

CAPITOLO X - Montecarlo - Simulazione

Monte Carlo: metodo probabilistico per la soluzione di problemi non probabilistici.

Monte Carlo per il calcolo degli integrali: prese X_1, X_2, \dots, X_n variabili IID, si considera la statistica:

$$G = \frac{1}{n} \sum_{i=1}^n g(X_i)$$
$$E(G) = E\left(\frac{1}{n} \sum_{i=1}^n g(X_i)\right) = \frac{1}{n} \sum_{i=1}^n E[g(X_i)] = E[g(X_i)]$$

cioè G (media aritmetica o sample mean delle $g(X_i)$) ha la stessa media di $g(X_i)$. In questo caso si dice che G è uno **stimatore non distorto** della media di $g(X_i)$. La varianza di G è:

$$\text{var}(G) = E(G^2) - E(G)^2 = \text{var}\left(\frac{1}{n} \sum_{i=1}^n g(X_i)\right) = \sum_{i=1}^n \frac{1}{n^2} \text{var}[g(X_i)]$$

quindi, al crescere del rango del campione n , la varianza della media di $g(X_i)$ decresce come $1/n$. Questo risultato contiene l'idea essenziale per il calcolo degli integrali tramite la simulazione Monte Carlo, cioè un integrale può essere stimato da una somma:

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx = E\left(\frac{1}{n} \sum_{i=1}^n g(X_i)\right)$$

Distribuzione della media di una v.a. - un teorema fondamentale: tutto si basa sulla “**legge dei grandi numeri**” della teoria della probabilità. Date n variabili random IID con μ valore d'aspettazione di ciascuna variabile X_i , per $n \rightarrow \infty$ la sample mean delle X_i è:

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

converge a μ quasi certamente (e quindi anche in probabilità):

$$P\left\{\lim_{n \rightarrow \infty} \bar{X} = \mu\right\} = 1$$

Questo teorema implica che la media di n variabili random campionate converge (in probabilità) al suo valore d'aspettazione. Si può usare il teorema di Chebyshev per avere una stima della possibilità di ottenere una grande (e non desiderata) deviazione in un calcolo Monte Carlo. Dato lo stimatore G e supposto che esistano la media e la varianza di G , la disuguaglianza di Chebyshev porta a:

$$P\left(|G - \mu_G| \geq \sqrt{\sigma_G^2 / \delta}\right) \leq \delta \quad \text{con } \delta > 0$$

con μ_G e σ_G media e deviazione standard di G . Considerando un valore di n sufficientemente grande, la varianza di G diventa piccola e quindi la probabilità di ottenere una deviazione grande rispetto a δ tra la stima dell'integrale e il valore vero è molto piccola; questo fatto sta al cuore del metodo Monte Carlo per la stima degli integrali.

Trasformazioni basate sull'integrale di probabilità: xxx

Checking Goodness of Fit: xxx

Simulazione di distribuzioni continue: xxx

CAPITOLO XI - Modello statistico - Metodi di verosimiglianza

I problemi della statistica: applicazioni di modelli di probabilità in problemi di analisi e interpretazione dati.

Stima di massima verosimiglianza o MLE: si suppone che un modello di probabilità per un dato esperimento sia stato formulato e che esso coinvolga un singolo parametro incognito θ . Viene eseguito l'esperimento e alcuni dati vengono raccolti in modo da stimare il valore di θ , ossia si vuole vedere quali dei possibili valori di θ siano plausibili o verosimili alla luce delle osservazioni. Si chiama spazio del parametro o **parameter space** l'insieme Ω dei possibili valori di θ . I dati osservati sono visti come un evento E , nello spazio campionario relativo al modello di probabilità, con probabilità $P(E; \theta)$. La MLE

(maximum likelihood estimate) $\hat{\theta}$ è il valore di θ che rende massima la $P(E; \theta)$, cioè il valore del parametro che meglio spiega i dati.

Esempio 11.4.a Supponiamo di volere determinare θ , la proporzione di persone con tubercolosi in una grande e omogenea popolazione. A tal fine selezioniamo casualmente n individui per il test, e troviamo che x di questi risultano ammalati. Poiché la popolazione è grande ed omogenea, possiamo supporre indipendenti gli n individui (approssimazione binomiale dell'ipergeometrica) e che ciascuno ha probabilità θ di avere la malattia; stiamo quindi ipotizzando valido il modello di probabilità della binomiale. La probabilità dell'evento E osservato (i dati) è data quindi da:

$$P(E; \theta) \equiv P(x \text{ su } n \text{ hanno la tubercolosi}) = \binom{n}{x} \theta^x (1 - \theta)^{n-x} \quad \text{con } 0 \leq \theta \leq 1$$

La MLE di θ è $\hat{\theta} = \frac{x}{n}$ (vedi esempio 11.4b)

Funzione di Verosimiglianza o LF o Likelihood Function: per semplificare le espressioni, visto che le costanti non hanno alcun effetto nel rendere massima la probabilità, si omettono le costanti e si ha:

$$L(\theta) = c \cdot P(E; \theta) \quad \text{con } c \text{ costante positiva non dipende da } \theta \text{ ma può dipendere dai dati}$$

Logaritmo della Funzione di Verosimiglianza: solitamente $L(\theta)$ e $P(E, \theta)$ sono prodotti di termini, e risulta più conveniente lavorare con il logaritmo naturale di L :

$$l(\theta) = \ln[L(\theta)] = c' + \ln[P(E; \theta)] \quad \text{con } c' = \ln(c)$$

Equazione di massima verosimiglianza:

$$\text{Score function} \Rightarrow S(\theta) = \frac{d[l(\theta)]}{d\theta}$$

Soluzione dell'equazione di verosimiglianza $\Rightarrow S(\theta) = 0$

$$\text{Information function} \Rightarrow Y(\theta) = -\frac{d^2[l(\theta)]}{d\theta^2}$$

$$a < \hat{\theta} < b \Rightarrow \begin{cases} S(\theta) = 0 \\ Y(\theta) > 0 \end{cases} \quad \text{con } \Omega = [a, b] \in \mathbb{R}$$

$$\hat{\theta} = a, b \Rightarrow \begin{cases} S(\theta) = 0 \\ Y(\theta) > 0 \end{cases} \quad \text{con } \Omega = [a, b] \in \mathbb{R}$$

Esempio 11.4.b Per questo esempio la funzione S e la funzione Y sono:

$$S(\theta) = \frac{d l(\theta)}{d\theta} = \frac{x}{\theta} - \frac{(n-x)}{1-\theta} \quad \text{con } 0 < \theta < 1$$

$$Y(\theta) = -\frac{d S(\theta)}{d\theta} = \frac{x}{\theta^2} + \frac{(n-x)}{(1-\theta)^2} \quad \text{con } 0 < \theta < 1$$

$$\text{Massimo} = \left. \begin{aligned} &\text{per } x \neq 0, n \Leftrightarrow 1 \leq x \leq n-1 \Rightarrow \left\{ \begin{aligned} &\left\{ \begin{aligned} &S(\theta) = 0 \\ &Y(\theta) > 0 \end{aligned} \right\} \Rightarrow \theta = \frac{x}{n} \\ &L(\theta) = 0 \quad \text{per } \theta = 0, 1 \end{aligned} \right\} \Rightarrow \hat{\theta} = \frac{x}{n} \\ &\text{per } x = 0, 1 \Rightarrow \left\{ \begin{aligned} &\text{Max}[P(E; \theta) = (1-\theta)^n] \Leftrightarrow \theta = 0 \\ &\text{Max}[P(E; \theta) = \theta^n] \Leftrightarrow \theta = 1 \end{aligned} \right\} \Rightarrow \hat{\theta} = 0 \end{aligned} \right\} \Rightarrow \hat{\theta} = \frac{x}{n} \quad \forall x$$

Esempio 11.5 Per vedere se l'acqua di un fiume è sicura, alcuni test stabiliscono la concentrazione batterica. Viene determinato il numero di batteri per ciascuno degli n campioni di volume unitario, cioè n conteggi osservati x_1, \dots, x_n . Si deve stimare μ , numero medio di batteri nel fiume per unità di volume.

Assumiamo che i batteri siano distribuiti a caso e uniformemente (processo di Poisson); la probabilità di osservare x_i batteri nell' i -esimo campione è data quindi dalla poissoniana:

$$f(x_i) = \frac{\mu^{x_i}}{x_i!} e^{-\mu} \quad \text{con } x_i = 0, 1, \dots$$

Poiché volumi disgiunti sono indipendenti, la probabilità degli n conteggi osservati è data dal prodotto:

$$P(E; \mu) = f(x_1) \cdot f(x_2) \cdot \dots \cdot f(x_n) = \prod_{i=1}^n \frac{\mu^{x_i}}{x_i!} e^{-\mu} = \frac{\mu^{\sum x_i} e^{-n\mu}}{x_1! x_2! \dots x_n!}$$

Per la likelihood function scegliamo $c = 1/x_1! x_2! \dots x_n!$, quindi

$$L(\mu) = \mu^{\sum x_i} e^{-n\mu} \quad \text{con } 0 \leq \mu < \infty$$

$$l(\mu) = \sum x_i \cdot \ln(\mu) - n\mu$$

$$S(\mu) = \frac{1}{\mu} \sum x_i - n$$

$$Y(\mu) = \sum x_i / \mu^2$$

Per determinare la MLE di μ dobbiamo distinguere i due casi:

$$\sum x_i > 0 \Rightarrow \left\{ \begin{array}{l} \left\{ \begin{array}{l} S(\mu) = 0 \Leftrightarrow \mu = \frac{1}{n} \sum x_i \equiv \bar{x} \\ Y(\mu = \bar{x}) > 0 \end{array} \right\} \\ \left\{ \begin{array}{l} L(0) = 0 \\ L(\mu) \rightarrow 0 \quad \text{per } \mu \rightarrow \infty \end{array} \right\} \end{array} \right\} \Rightarrow \mu = \bar{x}$$

Se $\sum x_i = 0$, l'equazione $S(\mu) = 0$ non ha soluzione. Guardando alla $L(\mu)$ si ha che il massimo è per $\mu=0$; possiamo concludere che in entrambi i casi la MLE di μ è $\mu = \bar{x}$. Seguendo il criterio di rendere massima la probabilità dell'evento osservato x_1, \dots, x_n , otteniamo una stima della media μ relativa all'intera popolazione pari alla media \bar{x} sul campione.

Esempio 11.6 Di solito si può soltanto determinare se batteri sono presenti o meno nel campione. Vengono testati n campioni, ciascuno contenente un volume v di acqua; un test negativo significa assenza di batteri, un test positivo vuol dire che almeno un batterio è presente nel campione. Supponiamo che su n campioni testati y di questi risultano negativi: qual è la MLE di μ ?

L'ipotesi che i batteri siano distribuiti a caso e uniformemente rende plausibile che la distribuzione batterica segua la legge di Poisson; la probabilità che vi siano x batteri in un volume v d'acqua è data da:

$$f(x) = \frac{(\mu v)^x}{x!} e^{-\mu v} \quad \text{con } \begin{cases} x_i = 0, 1, \dots \\ \text{media} = \mu \cdot v \end{cases}$$

Probabilità p di una reazione negativa (nessun batterio) $\Rightarrow p = f(0) = e^{-\mu v}$

Probabilità di una reazione positiva (almeno un batterio) $\Rightarrow 1 - p = 1 - e^{-\mu v}$

Poiché volumi disgiunti sono indipendenti, n test costituiscono n prove indipendenti. La probabilità di osservare y reazioni negative su n è data dalla binomiale $b(n, p)$:

$$P(E; \mu) = \binom{n}{y} p^y (1-p)^{n-y}$$

$$c = 1 / \binom{n}{y}$$

$$L(\mu) = p^y (1-p)^{n-y} \Rightarrow \hat{p} = \frac{y}{n}$$

Il valore corrispondente per μ è:

$$\hat{\mu} = -\frac{1}{v} \ln(\hat{p}) = -\frac{1}{v} \ln\left(\frac{y}{n}\right) = \frac{\ln(n) - \ln(y)}{v}$$

Verosimiglianza per tabelle di frequenze: consideriamo un esperimento che dia luogo a differenti

risultati, e supponiamo di essere interessati a particolari sottoinsiemi A_1, \dots, A_k che formano una partizione dello spazio campionario S :

$$S = A_1 \cup A_2 \cup \dots \cup A_k \quad A_i \cap A_j = \emptyset \quad \text{per } i \neq j$$

Indichiamo con A_i l'evento generico che si avvera nel singolo esperimento e consideriamo n ripetizioni indipendenti dell'esperimento. Sia X_i la v.a. che descrive il numero di volte in cui si presenta l'evento A_i , e x_i un possibile valore di X_i , come mostrato in tabella ($\sum X_i = n$):

Eventi o classi	A_1	A_2	...	A_k	totale
Frequenze osservate	x_1	x_2	...	x_k	n
Frequenze attese	np_1	np_2	...	np_k	n

Sia p_i la probabilità che si avveri A_i in ciascuna ripetizione, con $\sum p_i = 1$. Le p_i possono essere determinate da un modello di probabilità e saranno generalmente funzioni di un parametro incognito θ . Il modello può essere quello della distribuzione multinomiale: a differenza del tipo di esperimento che porta alla binomiale, in questo caso i possibili risultati di ciascuna prova (indipendente) sono k e non due. Abbiamo visto che la probabilità di ottenere una data tabella di frequenza (cioè x_1 volte A_1 , x_2 volte A_2 ecc.) è data da:

$$P(E; \theta) = \binom{n}{x_1 x_2 \dots x_k} p_1^{x_1} p_2^{x_2} \dots p_k^{x_k}$$

$$\text{LF} \Rightarrow L(\theta) = c p_1^{x_1} p_2^{x_2} \dots p_k^{x_k}$$

$$\text{MLE} \Rightarrow \hat{\theta} = \bar{x} = \frac{x}{n}$$

Usando $\hat{\theta}$ così ottenuto ($\hat{\theta}$ è contenuto nelle p_i) possiamo stimare le frequenze attese $n\hat{p}_i$ e confrontarle con quelle osservate.

Combinando esperimenti indipendenti: un esperimento che dia informazioni sul parametro θ può essere ripetuto, ottenendo così ulteriore informazione. Supponiamo che due esperimenti indipendenti diano informazioni su θ . L'esperimento 1 dà i dati E_1 con probabilità $P(E_1; \theta)$, e la LF di θ basata sul primo esperimento è:

$$L_1(\theta) = c_1 \cdot P(E_1; \theta)$$

Analogamente, per il secondo esperimento si ha:

$$L_2(\theta) = c_2 \cdot P(E_2; \theta)$$

I due esperimenti vengono considerati come componenti di un unico esperimento composito. Lo spazio campionario è un prodotto cartesiano, e i dati relativi all'esperimento composito corrispondono all'evento $E_1 E_2$ in questo spazio. La probabilità dei dati è $P(E_1 E_2; \theta)$ e la LF basata su entrambi gli esperimenti è:

$$L(\theta) = c \cdot P(E_1 E_2; \theta)$$

Poiché gli esperimenti sono indipendenti si ha:

$$P(E_1 E_2; \theta) = P(E_1; \theta) \cdot P(E_2; \theta)$$

$$L(\theta) = L_1(\theta) \cdot L_2(\theta)$$

$$l(\theta) = l_1(\theta) + l_2(\theta)$$

$$S(\theta) = S_1(\theta) + S_2(\theta)$$

$$Y(\theta) = Y_1(\theta) + Y_2(\theta)$$

Siano $\hat{\theta}_1$, $\hat{\theta}_2$ e $\hat{\theta}$ le MLE basate sul primo, sul secondo, e su entrambi gli esperimenti; $\hat{\theta}_1$ rende massima la $l_1(\theta)$, $\hat{\theta}_2$ rende massima la $l_2(\theta)$ e $\hat{\theta}$ rende massima la $l(\theta)$. Tranne che in casi particolari, non è possibile calcolare $\hat{\theta}$ da $\hat{\theta}_1$ e $\hat{\theta}_2$: è necessario calcolare il massimo della $l(\theta)$. Se $\hat{\theta}_1 = \hat{\theta}_2$ allora $\hat{\theta} = \hat{\theta}_1 = \hat{\theta}_2$. In generale $\hat{\theta}$ è compreso tra $\hat{\theta}_1$ e $\hat{\theta}_2$.

Funzione di Verosimiglianza Relativa o RLF o Relative Likelihood Function:

$$R(\theta) = \frac{L(\theta)}{L(\hat{\theta})} = \frac{cP(E; \theta)}{cP(E; \hat{\theta})} = \frac{P(E; \theta)}{P(E; \hat{\theta})}$$

Essendo $L(\theta) \leq L(\hat{\theta}) \quad \forall \theta \Rightarrow 0 \leq R(\theta) \leq 1$

Logaritmo della Funzione di Verosimiglianza Relativa:

$$r(\theta) = \ln[R(\theta)] = \ln[L(\theta)] - \ln[L(\hat{\theta})] = l(\theta) - l(\hat{\theta}) \quad \text{con } -\infty \leq r(\theta) \leq 0$$

Fissato un valore θ_1 del parametro, possiamo dire che:

$$R(\theta_1) = \frac{\text{probabilità dei dati } E \text{ quando } \theta = \theta_1}{\text{probabilità massima di } E}$$

Se, per es., $R(\theta_1) = 0.1$, diciamo che θ_1 è un valore del parametro piuttosto non plausibile (10%); se $R(\theta_1) = 0.5$ allora θ_1 è un valore del parametro abbastanza plausibile (50%).

Regioni di Verosimiglianza al 100q%: il set di valori di θ per cui:

$$R(\theta) \geq q$$

Intervalli di Verosimiglianza o LI o Likelihood Interval: usualmente tali regioni consistono di intervalli di numeri reali. Gli intervalli usualmente considerati sono al 50%, 10% e 1%. Valori di θ interni all'intervallo al 10% vengono detti "plausibili", e valori esterni a tale intervallo "non plausibili". Valori interni all'intervallo al 50% sono detti "molto plausibili", mentre valori esterni all'intervallo al 1% "molto non plausibili". Gli intervalli possono essere determinati dal grafico di $R(\theta)$ o del suo ln:

$$R(\theta) \geq q \Leftrightarrow r(\theta) \geq \ln(q) \Leftrightarrow \begin{cases} \text{LI al 50\%} \Leftrightarrow r(\theta) \geq \ln(0.5) = -0.69 \\ \text{LI al 10\%} \Leftrightarrow r(\theta) \geq \ln(0.1) = -2.30 \\ \text{LI al 1\%} \Leftrightarrow r(\theta) \geq \ln(0.01) = -4.61 \end{cases}$$

Verosimiglianza per modelli continui: distribuzioni di probabilità continue sono usate come modelli di probabilità per esperimenti che coinvolgono misure di tempi, pesi, lunghezze, ecc. Sia X una v.a. continua con densità f e cumulativa F , dipendenti da un parametro θ . Viene eseguito l'esperimento e valori di X vengono osservati; il problema è di determinare quali valori di θ sono plausibili alla luce dei dati. Un valore osservato x corrisponderà ad un qualche piccolo intervallo di valori reali, diciamo $a < X \leq b$. La probabilità di osservare il valore x è allora:

$$P(a < X \leq b) = \int_a^b f(x) \cdot dx = F(b) - F(a)$$

Supponiamo che in n ripetizioni indipendenti si osservino n valori x_1, x_2, \dots, x_n , con x_i corrispondente all'intervallo reale $[a_i, b_i]$. Poiché le ripetizioni sono indipendenti, la probabilità dei dati si ottiene da:

$$P(E; \theta) = \prod_{i=1}^n P(a_i < X \leq b_i) = \prod_{i=1}^n [F(b_i) - F(a_i)]$$

La funzione di verosimiglianza per θ è proporzionale a questa espressione. Quando l'intervallo $\Delta_i = b_i - a_i$ è sufficientemente piccolo possiamo usare l'approssimazione:

$$P(a_i < X \leq b_i) = F(b_i) - F(a_i) \cong f(x_i) \cdot \Delta_i$$

Se questa approssimazione vale per tutti gli intervalli, si ha:

$$P(E; \theta) \cong \prod_{i=1}^n f(x_i) \Delta_i = \left[\prod_{i=1}^n \Delta_i \right] \prod_{i=1}^n f(x_i)$$

$$L(\theta) = c \prod_{i=1}^n f(x_i)$$

CAPITOLO XII – Distribuzione campionaria

Distribuzione campionaria o Sapling Distributions: per valutare e confrontare procedure statistiche, risulta utile esaminare come queste si comporterebbero in una ipotetica serie di ripetizioni

dell'esperimento. Assumiamo che in tutte le ripetizioni θ abbia lo stesso valore θ_0 , detto valore vero di θ . Se l'esperimento fosse realmente ripetuto, noi otterremmo molto probabilmente un set differente di dati. La likelihood function dipende dai dati, e quindi la ripetizione dell'esperimento darebbe luogo verosimilmente ad una diversa $r(\theta)$, come pure cambierebbero la MLE e gli estremi dell'intervallo di verosimiglianza. Tali quantità diventano quindi v.a. e le corrispondenti distribuzioni di probabilità, che hanno origine dalle ipotetiche ripetizioni dell'esperimento, sono dette distribuzioni campionarie con θ fissato.

Statistica del Rapporto di Verosimiglianza o LRS o Likelihood Ratio Statistics: teoricamente è possibile ricavare dal modello di probabilità distribuzioni campionarie esatte, tuttavia ciò può risultare molto difficile e si ricorre a delle approssimazioni. Differenti quantità associate al log della RLF possono essere studiate; la più importante di queste applicazioni statistiche è:

$$D \equiv -2 \cdot r(\theta_0) \equiv -2[l(\theta_0) - l(\hat{\theta})]$$

Concettualmente la statistica D è identica alla statistica $r(\theta_0)$; risulta più conveniente di $r(\theta_0)$ perché D è non negativa, mentre $r(\theta_0)$ è non positiva. Il fattore 2 viene incluso perché semplifica un po' le cose negli esempi di distribuzioni normali. Nel ripetere l'esperimento, $r(\theta_0)$ e D variano in accordo ai dati ottenuti; D sarà piccola quando i dati sono tali che il valore vero θ_0 è un valore verosimile di θ , e viceversa. Possiamo considerare D come una v.a. la cui distribuzione campionaria può essere derivata dal modello. Come si è detto, le ripetizioni cui si fa riferimento sono puramente ipotetiche.

Esempio 12.3 Consideriamo n misure x_1, x_2, \dots, x_n che modelliamo come valori osservati di n variabili indipendenti (IID) $N(\mu, \sigma^2)$. Vogliamo determinare la distribuzione campionaria della media ignota μ , assumendo che σ^2 sia nota. Se gli intervalli di misura sono piccoli, la LF di μ è:

$$L(\mu) = c \prod_{i=1}^n f(x_i) = c \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}} = e^{-\frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2}}$$

$$l(\mu) = -\frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2}$$

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

$$S(\mu) = 0 \Leftrightarrow \hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}$$

$$Y(\mu) = \frac{n}{\sigma^2}$$

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

L'intervallo di verosimiglianza al 100 q % è dato da:

$$\bar{x} - \frac{c \cdot \sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + \frac{c \cdot \sigma}{\sqrt{n}} \quad \text{con } c = \sqrt{-2 \cdot \ln(q)}$$

Immaginiamo ora che l'esperimento venga ripetuto diverse volte con $\mu = \mu_0$. Ciascuna delle osservazioni x_i varia di volta in volta, e \bar{x} diventa una v.a.:

$$\bar{X} = \frac{1}{n} (X_1 + X_2 + \dots + X_n) \sim N\left(\mu_0, \frac{\sigma^2}{n}\right)$$

Anche la stima di massima verosimiglianza $\hat{\mu}$ diventa una v.a., e da:

$$S(\mu) = 0 \Leftrightarrow \hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x} \Rightarrow \hat{\mu} \equiv \bar{X} \sim N\left(\mu_0, \frac{\sigma^2}{n}\right)$$

Questa è la distribuzione campionaria della MLE per ripetizioni indipendenti dell'esperimento con $\mu = \mu_0$. Allo stesso modo, gli estremi dell'LI al 100q% vengono ora guardati come v.a., $\bar{X} - c \cdot \sigma / \sqrt{n}$ e $\bar{X} + c \cdot \sigma / \sqrt{n}$.

$$D \equiv -2r(\mu_0) \equiv \frac{n}{\sigma^2} (\bar{X} - \mu_0)^2 \equiv Z^2 \quad \text{con } Z \equiv (\bar{X} - \mu_0) / \sqrt{\sigma^2/n} \sim N(0,1) \text{ forma standard di } \bar{X}$$

$$Z \sim N(0,1) \Rightarrow Z^2 \sim \chi_{(1)}^2 \Rightarrow D \equiv Z^2 \sim \chi_{(1)}^2 \quad \forall \mu_0$$

$$\lim_{n \rightarrow \infty} P(D \leq d; \theta = \theta_0) = P(\chi_{(1)}^2 \leq d)$$

Probabilità di Copertura CP o Coverage Probability: supponiamo che il modello di probabilità per un dato esperimento coinvolga un solo parametro ignoto θ . Immaginiamo una serie di ripetizioni dell'esperimento, con θ fissato a θ_0 . Supponiamo che si debba determinare un intervallo $[A, B]$ dai dati, che per es. potrebbe essere l'intervallo di verosimiglianza al 10% per θ . A causa del carattere aleatorio dei risultati, l'intervallo $[A, B]$ varia da una ripetizione all'altra, ed i suoi estremi possono essere modellati come variabili random, con distribuzione da determinare dal modello di probabilità e, verosimilmente, dipendente da θ_0 . Poiché $[A, B]$ varia da una ripetizione all'altra, può succedere (sfortunatamente) che l'intervallo non contenga il valore vero θ_0 . Definiamo probabilità di copertura dell'intervallo aleatorio $[A, B]$ la probabilità che lo stesso intervallo contenga il valore vero θ_0 del parametro:

$$CP(\theta_0) = P(A \leq \theta_0 \leq B; \theta = \theta_0)$$

La CP può essere vista come la frazione di tempo in cui l'intervallo $[A, B]$ contiene il valore vero θ_0 in un gran numero di ripetizioni dell'esperimento con $\theta = \theta_0$. Da notare che A e B sono variabili random, mentre θ_0 è fissato.

CP in termini della statistica D: l'intervallo di verosimiglianza al 100q% per θ è il set di valori di θ tali che $R(\theta) \geq q \Leftrightarrow r(\theta) \geq \ln(q)$. Essendo $D \equiv -2 \cdot r(\theta_0)$, segue che θ_0 appartiene a tale intervallo se e solo se $D \leq -2 \cdot \ln(q)$. Quindi, la CP dell'intervallo di verosimiglianza al 100q% è:

$$CP(\theta_0) = P(\theta_0 \text{ appartenga all'LI al } 100q\%; \theta = \theta_0) = P(D \leq -2 \cdot \ln(q); \theta = \theta_0) = x$$

$$\begin{cases} -2 \cdot \ln(q) \text{ conosciuto} \Rightarrow \text{si trova la \%} \\ x \text{ conosciuto} \Rightarrow \text{si trova l'estremo} \end{cases}$$

$$q = e^{-d/2}$$

Esempio 12.6 Nell'Esempio 12.3 l'LI al 100q% per μ ha la forma:

$$\bar{X} - \frac{c \cdot \sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + \frac{c \cdot \sigma}{\sqrt{n}} \quad \text{con } c = \sqrt{-2 \cdot \ln(q)}$$

La CP di quest'intervallo in ripetizioni dell'esperimento con $\mu = \mu_0$ è:

$$CP(\mu_0) = P\left(\bar{X} - \frac{c \cdot \sigma}{\sqrt{n}} \leq \mu_0 \leq \bar{X} + \frac{c \cdot \sigma}{\sqrt{n}}; \mu = \mu_0\right) = P\left(-c \leq \frac{\bar{X} - \mu_0}{\sqrt{\sigma^2/n}} \leq c; \mu = \mu_0\right) =$$

La distribuzione di \bar{X} quando $\mu = \mu_0$ è:

$$N\left(\mu_0, \frac{\sigma^2}{n}\right) \Rightarrow Z \equiv \frac{\bar{X} - \mu_0}{\sqrt{\sigma^2/n}} \sim N(0,1)$$

$$= CP(\mu_0) = P(-c \leq Z \leq c) = F(c) - F(-c) = F(c) - [1 - F(c)] = 2F(c) - 1 \quad \text{con } Z \sim N(0,1)$$

$$F(c) = \frac{CP(\mu_0) + 1}{2}$$

Agli stessi risultati si perviene usando:

$$CP(\theta_0) = P(D \leq -2 \cdot \ln(q); \theta = \theta_0) \Rightarrow \{D \sim \chi_{(1)}^2 \quad \forall \mu_0\} \Rightarrow CP(\mu_0) = P(\chi_{(1)}^2 \leq -2 \cdot \ln(q))$$

infatti:

$$P(\chi_{(1)}^2 \leq -2 \cdot \ln(q)) = P(D \leq -2 \cdot \ln(q)) = P(Z^2 \leq -2 \cdot \ln(q)) = P(-c \leq Z \leq c)$$

Esempio 12.7 Supponiamo che un esperimento dia n conteggi che sono modellati come valori osservati di v.a. X_1, X_2, \dots, X_n indipendenti e di Poisson, con valore atteso μ :

$$l(\mu) = t \cdot \ln(\mu) - n\mu \quad \text{con } t = \sum x_i$$

$$\hat{\mu} = \frac{t}{n}$$

$$r(\mu) = l(\mu) - l(\hat{\mu}) = t \cdot \ln\left(\frac{\mu}{\hat{\mu}}\right) - n\mu + n\hat{\mu} = -t \cdot \ln\left(\frac{t}{n\mu}\right) - n\mu + t$$

Immaginiamo ora di ripetere l'esperimento con μ fissato a μ_0 ; il conteggio totale $T \equiv \sum X_i$ diventa una v.a. la cui distribuzione è di Poisson con media $m = n\mu_0$. La D risulta quindi:

$$D \equiv -2r(\mu_0) \equiv 2 \left[T \cdot \ln\left(\frac{T}{m}\right) + m - T \right]$$

In questo caso D dipende dal valore vero μ_0 tramite m , ed è evidente che la sua distribuzione non è quella del chi-quadrato. Vogliamo, tuttavia, vedere se la distribuzione $\chi_{(1)}^2$ approssima bene la distribuzione esatta di D che andiamo a calcolare. La distribuzione di D dipende dalla legge di T che è Poissoniana:

$$P(T = t) = \frac{m^t e^{-t}}{t!} \quad \text{con } t = 0, 1, 2, \dots$$

Fissato per es., $m=10$, si ha

$$D \equiv 2 \left[T \cdot \ln\left(\frac{T}{10}\right) + 10 - T \right]$$

Calcoliamo le probabilità di copertura per gli intervalli di verosimiglianza al 25.8%, 14.7% e 3.6%, corrispondenti nel caso normale a CP pari a 0.9, 0.95, e 0.99, rispettivamente. Dobbiamo considerare gli eventi $(D \leq 2.706)$, $(D \leq 3.841)$ e $(D \leq 6.635)$, cui corrispondono gli eventi $(6 \leq T \leq 15)$, $(5 \leq T \leq 16)$ e $(4 \leq T \leq 19)$, rispettivamente. Usando che

$$P(T = t) = \frac{m^t e^{-t}}{t!} \quad \text{con } t = 0, 1, 2, \dots$$

si ottiene

$$P(D \leq 2.706) = (6 \leq T \leq 15) = \sum_{t=6}^{15} 10^t e^{-10} / t! = 0.884$$

$$P(D \leq 3.841) = (5 \leq T \leq 16) = \sum_{t=5}^{16} 10^t e^{-10} / t! = 0.944$$

$$P(D \leq 6.635) = (4 \leq T \leq 19) = \sum_{t=4}^{19} 10^t e^{-10} / t! = 0.986$$

Intervalli di confidenza o fiducia o CI: è l'intervallo random $[A, B]$ per θ se la relativa CP:

$$CP(\theta_0) = P(A \leq \theta_0 \leq B; \theta = \theta_0)$$

è la stessa per ogni valore del parametro θ_0 . La stessa CP prende il nome di **coefficiente di confidenza**.

Per es., $[A, B]$ è un intervallo di confidenza al 95% per θ se:

$$P(A \leq \theta_0 \leq B; \theta = \theta_0) = 0.95$$

per tutti i valori del parametro θ_0 . Un CI al 95% includerà il valore vero del parametro θ_0 nel 95% delle ripetizioni dell'esperimento con θ fissato.

Nell'esempio 12.6 abbiamo trovato che la CP degli intervalli di verosimiglianza al $100q\%$ non dipende

dal valore vero del parametro. Nell'esempio l'LI al 100q% è un CI. In particolare, il coefficiente di confidenza dell'LI al 14.7% è esattamente 0.95. Gli intervalli di verosimiglianza non sono intervalli di confidenza nell'esempio 12.7, poiché le probabilità di copertura dipendono da θ_0 .

A causa dell'approssimazione χ^2 , gli intervalli di verosimiglianza sono intervalli di confidenza esatti o approssimati in gran parte delle applicazioni. Quando si approssima la D , il coefficiente di confidenza approssimato dell'intervallo di verosimiglianza al 100q% è dato da:

$$CP \approx P(\chi_{(1)}^2 \leq -2 \cdot \ln(q))$$

N.B.: abbiamo detto che la CP può essere interpretata come la frazione di tempo in cui l'intervallo aleatorio $[A, B]$ include il valore vero θ_0 . La stessa considerazione vale per CI. Tuttavia, questa affermazione si riferisce alle ipotetiche ripetizioni dell'esperimento; non è corretto concludere che un particolare osservato intervallo di confidenza $[a, b]$, per es., al 95%, ha il 95% di probabilità di includere il valore vero di θ . Può succedere, infatti, che il particolare intervallo contenga tutti i possibili valori di θ , e quindi copra θ_0 con probabilità 100%.

Stimatore non distorto: xxx

Stimatore non distorto della media $\Rightarrow \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$

Stimatore non distorto della varianza $var_{\theta}(X_i)$ con media nota $\Rightarrow \bar{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2$

Stimatore non distorto della varianza $var_{\theta}(X_i)$ con media non nota $\Rightarrow S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$

CAPITOLO XIII - Test

Test: un test di significatività è una procedura per valutare l'evidenza fornita dai dati contro un'ipotesi H , simile alla prova per assurdo in matematica. Si assume vera l'ipotesi e si controlla se tale assunzione porta ad una inconsistenza. Se una contraddizione viene riscontrata, l'ipotesi è scartata. Se nessuna contraddizione viene riscontrata, il metodo della prova fallisce e l'ipotesi potrebbe essere vera o falsa. A differenza della prova matematica per assurdo, nelle applicazioni statistiche non c'è una inconsistenza logica tra i dati e l'ipotesi. Più piccola è la probabilità, più forte è l'evidenza che l'ipotesi è falsa.

Livello di Significatività o SL o Significante Level: poiché in un test bisogna scegliere tra due possibilità (respingere o no l'ipotesi), una regola di decisione equivale a stabilire quali siano i valori delle osservazioni che conducono al rigetto dell'ipotesi. Chiameremo questo insieme C di valori delle osservazioni **regione critica o di rigetto** del test. In generale, qualunque sia la scelta della regione critica, se l'ipotesi è vera vi è una probabilità positiva di avere un'osservazione in C e quindi di respingere a torto l'ipotesi (errore di prima specie). Allo stesso modo vi è una probabilità positiva di non respingere l'ipotesi falsa (errore di seconda specie). Da notare che i due errori non giocano un ruolo simmetrico: ad esempio, quando si controlla l'efficacia di un farmaco si procede facendo l'ipotesi che esso non sia efficace, e si considera più grave attribuire a torto al farmaco delle proprietà che esso non possiede piuttosto che il contrario. Ricordiamo che l'ipotesi è l'eventualità più sfavorevole che si spera, sulla base delle osservazioni, di respingere.

Per costruire regioni di rigetto, e più in generale, per testare un'ipotesi, cerchiamo una statistica D opportuna, come ad es. $D \equiv |X - 50|$. Un piccolo valore di D indica accordo tra risultati e ipotesi, mentre un valore alto di D indica disaccordo. Una volta eseguito l'esperimento e raccolti i dati, possiamo calcolare il valore osservato di D . Quindi, assumendo H vera, calcoliamo la probabilità di ottenere un valore di D grande almeno quanto quello osservato. Questa probabilità è detta livello di significatività dei dati in relazione all'ipotesi:

$$SL = P(D \geq D_{obs}; H \text{ vera})$$

dove D_{obs} è il valore osservato di D . Se SL è molto piccolo, allora un accordo tanto povero non dovrebbe avverarsi quasi mai se l'ipotesi fosse vera, e si ha quindi evidenza che H è falsa. Più piccolo è il livello di significatività, più grande è l'evidenza contro l'ipotesi. Al contrario, un grande valore di SL indica solo la mancanza di evidenza contro l'ipotesi. Anche un livello del 90% o 100% non implica che l'ipotesi è

“probabilmente vera”. L'affermazione di probabilità si riferisce ai dati, non all'ipotesi. Convenzionalmente, il valore 0.05 è preso come linea di demarcazione tra livello di significatività “piccolo” e “grande”. Se $SL \leq 0.05$ diciamo che l'ipotesi è contraddetta dai dati al livello 5%, mentre se $SL > 0.05$, l'ipotesi è detta consistente o compatibile coi dati, al livello 5%. Naturalmente, questa è una convenzione, da non prendere in modo molto rigido.

Alternativamente, possiamo fissare il livello di significatività, per es., $SL = 0.05$ e determinare la regione di rigetto a questo livello, cioè l'insieme dei valori del parametro che contraddicono l'ipotesi a questo livello.