} \left(\log f(X_i, \hat{\theta}') - \log f(X_i, \hat{\theta}) \right)$$

3.1 Convergenza nel caso unidimensionale

Teorema 3.1. *Sia $X_1, \dots, X_n$ campione di variabili aleatorie indipendenti ed identicamente distribuite a valori reali con funzione di distribuzione $F(x, \theta)$, $\theta \in \Theta$ unidimensionale, $F(x, \theta)$ regolare rispetto alle prime due derivate. Supposiamo esista unico $\hat{\theta}$ stimatore di massima verosimiglianza di θ per $n > n_0$. Sia $H_0 : \theta = \theta_0$ ipotesi semplice e sia Λ il rapporto di verosimiglianza del test. Se H_0 è verificata, allora $-2 \log \Lambda$ converge in distribuzione a una $\chi^2(1)$.*

Dimostrazione. Se H_0 è verificata, allora $\hat{\theta}$ è lo stimatore di massima verosimiglianza di θ_0 e valgono le conclusioni del teorema 2.7, per il quale $\sqrt{n}B(\theta_0, \theta_0)(\theta_0 - \hat{\theta})$ converge in distribuzione a una normale standard. Ne segue che $nB^2(\theta_0, \theta_0)(\theta_0 - \hat{\theta})^2$ converge a una $\chi^2(1)$. Consideriamo ora il logaritmo della verosimiglianza in θ_0.

$$\log L_{\theta_0}(X_1, \dots, X_n) = \sum_{i=1}^{n} \log f(X_i, \theta_0)$$

$\hat{\theta}$ converge a θ_0 quasi certamente per $n \to \infty$. Utilizzando lo sviluppo in serie di Taylor, possiamo scrivere

$$\begin{aligned}\sum_{i=1}^{n} \log f(X_i, \theta_0) =& \sum_{i=1}^{n} \log f(X_i, \hat{\theta}) + \sum_{i=1}^{n} \left[\frac{\partial}{\partial\theta} \log f(X_i, \hat{\theta})\right] (\theta_0 - \hat{\theta}) + \\ &+ \frac{1}{2} \sum_{i=1}^{n} \left[\frac{\partial^2}{\partial\theta^2} \log f(X_i, \theta)\right]_{\theta=\overline{\theta}} (\theta_0 - \hat{\theta})^2\end{aligned}$$

con $\overline{\theta}$ tale che $|\theta - \overline{\theta}| < |\theta - \hat{\theta}|$.
$\hat{\theta}$ è soluzione dell'equazione $S_n(X_1, \dots, X_n, \theta) = 0$, perciò

$$\sum_{i=1}^{n} \left[\frac{\partial}{\partial\theta} \log f(X_i, \hat{\theta})\right] (\theta_0 - \hat{\theta}) = S_n(X_1, \dots, X_n, \hat{\theta})(\theta_0 - \hat{\theta}) = 0$$

Si può ottenere dunque l'uguaglianza

$$-2 \sum_{i=1}^{n} \log \frac{f(X_i, \theta_0)}{f(X_i, \hat{\theta})} = -\frac{1}{n} \sum_{i=1}^{n} \left[\frac{\partial^2}{\partial\theta^2} \log f(X_i, \theta)\right]_{\theta=\overline{\theta}} [\sqrt{n}(\theta_0 - \hat{\theta})]^2$$

La parte sinistra dell'uguaglianza è $-2\log\Lambda$. Analizziamo ora la parte destra.

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2}{\partial\theta^2}\log f(X_i,\theta)$$

è una media campionaria e converge in distribuzione all'aspettativa

$$\mathbb{E}\left(\frac{\partial^2}{\partial\theta^2}\log f(X_i,\theta)\right)=\mathbb{E}(S'(X_i,\theta))$$

Per regolarità rispetto alla seconda derivata

$$\mathbb{E}(S'(X_i,\theta))=-\mathbb{E}(S(X_i,\theta)^2)=-B^2(\theta,\theta)$$

$\hat{\theta}\to\theta_0\Rightarrow\overline{\theta}\to\theta_0$, dunque

$$-\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\partial^2}{\partial\theta^2}\log f(X_i,\theta)\right]_{\theta=\overline{\theta}}\longrightarrow B^2(\theta_0,\theta_0)$$

La parte destra converge perciò a una variabile con la stessa distribuzione di $nB^2(\theta_0,\theta_0)(\theta_0-\hat{\theta})^2$, cioè una $\chi^2(1)$. Ciò ci permette di affermare che anche la parte sinistra, cioè $-2\log\Lambda$ converge a una distribuzione $\chi^2(1)$.

□

3.2 Convergenza nel caso r-dimensionale

Proposizione 3.2. *Sia* $Y=(Y_1,\dots,Y_r)$ *una variabile aleatoria r-dimensionale distribuita normalmente con media* 0 *e matrice di covarianza* Σ. *Allora la variabile aleatoria* $<\Sigma^{-1}Y,Y>$ *ha distribuzione* $\chi^2(r)$.

Dimostrazione. Σ è simmetrica definita positiva, esistono perciò Q ortogonale e Λ diagonale tali che $\Sigma=Q^T\Lambda Q$. Se poniamo $Z=QY$ e $\mathrm{diag}(\Lambda)=(\lambda_1^2,\dots,\lambda_r^2)$ si ha:

$$<\Sigma^{-1}Y,Y>=<Q^T\Lambda^{-1}QY,Y>=<\Lambda^{-1}QY,QY>=$$
$$=<\Lambda^{-1}Z,Z>=\sum_{k=1}^{r}\lambda_k^{-2}Z_k^2$$

Per ottenere la tesi è mostrare che le variabili $\lambda_k^{-1} Z_k$ sono normali standard fra loro indipendenti.

$$\mathbb{E}(Z) = \mathbb{E}(QY) = Q\mathbb{E}(Y) = 0 \Rightarrow \mathbb{E}(\lambda_k^{-1} Z_k) = 0 \ \forall k$$

$$\begin{aligned} \operatorname{cov}(Z_k, Z_h) =& \operatorname{cov}(\sum_i Q_{k,i} Y_i, \sum_j Q_{h,j} Y_j) = \sum_{i,j} Q_{k,i} \operatorname{cov}(Y_i, Y_j) Q_{h,j} = \\ =& \sum_{i,j} Q_{k,i} \Sigma_{i,j} Q^T_{j,h} = (Q\Sigma Q^T)_{k,h} = \Lambda_{k,h} \\ & \operatorname{cov}(\lambda_k^{-1} Z_k, \lambda_h^{-1} Z_h) = \lambda_k^{-1}\lambda_h^{-1}\Lambda_{k,h} = \delta_{k,h} \end{aligned}$$

□

Teorema 3.3. *Sia $X_1, \dots, X_n$ campione di variabili aleatorie indipendenti ed identicamente distribuite a valori reali con funzione di distribuzione $F(x, \theta)$, $\theta \in \Theta$ r-dimensionale, $F(x, \theta)$ regolare rispetto alle prime due derivate. Supposiamo esista unico $\hat{\theta}$ stimatore di massima verosimiglianza di θ per $n > n_0$. Sia $H_0 : \theta = \theta_0$ ipotesi semplice e sia Λ il rapporto di verosimiglianza del test. Se H_0 è verificata, allora $-2 \log \Lambda$ converge in distribuzione a una $\chi^2(r)$.*

Dimostrazione. Se vale H_0 si ha che $\hat{\theta}$ converge quasi certamente a θ_0, dunque $\sqrt{n}(\theta_0 - \hat{\theta})$ è una normale r-dimensionale con media 0 e matrice di covarianza B^{-1}. Ne segue che $< B\sqrt{n}(\theta_0 - \hat{\theta}), \sqrt{n}(\theta_0 - \hat{\theta}) >$ converge in distribuzione a una $\chi^2(r)$ per la proposizione 3.2. Essendo $\hat{\theta}$ soluzione del sistema $S^{(n)}(X_1, \dots, X_n, \theta)$, attraverso lo sviluppo di Taylor di $\log L_{\theta_0}$ si può ottenere, analogamente a come dimostrato per il teorema 3.1:

$$-2 \log \Lambda(X_1, \dots, X_n) =< \left(-\frac{1}{n} \sum_i H(X_i, \overline{\theta}) \right) \sqrt{n}(\theta_0 - \hat{\theta}), \sqrt{n}(\theta_0 - \hat{\theta}) >$$

con $H(X_i, \bullet)$ matrice hessiana di $\log f(X_i, \bullet)$ e $\overline{\theta}$ sulla congiungente di $\hat{\theta}$ e θ_0. Considiamo la parte destra dell'equazione. Per il teorema centrale del limite

$$\left(\frac{1}{n} \sum_i H(X_i, \overline{\theta}) \right) \longrightarrow D$$

con $D_{p,q} = \mathbb{E}(\frac{\partial^2}{\partial\theta_p \partial\theta_q} \log f(X, \theta_0))$. Per regolarità si ha $D = -B$. Dunque la parte destra dell'equazione converge alla stessa distribuzione di $< B\sqrt{n}(\theta_0 - \hat{\theta}), \sqrt{n}(\theta_0 - \hat{\theta}) >$, cioè una $\chi^2(r)$. Lo stesso vale dunque per $-2 \log \Lambda$. □

3.3 Generalizzazione ad ipotesi composite

É possibile generalizzare ulteriormente i teoremi 3.1 e 3.3 al caso di un'ipotesi composita. A questo proposito, si presenta il seguente teorema senza dimostrazione.

Teorema 3.4. *Sia $X_1, \dots, X_n$ campione di variabili aleatorie indipendenti ed identicamente distribuite a valori reali con funzione di distribuzione $F(x, \theta)$, $\theta \in \Theta$ r-dimensionale, $F(x, \theta)$ regolare rispetto alle prime due derivate. Sia $H_0 : \theta \in \Theta_0$ ipotesi composita, $dim\Theta_0 = r'$. Supposiamo che per $n > n_0$ esistano unici $\hat{\theta}$ stimatore di massima verosimiglianza di θ in Θ e $\hat{\theta}'$ stimatore di massima verosimiglianza di θ in Θ_0. Sia Λ il rapporto di verosimiglianza del test. Se H_0 è verificata, allora $-2 \log \Lambda$ converge in distribuzione a una $\chi^2(r - r')$*

Capitolo 4

Test del Chi-Quadro

Questo capitolo tratta il *test del Chi-Quadro*, anche noto come *test di Pearson*, una delle più comuni applicazioni dei teoremi 3.3 e 3.4. Il test utilizza la convergenza normale del logaritmo del rapporto di verosimiglianza per verificare ipotesi su variabili aleatorie discrete. A questo proposito viene effettuata una approssimazione del rapporto di verosimiglianza attraverso l'uso della distribuzione multinomiale. Nella prima sezione definiremo la distribuzione multinomiale. Nelle due sezioni successive definiremo infine due casi notevoli del test del Chi-Quadro riguardanti ipotesi sulla *bontà di adattamento* e la *indipendenza di variabili.*

4.1 Distribuzione multinomiale

In questa sezione è definita la distribuzione multinomiale con le sue principali proprietà.

Definizione 4.1. Siano $p, n \in \mathbb{N}$, $\Omega = \{\mathbf{k} = (k_1, \ldots, k_p) \in \mathbb{N}^p \| \sum_{i=1}^p k_i = n\}$ e $\theta \in \mathbb{R}^p$ tale che $\sum_{i=1}^p \theta_i = 1$ e $\theta_i \geq 0\ \forall i$ La *distribuzione multinomiale di parametro scalare n e parametro vettoriale θ* su Ω è la distribuzione tale che:

$$\mathcal{M}_{n,\theta}(\{\mathbf{k}\}) = \binom{n}{\mathbf{k}} \prod_{i=1}^{p} \theta_i^{k_i}$$

Definizione 4.2. Siano $X_1, \dots, X_n$ variabili aleatorie indipendenti ed identicamente distribuite a valori in $\{1, \dots, p\}$. Definisco $F_i := \#\{j \in \{1, \dots, n\} \| X_j = i\}$ La variabile aleatoria $F = (F_1, \dots, F_p)$ è il *vettore delle frequenze* di $X_1, \dots, X_n$

Proposizione 4.3. *Siano $X_1, \dots, X_n$ variabili aleatorie indipendenti ed identicamente distribuite a valori in $\{1, \dots, p\}$ tali che $\mathbb{P}(X_j = i) = \theta_i \ \forall i, j$. Il vettore F delle frequenze di $X_1, \dots, X_n$ ha distribuzione multinomiale di parametro scalare n e parametro vettoriale θ.*

Dimostrazione. Sia $\mathbf{k} = (k_1, \dots, k_p)$ un esito di F. L'esito k si realizza ogni volta che l'esito $\mathbf{x} = (x_1, \dots, x_n)$ di $\mathbf{X} = (X_1, \dots, X_n)$ è tale per cui $\#\{j \| x_j = i\} = k_i \ \forall i$. Sia $\Omega_\mathbf{k}$ l'insieme di tali esiti. Osserviamo che $\#\Omega_\mathbf{k} = \binom{n}{\mathbf{k}}$.
Si può affermare che:

$$\mathbb{P}(F = \mathbf{k}) = \sum_{\mathbf{x} \in \Omega_\mathbf{k}} \mathbb{P}(\mathbf{X} = \mathbf{x})$$

Sia ora $\mathbf{x}$ un esito di $\mathbf{X}$ e $\mathbf{k}$ il corrispondente esito del vettore delle frequenze.

$$\mathbb{P}(\mathbf{X} = \mathbf{x}) = \prod_{j=1}^{n} \mathbb{P}(X_j = x_j) = \prod_{j=1}^{n} \theta_{x_j} = \prod_{i=1}^{p} \theta_i^{k_i}$$

Concludiamo pertanto che

$$\mathbb{P}(F = \mathbf{k}) = \sum_{\mathbf{x} \in \Omega_\mathbf{k}} \mathbb{P}(\mathbf{X} = \mathbf{x}) = \#\Omega_\mathbf{k} \prod_{i=1}^{p} \theta_i^{k_i} = \binom{n}{\mathbf{k}} \prod_{i=1}^{p} \theta_i^{k_i}$$

Ciò dimostra che F ha dimostrazione multinomiale con parametri n, θ. □

Corollario 4.4. *Siano $X_1, \dots, X_n$ variabili aleatorie indipendenti ed identicamente distribuite a valori in $\{1, \dots, p\}$ tali che $\mathbb{P}(X_j = i) = \theta_i \ \forall i, j$. Allora per ogni $\boldsymbol{x}$ esito di $\boldsymbol{X}$, dato il relativo esito $\boldsymbol{k}$ del vettore delle frequenze F, si ha:*

$$L_\theta(x_1, \dots, x_n) = \binom{n}{\boldsymbol{k}}^{-1} \mathbb{P}_\theta(F = \boldsymbol{k})$$

Dimostrazione.

$$L_\theta(x_1, \ldots, x_n) = \prod_{j=1}^{n} \mathbb{P}(X_j = x_j)$$

Nella dimostazione di 4.3 abbiamo mostrato come

$$\prod_{j=1}^{n} \mathbb{P}(X_j = x_j) = \prod_{j=1}^{n} \theta_{x_j} = \prod_{i=1}^{p} \theta_i^{k_i}$$

Ne segue

$$L_\theta(x_1, \ldots, x_n) = (\#\Omega_{\mathbf{k}})^{-1} \mathbb{P}_\theta(F = \mathbf{k}) = \binom{n}{\mathbf{k}}^{-1} \mathbb{P}_\theta(F = \mathbf{k}) = \prod_{i=1}^{p} \theta_i^{k_i}$$

□

Proposizione 4.5. *Siano $X_1, \ldots, X_n$ variabili aleatorie indipendenti ed identicamente distribuite a valori in $\{1, \ldots, p\}$ tali che $\mathbb{P}(X_j = i) = \theta_i \ \forall i, j$. Per ogni $\boldsymbol{x}$ esito di $\boldsymbol{X}$, dato il relativo esito $\boldsymbol{k}$ del vettore delle frequenze F, lo stimatore di massima verosimiglianza di L_θ è $\hat{\theta} = (k_1/n, \ldots, k_p/n)$.*

Dimostrazione. Si tratta di massimizzare

$$L_\theta(x_1, \ldots, x_n) = \binom{n}{\mathbf{k}}^{-1} \mathbb{P}_\theta(F = \mathbf{k}) = \prod_{i=1}^{p} \theta_i^{k_i}$$

Applichiamo il Teorema dei moltiplicatori di Lagrange al logaritmo della verosimiglianza

$$\log L_\theta(x_1, \ldots, x_n) = \sum_{i=1}^{p} k_i \log \theta_i$$

tenendo in considerazione il vincolo $\sum_{i=1}^{p} \theta_i = 1$:

$$\sum_{i=1}^{p} k_i \log \theta_i - \lambda \left(\sum_{i=1}^{p} \theta_i - 1 \right)$$

Derivando rispetto a θ_i otteniamo

$$\frac{\partial}{\partial \theta_i} = \frac{k_i}{\theta_i} - \lambda \quad ; \quad \frac{k_i}{\hat{\theta}_i} - \lambda = 0 \Rightarrow \hat{\theta}_i = \frac{k_i}{\lambda}$$

Sommando $\forall i$ si ha

$$1 = \frac{n}{\lambda} \Rightarrow \lambda = 1 \Rightarrow \hat{\theta}_i = \frac{k_i}{n} \ \forall i$$

□

4.2 Test del Chi-Quadro per la bontà di adattamento

I risultati di questa sezione fanno riferimento a [1], capitolo **III.3**. Sia $\mathcal{A}$ una popolazione statistica di elementi raggruppati in p classi. Consideriamo $\Theta = \{\theta \in \mathbb{R}^p \| \ \theta_1 + \cdots + \theta_p = 1 \text{ e } \theta_1, \ldots, \theta_p \geq 0\}$ e definiamo $f_\theta(i) := \theta_i$. $(f_\theta)_{\theta \in \Theta}$ è una famiglia di distribuzioni discrete finite.
Supponiamo $\mathcal{A}$ abbia distribuzione f_θ ignota e, dato $\theta_0 \in \Theta$, vogliamo verificare se f_{θ_0} ben descrive $\mathcal{A}$. Costruiamo un test d'ipotesi

$$H_0 : \theta \in \Theta_0 \qquad H_1 : \theta \in \Theta_1 \qquad \Theta_0 = \{\theta_0\}$$

Sia $X_1, \ldots, X_n$ un campione della popolazione e $F = (F_1, \ldots, F_p)$ il corrispettivo vettore delle frequenze. F ha distribuzione multinomiale con parametro scalare n e parametro vettoriale θ.

Verosimiglianza

Sia $(i_1, \ldots, i_n)$ l'esito del campione e $(k_1, \ldots, k_p)$ le frequenze corrispondenti. Per la proposizione 4.4 la verosimiglianza è:

$$L_\theta(i_1, \ldots, i_n) = \binom{n}{k_1, \ldots, k_n}^{-1} \mathbb{P}_\theta(F = (k_1, \ldots, k_n)) = \theta_1^{k_1} \cdots \theta_p^{k_p}$$

Da cui si ottiene il rapporto di verosimiglianza:

$$\Lambda(k_1, \ldots, k_p) = \frac{\mathbb{P}_{\theta_0}(F = (k_1, \ldots, k_n))}{\sup_{\theta \in \Theta} \mathbb{P}_\theta(F = (k_1, \ldots, k_n))}$$

Teorema 4.6. *Dato l'esito* $(k_1, \ldots, k_p)$ *del vettore frequenza, il rapporto di verosimiglianza ha valore*

$$\Lambda(k_1, \ldots, k_p) = \left(\frac{n\theta_{1,0}}{k_1}\right)^{k_1} \cdots \left(\frac{n\theta_{p,0}}{k_p}\right)^{k_p}$$

Dimostrazione. Consideriamo il rapporto di verosimiglianza:

$$\Lambda(k_1, \ldots, k_p) = \frac{\mathbb{P}_{\theta_0}(F = (k_1, \ldots, k_n))}{\sup_{\theta \in \Theta} \mathbb{P}_\theta(F = (k_1, \ldots, k_n))}$$

Il numeratore della frazione è:

$$\mathbb{P}_{\theta_0}(F = (k_1, \ldots, k_n)) = \binom{n}{k_1, \ldots, k_n} \theta_{1,0}^{k_1} \cdots \theta_{n,0}^{k_n}$$

Per la proposizione 4.5 l'estremo superiore al denominatore si realizza per $\hat{\theta} = (k_1/n, \ldots, k_p/n)$. Il denominatore è dunque:

$$\begin{aligned} \sup_{\theta \in \Theta} \mathbb{P}_\theta(F = (k_1, \ldots, k_n)) &= \mathbb{P}_{\hat{\theta}}(F = (k_1, \ldots, k_n)) = \\ &= \binom{n}{k_1, \ldots, k_n} \left(\frac{k_1}{n}\right)^{k_1} \cdots \left(\frac{k_p}{n}\right)^{k_p} \end{aligned}$$

Concludiamo semplificando il rapporto

$$\begin{aligned} \Lambda(k_1, \ldots, k_p) &= \frac{\binom{n}{k_1, \ldots, k_n} \theta_{1,0}^{k_1} \cdots \theta_{n,0}^{k_n}}{\binom{n}{k_1, \ldots, k_n} \left(\frac{k_1}{n}\right)^{k_1} \cdots \left(\frac{k_p}{n}\right)^{k_p}} = \\ &= \left(\frac{n\theta_{1,0}}{k_1}\right)^{k_1} \cdots \left(\frac{n\theta_{p,0}}{k_p}\right)^{k_p} \end{aligned}$$

□

Considerando il rapporto di verosimiglianza come variabile aleatoria, otteniamo:

$$\Lambda(F_1, \ldots, F_p) = \prod_{i=1}^{p} \left(\frac{n\theta_{i,0}}{F_i}\right)^{F_i} = \prod_{i=1}^{p} \left(\frac{\mathbb{E}(F_i)}{F_i}\right)^{F_i}$$

Calcoliamo infine il logaritmo del rapporto di verosimiglianza:

$$\log \Lambda(F_1, \ldots, F_p) = \log \prod_{i=1}^{p} \left(\frac{\mathbb{E}(F_i)}{F_i}\right)^{F_i} = \sum_{i=1}^{p} F_i \log \left(\frac{\mathbb{E}(F_i)}{F_i}\right)$$

Proposizione 4.7. *Se vale H_0 allora la variabile aleatoria $-2 \log \Lambda(F_1, \ldots, F_p)$ ha distribuzione asintotica $\chi^2(p-1)$.*

Dimostrazione. H_0 è un'ipotesi semplice.
Per il teorema 3.3, $-2 \log \Lambda(F_1, \ldots, F_p) = -2 \log \Lambda(X_1, \ldots, X_n)$ ha distribuzione asintotica $\chi^2(r)$ con $r = \dim\Theta$.

$$\Theta = \{\theta \in \mathbb{R}^p \| \ \theta_1 + \cdots + \theta_p = 1 \text{ e } \theta_1, \ldots, \theta_p \geq 0\} \Rightarrow \dim\Theta = p - 1$$

□

Il test

Teorema 4.8. *Se vale H_0 allora, per n grande, la variabile*

$$\sum_{i=1}^{p} \frac{(F_i - \mathbb{E}(F_i))^2}{\mathbb{E}(F_i)}$$

ha approssimativamente una distribuzione χ^2 con $p-1$ gradi di libertà.

Dimostrazione. Si fornisce un argomento non rigoroso attraverso approssimazioni successive. $-2\log\Lambda(F_1,\ldots,F_p)$ ha distribuzione asintotica $\chi^2(p-1)$. È sufficiente dimostrare che per n grandi:

$$-2\log\Lambda(F_1,\ldots,F_p) = -2\sum_{i=1}^{p} F_i \log\left(\frac{\mathbb{E}(F_i)}{F_i}\right) \approx \sum_{i=1}^{p} \frac{(F_i - \mathbb{E}(F_i))^2}{\mathbb{E}(F_i)}$$

Analizziamo la variabile F_i/n:

$$\mathbb{E}\left(\frac{F_i}{n}\right) = \theta_{i,0} \qquad \mathrm{var}\left(\frac{F_i}{n}\right) = \frac{\theta_{i,0}(1-\theta_{i,0})}{n^2} \approx 0 \quad n >> 1$$

Per n grande possiamo dunque approssimare $F_i/n \approx \theta_{i,0}$ da cui otteniamo:

$$\frac{\mathbb{E}(F_i)}{F_i} = \frac{\theta_{i,0}}{F_i/n} \approx 1 \qquad \frac{\mathbb{E}(F_i)}{F_i} - 1 \approx 0$$

Utilizziamo ora lo sviluppo di Taylor al secondo ordine del logaritmo attorno al punto $x = 0$. Per $x \approx 0$ si ha l'approssimazione:

$$\log(1+x) \approx x - \frac{1}{2}x^2$$

Applicando al logaritmo di $\mathbb{E}(F_i)/F_i$ abbiamo:

$$\log\frac{\mathbb{E}(F_i)}{F_i} = \log\left[1 + \left(\frac{\mathbb{E}(F_i)}{F_i} - 1\right)\right] \approx \left(\frac{\mathbb{E}(F_i)}{F_i} - 1\right) - \frac{1}{2}\left(\frac{\mathbb{E}(F_i)}{F_i} - 1\right)^2$$

Da cui, moltiplicando per F_i e sommando per $i = 1,\ldots,p$, otteniamo:

$$\sum_{i=1}^{p} F_i \log\frac{\mathbb{E}(F_i)}{F_i} \approx \sum_{i=1}^{p} (\mathbb{E}(F_i) - F_i) - \frac{1}{2}\sum_{i=1}^{p} \frac{(\mathbb{E}(F_i) - F_i)^2}{F_i} \approx$$
$$\approx -\frac{1}{2}\sum_{i=1}^{p} \frac{(\mathbb{E}(F_i) - F_i)^2}{F_i}$$

Ricordando che $F_i/\mathbb{E}(F_i) \approx 1$ concludiamo:

$$-2\log\Lambda(F_1,\dots,F_p) \approx \sum_{i=1}^{p}\frac{(\mathbb{E}(F_i)-F_i)^2}{F_i} \approx \sum_{i=1}^{p}\frac{F_i}{\mathbb{E}(F_i)}\frac{(\mathbb{E}(F_i)-F_i)^2}{F_i} \approx$$
$$\approx \sum_{i=1}^{p}\frac{(\mathbb{E}(F_i)-F_i)^2}{\mathbb{E}(F_i)}$$

□

Utilizziamo il teorema 4.8 per verificare H_0. Se H_0 è valida allora

$$T(F_1,\dots,F_p) = \sum_{i=1}^{p}\frac{(F_i-\mathbb{E}(F_i))^2}{\mathbb{E}(F_i)}$$

ha distribuzione $\chi^2(p-1)$.
Fissata α significatività del test, definisco la zona critica:

$$G_\alpha = \{(k_1,\dots,k_p)\|T(k_1,\dots,k_p) > y_\alpha\} \qquad \text{con } y_\alpha \ t.c.\ \mathbb{P}(T > y_\alpha) = \alpha$$

Il test di ipotesi $(X_1,\dots,X_n,H_0,H_1,G_\alpha)$ così ottenuto è detto *test del Chi-Quadro per la bontà di adattamento.*

4.3 Test del Chi-Quadro per l'indipendenza

Questa sezione è tratta da [1], capitolo **III.4**. Sia $\mathcal{A}$ una popolazione statistica di elementi raggruppati in p classi rispetto alla variabile aleatoria X e in q classi rispetto a Y. Definito

$$\Theta = \left\{\theta \in \mathbb{R}^{p,q}\| \sum_{i,j}\theta_{i,j} = 1 \text{ e } \theta_{i,j} \geq 0\ \forall i,j\right\}$$

$\mathcal{A}$ ha distribuzione $f_\theta(i,j) := \theta_{i,j}$. Vogliamo costruire un test per verificare l'indipendenza di X e Y. Se le variabili sono statisticamente indipendenti, allora $\theta \in \Theta_0$ con

$$\Theta_0 = \left\{\theta \in \Theta\|\forall(i,j),\ \theta_{i,j} = \theta_{i,\bullet}\theta_{\bullet,j},\ \theta_{i,\bullet} = \sum_j \theta_{i,j},\ \theta_{\bullet,j} = \sum_i \theta_{i,j}\right\}$$

Si hanno le ipotesi

$$H_0 : \theta \in \Theta_0 \qquad H_1 : \theta \in \Theta \setminus \Theta_0$$

Sia $(X_1, Y_1), \ldots, (X_n, Y_n)$ un campione ed F la corrispettiva matrice delle frequenze. F ha distribuzione multinomiale con parametro scalare n e parametro vettoriale θ. Siano infine

$$F_{i,\bullet} = \sum_j F_{i,j} \quad F_{\bullet,j} = \sum_i F_{i,j} \quad E_{i,j} = \frac{F_{i,\bullet} F_{\bullet,j}}{n}$$

Verosimiglianza

Sia $((i_1, j_1), \ldots, (i_n, j_n))$ un esito del campione e $\mathbf{k}$ la corrispondente matrice delle frequenze. Per il corollario 4.4 la verosimiglianza è:

$$L_\theta\left((i_1, j_1), \ldots, (i_n, j_n)\right) = \binom{n}{\mathbf{k}}^{-1} \mathbb{P}(F = \mathbf{k}) = \prod_{i,j} \theta_{i,j}^{k_{i,j}}$$

Da cui si ottiene il rapporto di verosimiglianza:

$$\Lambda(\mathbf{k}) = \frac{\sup_{\theta \in \Theta_0} \mathbb{P}_\theta(F = \mathbf{k})}{\sup_{\theta \in \Theta} \mathbb{P}_\theta(F = \mathbf{k})}$$

Teorema 4.9. *Dato l'esito $\boldsymbol{k}$ del vettore frequenza, il rapporto di verosimiglianza ha valore*

$$\Lambda(\boldsymbol{k}) = \prod_{i,j} \left(\frac{k_{i,\bullet} k_{\bullet,j}}{n k_{i,j}} \right)^{k_{i,j}}$$

Dimostrazione. Il rapporto di verosimiglianza

$$\Lambda(\mathbf{k}) = \frac{\sup_{\theta \in \Theta_0} \mathbb{P}_\theta(F = \mathbf{k})}{\sup_{\theta \in \Theta} \mathbb{P}_\theta(F = \mathbf{k})}$$

consiste in due estremi superiori. Calcoliamo prima il denominatore: per 4.5 si ha

$$\sup_{\theta \in \Theta} \mathbb{P}_\theta(F = \mathbf{k}) = \binom{n}{\mathbf{k}} \sum_{i,j} \left(\frac{k_{i,j}}{n} \right)^{k_{i,j}}$$

Consideriamo il numeratore. Siccome X e Y sono indipendenti è sufficiente calcolare gli stimatori di massima verosimiglianza $\hat{\theta}_{i,\bullet}$ e $\hat{\theta}_{\bullet,j}$, da cui otteniamo $\hat{\theta}_{i,j} = \hat{\theta}_{i,\bullet}\hat{\theta}_{\bullet,j}$. Definita la funzione h_θ:

$$h_\theta(\mathbf{k}) := \binom{n}{\mathbf{k}}^{-1} \mathbb{P}_\theta(F = \mathbf{k}) = \prod_{i,j} (\theta_{i,j})^{k_{i,j}} = \prod_{i,j} (\theta_{i,\bullet}\theta_{\bullet,j})^{k_{i,j}}$$

applichiamo il Teorema dei moltiplicatori di Lagrange al suo logaritmo

$$\log h_\theta(\mathbf{k}) = \sum_{i,j} [k_{i,j}(\log \theta_{i,\bullet} + \log \theta_{\bullet,j})]$$

sfruttando i vincoli $\sum_j \theta_{i,\bullet} = 1$ e $\sum_i \theta_{\bullet,j} = 1$:

$$\sum_{i,j} [k_{i,j}(\log \theta_{i,\bullet} + \log \theta_{\bullet,j})] - \lambda_1 \left(\sum_j \theta_{i,\bullet} - 1 \right) - \lambda_2 \left(\sum_i \theta_{\bullet,j} - 1 \right)$$

Deriviamo rispetto a $\theta_{i,\bullet}$:

$$\frac{\partial}{\partial \theta_{i,\bullet}} = \sum_j \left(\frac{k_{i,j}}{\theta_{i,\bullet}} \right) - \lambda_1 = \frac{k_{i,\bullet}}{\theta_{i,\bullet}} - \lambda_1$$

$$\frac{k_{i,\bullet}}{\hat{\theta}_{i,\bullet}} - \lambda_1 = 0 \Rightarrow \hat{\theta}_{i,\bullet} = \frac{k_{i,\bullet}}{\lambda_1}$$

Sommando per ogni i otteniamo

$$1 = \frac{n}{\lambda_1} \Rightarrow \lambda_1 = n \Rightarrow \hat{\theta}_{i,\bullet} = \frac{k_{i,\bullet}}{n} \quad \forall i$$

Derivando rispetto a $\theta_{\bullet,j}$ si ottiene analogamente:

$$\hat{\theta}_{\bullet,j} = \frac{k_{\bullet,j}}{n} \quad \forall j$$

Ne segue pertanto che

$$\hat{\theta}_{i,j} = \hat{\theta}_{i,\bullet}\hat{\theta}_{\bullet,j} = \frac{k_{i,\bullet}k_{\bullet,j}}{n^2}$$

$$\sup_{\theta \in \Theta_0} \mathbb{P}_\theta(F = \mathbf{k}) = \binom{n}{\mathbf{k}} \prod_{i,j} \left(\frac{k_{i,\bullet}k_{\bullet,j}}{n^2} \right)^{k_{i,j}}$$

Semplificando il rapporto otteniamo dunque

$$\Lambda(\mathbf{k}) = \prod_{i,j} \left(\frac{k_{i,\bullet}k_{\bullet,j}}{nk_{i,j}} \right)^{k_{i,j}}$$

□

Trattando Λ come variabile aleatoria e calcolandone il logaritmo si ottiene

$$\Lambda(F) = \prod_{i,j} \left(\frac{F_{i,\bullet} F_{\bullet,j}}{n F_{i,j}} \right)^{F_{i,j}} = \prod_{i,j} \left(\frac{E_{i,j}}{F_{i,j}} \right)^{F_{i,j}}$$

$$\log \Lambda(F) = \sum_{i,j} F_{i,j} \log \left(\frac{E_{i,j}}{F_{i,j}} \right)$$

Proposizione 4.10. *Se vale H_0 allora la variabile aleatoria $-2 \log \Lambda(F)$ ha distribuzione asintotica $\chi^2((p-1)(q-1))$.*

Dimostrazione. H_0 è un'ipotesi composita.
Per il teorema 3.4, $-2 \log \Lambda(F_1, \ldots, F_n)$ ha distribuzione asintotica $\chi^2(r - r')$ con $r = dim\Theta$ e $r' = dim\Theta_0$. Il calcolo della dimensione di Θ è molto semplice

$$\Theta = \left\{ \theta \in \mathbb{R}^{p,q} \| \sum_{i,j} \theta_{i,j} = 1 \text{ e } \theta_{i,j} \geq 0 \ \forall i,j \right\} \Rightarrow dim\Theta = pq - 1$$

Consideriamo ora Θ_0:

$$\Theta_0 = \left\{ \theta \in \Theta \| \forall (i,j),\ \theta_{i,j} = \theta_{i,\bullet} \theta_{\bullet,j},\ \theta_{i,\bullet} = \sum_j \theta_{i,j},\ \theta_{\bullet,j} = \sum_i \theta_{i,j} \right\}$$

Definisco gli insiemi:

$$\Theta_X = \left\{ \theta \in \mathbb{R}^p \| \sum_i \theta_i = 1 \text{ e } \theta_i \geq 0 \ \forall i \right\} \Rightarrow dim\Theta_X = p - 1$$

$$\Theta_Y = \left\{ \theta \in \mathbb{R}^q \| \sum_j \theta_j = 1 \text{ e } \theta_j \geq 0 \ \forall j \right\} \Rightarrow dim\Theta_Y = q - 1$$

Si può dunque affermare

$$\Theta_0 = \{ \theta \in \mathbb{R}^{p,q} \| \theta_{i,j} = (\theta_X)_i (\theta_Y)_j,\ \theta_X \in \Theta_X, \theta_Y \in \Theta_Y \}$$

La dimensione di Θ_0 è pertanto $dim\Theta_0 = dim\Theta_X + dim\Theta_Y = p + q - 2$
Si conclude verificando che $r - r' = (pq - 1) - (p + q - 2) = (p - 1)(q - 1)$. $\square$

Il test

Teorema 4.11. *Se vale H_0 allora, per n grande, la variabile*

$$\sum_{i,j} \frac{(F_{i,j} - E_{i,j})^2}{E_{i,j}}$$

ha approssimativamente una distribuzione χ^2 con $(p-1)(q-1)$ gradi di libertà.

Dimostrazione. Si fornisce un argomento non rigoroso attraverso approssimazioni successive. $-2\log\Lambda(F)$ ha distribuzione asintotica $\chi^2((p-1)(q-1))$. É sufficiente dimostrare che per n grandi:

$$-2\log\Lambda(F) = -2\sum_{i,j} F_{i,j}\log\left(\frac{E_{i,j}}{F_{i,j}}\right) \approx \sum_{i,j} \frac{(F_{i,j} - E_{i,j})^2}{E_{i,j}}$$

Fissiamo $\theta \in \Theta_0$. Analogamente a quanto visto nella dimostrazione del teorema 4.8, si può dimostrare come

$$\frac{F_{i,\bullet}}{\mathbb{E}_\theta(F_{i,\bullet})} \approx 1 \quad , \quad \frac{F_{\bullet,j}}{\mathbb{E}_\theta(F_{\bullet,j})} \approx 1 \quad , \quad \frac{\mathbb{E}_\theta(F_{i,j})}{F_{i,j}} \approx 1$$

Si ha dunque

$$\begin{aligned} \frac{E_{i,j}}{F_{i,j}} =& \frac{F_{i,\bullet}}{\mathbb{E}_\theta(F_{i,\bullet})} \cdot \frac{F_{\bullet,j}}{\mathbb{E}_\theta(F_{\bullet,j})} \cdot \frac{\mathbb{E}_\theta(F_{i,j})}{F_{i,j}} \cdot \frac{\mathbb{E}_\theta(F_{i,\bullet})\mathbb{E}_\theta(F_{\bullet,j})}{n\mathbb{E}_\theta(F_{i,j})} \approx \\ \approx& \frac{\mathbb{E}_\theta(F_{i,\bullet})\mathbb{E}_\theta(F_{\bullet,j})}{n\mathbb{E}_\theta(F_{i,j})} = \frac{n\theta_{i,\bullet} \cdot n\theta_{\bullet,j}}{n \cdot n\theta_{i,j}} = 1 \end{aligned}$$

Possiamo dunque sviluppare con Taylor il logaritmo

$$\log\frac{E_{i,j}}{F_{i,j}} = \log\left[1 + \left(\frac{E_{i,j}}{F_{i,j}} - 1\right)\right] \approx \left(\frac{E_{i,j}}{F_{i,j}} - 1\right) - \frac{1}{2}\left(\frac{E_{i,j}}{F_{i,j}} - 1\right)^2$$

Da cui, moltiplicando per $F_{i,j}$ e sommando per i, j, otteniamo:

$$\begin{aligned} \sum_{i,j} F_{i,j}\log\frac{E_{i,j}}{F_{i,j}} &\approx \sum_{i,j}(E_{i,j} - F_{i,j}) - \frac{1}{2}\sum_{i,j}\frac{(F_{i,j} - E_{i,j})^2}{F_{i,j}} \approx \\ &\approx -\frac{1}{2}\sum_{i,j}\frac{(F_{i,j} - E_{i,j})^2}{F_{i,j}} \end{aligned}$$

Siccome $F_{i,j}/E_{i,j} \approx 1$, si può concludere infine:

$$-2 \log \Lambda(F) \approx \sum_{i,j} \frac{(F_{i,j} - E_{i,j})^2}{F_{i,j}} \approx \sum_{i,j} \frac{(F_{i,j} - E_{i,j})^2}{E_{i,j}}$$

$\square$

Utilizziamo il teorema 4.11 per verificare H_0. Se H_0 è valida allora

$$T(F) = \sum_{i,j} \frac{(F_{i,j} - E_{i,j})^2}{E_{i,j}}$$

ha distribuzione $\chi^2((p-1)(q-1))$.
Fissata α la significatività del test, definisco la zona critica:

$$G_\alpha = \{\mathbf{k} \| T(\mathbf{k}) > y_\alpha\} \qquad \text{con } y_\alpha \text{ } t.c. \text{ } \mathbb{P}(T > y_\alpha) = \alpha$$

Il test $((X_1, Y_1), \ldots, (X_n, Y_n), H_0, H_1, G_\alpha)$ è detto *test del Chi-Quadro per l'indipendenza.*

Conclusioni

La convergenza normale della massima verosimiglianza e del logaritmo del rapporto garantisce la grande efficacia del test del Chi-Quadro nella verifica di ipotesi riguardanti variabili aleatorie discrete finite.
É possibile estendere l'utilizzo del test del Chi-Quadro a variabili numerabili o continue raggruppando la popolazione in un numero finito di categorie in base al valore della variabile aleatoria.
Ciò rende il test del Chi-Quadro uno dei più potenti strumenti statistici a nostra disposizione.

Bibliografia

[1] Pestman, Wiebe R., *Mathematical Statistics*, 2^{nd} Edition, de Gruyter, Berlin, 2010.

[2] Wilks, Samuel S., *Mathematical Statistics*, John Wiley & Sons Inc., New York, 1962.

[3] Lindley, Dennis V., *Introduction to Probability and Statistics from a Bayesian viewpoint, part 2: Inference*, Cambridge University Press, Cambridge, 1965.