Stima Parametrica

Stefania Bartoletti

4 Giugno 2021

Indice

Distribuzioni Parametriche

- Consideriamo un campione di dati che consiste in n realizzazioni $x_1, x_2, \ldots, x_n$ di una variabile aleatoria esponenziale.

Distribuzioni Parametriche

- Consideriamo un campione di dati che consiste in n realizzazioni $x_1, x_2, \ldots, x_n$ di una variabile aleatoria esponenziale.
- Sì, ma quale?

Distribuzioni Parametriche

- Consideriamo un campione di dati che consiste in n realizzazioni $x_1, x_2, \ldots, x_n$ di una variabile aleatoria esponenziale.
- Sì, ma quale?
- Ogni volta che ci riferiamo alla distribuzione espoenziale, oppure binomiale o Gaussiana, ci stiamo di fatto riferendo ad una famiglia di distribuzioni

Distribuzioni Parametriche

- Consideriamo un campione di dati che consiste in n realizzazioni $x_1, x_2, \ldots, x_n$ di una variabile aleatoria esponenziale.
- Sì, ma quale?
- Ogni volta che ci riferiamo alla distribuzione espoenziale, oppure binomiale o Gaussiana, ci stiamo di fatto riferendo ad una famiglia di distribuzioni
- Tali distribuzioni infatti dipendono da alcuni parametri (λ; p, n; μ, σ^2) e pertanto vengono dette parametriche

Distribuzioni Parametriche

- Consideriamo un campione di dati che consiste in n realizzazioni $x_1, x_2, \ldots, x_n$ di una variabile aleatoria esponenziale.
- Sì, ma quale?
- Ogni volta che ci riferiamo alla distribuzione espoenziale, oppure binomiale o Gaussiana, ci stiamo di fatto riferendo ad una famiglia di distribuzioni
- Tali distribuzioni infatti dipendono da alcuni parametri (λ; p, n; μ, σ^2) e pertanto vengono dette parametriche
- Per ogni valore di parametro infatti, abbiamo una funzione di distribuzione diversa.

Stima parametrica

- La stima parametrica si riferisce alla stima di un parametro di una distribuzione a partire da un campione di dati

Stima parametrica

- La stima parametrica si riferisce alla stima di un parametro di una distribuzione a partire da un campione di dati
- Gli stimatori puntuali associano alla distribuzione un valore singolo del parametro

Stima parametrica

- La stima parametrica si riferisce alla stima di un parametro di una distribuzione a partire da un campione di dati
- Gli stimatori puntuali associano alla distribuzione un valore singolo del parametro
- Gli stimatori ad intervallo, o intervalli di confidenza, associano alla distribuzione un intervallo di parametri

Stimatori

- Consideriamo

Stimatori

- Consideriamo
  - Un esperimento

Stimatori

- Consideriamo
  - Un esperimento
  - Un campione di dati

Stimatori

- Consideriamo
  - Un esperimento
  - Un campione di dati
  - Di essere interessati alla stima di un parametro della distribuzione dei dati

Stimatori

- Consideriamo
  - Un esperimento
  - Un campione di dati
  - Di essere interessati alla stima di un parametro della distribuzione dei dati
- Uno stimatore è una statistica (quindi una variabile aleatoria) che fornisce la stima di un parametro

Stimatori

- Consideriamo
  - Un esperimento
  - Un campione di dati
  - Di essere interessati alla stima di un parametro della distribuzione dei dati
- Uno stimatore è una statistica (quindi una variabile aleatoria) che fornisce la stima di un parametro
- La stima è un valore deterministico assunto dallo stimatore come risultato dello stimatore

Stimatore di Massima Verosimiglianza (Maximum Likelihood Estimator)

Lo stimatore di massima verosimiglianza risponde alla domanda:
Per quale valore del parametro il campione di dati ha probabilità (verosimiglianza) massima di essere osservato ?

Si tratta di uno stimatore puntuale, che ad un generico parametro θ associa un singolo valore $\hat{\theta}$

Esempio 1

In un esperimento che consiste in 100 lanci di una stessa moneta, otteniamo 55 teste. Qual è la stima di massima verosimiglianza per p ?

Esempio 1

In un esperimento che consiste in 100 lanci di una stessa moneta, otteniamo 55 teste. Qual è la stima di massima verosimiglianza per p ?

- Per un dato valorie di p, la probabilità di ottenere 55 teste in questo esperimento sarebbe

$$\mathbb{P}\{55 \text{ teste }\} = \binom{100}{55} p^{55}(1-p)^{45}$$

- Evidenziamo che si tratta di una funzione di p definendo la funzione di verosimiglianza come

$$\mathbb{P}\{55 \text{ teste }; p\} = \binom{100}{55} p^{55}(1-p)^{45}$$

- Cerchiamo p che massimizzare tale funzione di verosimiglianza

Definizione

Considerato un campione di dati $x_1, x_2, \ldots, x_n$, la stima di massima verosimigliannza di un parametro θ è il valore $\hat{\theta}$ che massimizza la funzione di *likelihood* $\mathbb{P}\left\{x_1, x_2, \ldots, x_n; \theta\right\}$. Ovvero

$$\hat{\theta} = \arg\max_{\theta} \mathbb{P}\left\{x_1, x_2, \ldots, x_n; \theta\right\}$$

Definizione

Considerato un campione di dati $x_1, x_2, \ldots, x_n$ estratti da n variabili aleatorie $\mathsf{X}_1, \mathsf{X}_2, \ldots, \mathsf{X}_n$, lo stimatore di massima verosimiglianza di un parametro θ è il valore $\hat{\theta}$ che massimizza la funzione di *likelihood* $\Lambda(\theta) = \mathbb{P}\left\{\mathsf{X}_1, \mathsf{X}_2, \ldots, \mathsf{X}_n; \theta\right\}$. Ovvero

$$\hat{\theta} = \arg\max_{\theta} \Lambda(\theta) = \arg\max_{\theta} \mathbb{P}\left\{\mathsf{X}_1, \mathsf{X}_2, \ldots, \mathsf{X}_n; \theta\right\}$$

Log-likelihood

Spesso ai fini del calcolo è più semplice lavorare conn il logaritmo naturale della funzione di likelihood. In questo caso parliamo di *log-likelihood*. Dal momento che $\ln(\cdot)$ è una funzione crescente, il massimo della likelihood e quello della log-likelihood coincidono.

$$\hat{\theta} = \arg\max_{\theta} \ \mathbb{P}\{X_1, X_2, \ldots, X_n; \theta\} = \arg\max_{\theta} \ \ln(\mathbb{P}\{X_1, X_2, \ldots, X_n; \theta\})$$

MLE per distribuzioni discrete

- Consideriamo delle variabili aleatorie discrete $X_1, x_2, \ldots, X_n$ la cui distribuzione congiunta sia nota a meno di un parametro incognito θ

MLE per distribuzioni discrete

- Consideriamo delle variabili aleatorie discrete $X_1, x_2, \ldots, X_n$ la cui distribuzione congiunta sia nota a meno di un parametro incognito θ
- Definiamo la funzione di *likelihood* $\Lambda(\theta) = f_{X_1, X_2, \ldots, X_n}(X_1, X_2, \ldots, X_n; \theta)$

MLE per distribuzioni discrete

- Consideriamo delle variabili aleatorie discrete $X_1, x_2, \ldots, X_n$ la cui distribuzione congiunta sia nota a meno di un parametro incognito θ
- Definiamo la funzione di *likelihood* $\Lambda(\theta) = f_{X_1, X_2, \ldots, X_n}(X_1, X_2, \ldots, X_n; \theta)$
- Lo stimatore di massima verosimiglianza è la statistica $\hat{\theta}$, ovvero una variabile aleatoria funzione del campione, che massimizza la funzione di likelihood

$$\begin{aligned}\hat{\theta} &= \arg\max_{\theta} \Lambda(\theta) \\ &= \arg\max_{\theta} f_{X_1, X_2, \ldots, X_n}(X_1, X_2, \ldots, X_n; \theta)\end{aligned}$$

MLE per distribuzioni continue

- Consideriamo delle variabili aleatorie continue $X_1, x_2, \dots, X_n$ la cui distribuzione congiunta sia nota a meno di un parametro incognito θ

MLE per distribuzioni continue

- Consideriamo delle variabili aleatorie continue $X_1, x_2, \ldots, X_n$ la cui distribuzione congiunta sia nota a meno di un parametro incognito θ
- Definiamo la funzione di *likelihood* $\Lambda(\theta) = f_{X_1, X_2, \ldots, X_n}(X_1, X_2, \ldots, X_n; \theta)$

MLE per distribuzioni continue

- Consideriamo delle variabili aleatorie continue $\mathsf{X}_1, \mathsf{x}_2, \ldots, \mathsf{X}_n$ la cui distribuzione congiunta sia nota a meno di un parametro incognito θ
- Definiamo la funzione di *likelihood* $\Lambda(\theta) = f_{\mathsf{X}_1,\mathsf{X}_2,\ldots,\mathsf{X}_n}(\mathsf{X}_1, \mathsf{X}_2, \ldots, \mathsf{X}_n; \theta)$
- Lo stimatore di massima verosimiglianza è la statistica $\hat{\uptheta}$, ovvero una variabile aleatoria funzione del campione, che massimizza la funzione di likelihood

$$\begin{aligned}\hat{\uptheta} &= \arg\max_{\theta} \Lambda(\theta) \\ &= \arg\max_{\theta} f_{\mathsf{X}_1,\mathsf{X}_2,\ldots,\mathsf{X}_n}(\mathsf{X}_1, \mathsf{X}_2, \ldots, \mathsf{X}_n; \theta)\end{aligned}$$

MLE per distribuzioni continue

Nel caso di distribuzioni continue, il calcolo della likelihood si ottiene tramite le pdf (vedremo tra poco la definizione formale).

MLE per distribuzioni continue

Nel caso di distribuzioni continue, il calcolo della likelihood si ottiene tramite le pdf (vedremo tra poco la definizione formale).

Esempio 2

- Immaginiamo che il tempo di vita di alcune lampadine segua una distribuzione esponenziale, di cui non conosciamo il parametro λ

MLE per distribuzioni continue

Nel caso di distribuzioni continue, il calcolo della likelihood si ottiene tramite le pdf (vedremo tra poco la definizione formale).

Esempio 2

- Immaginiamo che il tempo di vita di alcune lampadine segua una distribuzione esponenziale, di cui non conosciamo il parametro λ
- Tuttavia, collezioniamo un insieme di valori che corrispondono al tempo di vita di 5 lampadine

MLE per distribuzioni continue

Nel caso di distribuzioni continue, il calcolo della likelihood si ottiene tramite le pdf (vedremo tra poco la definizione formale).

Esempio 2

- Immaginiamo che il tempo di vita di alcune lampadine segua una distribuzione esponenziale, di cui non conosciamo il parametro λ
- Tuttavia, collezioniamo un insieme di valori che corrispondono al tempo di vita di 5 lampadine
- Il vettore di dati è X $= [2, 3, 1, 3, 4]$

MLE per distribuzioni continue

Nel caso di distribuzioni continue, il calcolo della likelihood si ottiene tramite le pdf (vedremo tra poco la definizione formale).

Esempio 2

- Immaginiamo che il tempo di vita di alcune lampadine segua una distribuzione esponenziale, di cui non conosciamo il parametro λ
- Tuttavia, collezioniamo un insieme di valori che corrispondono al tempo di vita di 5 lampadine
- Il vettore di dati è $\mathsf{X} = [2, 3, 1, 3, 4]$
- Consideriamo il tempo di vita di ogni lampadina come indipendente da quello delle altre

MLE per distribuzioni continue

Nel caso di distribuzioni continue, il calcolo della likelihood si ottiene tramite le pdf (vedremo tra poco la definizione formale).

Esempio 2

- Immaginiamo che il tempo di vita di alcune lampadine segua una distribuzione esponenziale, di cui non conosciamo il parametro λ
- Tuttavia, collezioniamo un insieme di valori che corrispondono al tempo di vita di 5 lampadine
- Il vettore di dati è $\mathsf{X} = [2, 3, 1, 3, 4]$
- Consideriamo il tempo di vita di ogni lampadina come indipendente da quello delle altre

MLE per distribuzioni continue

Nel caso di distribuzioni continue, il calcolo della likelihood si ottiene tramite le pdf (vedremo tra poco la definizione formale).

Esempio 2

- Immaginiamo che il tempo di vita di alcune lampadine segua una distribuzione esponenziale, di cui non conosciamo il parametro λ
- Tuttavia, collezioniamo un insieme di valori che corrispondono al tempo di vita di 5 lampadine
- Il vettore di dati è X $= [2, 3, 1, 3, 4]$
- Consideriamo il tempo di vita di ogni lampadina come indipendente da quello delle altre

Nota: La scelta di utilizzare le pdf per MLE si basa sempre sulla scelta di un intervallo molto piccolo in cui calcolare la probabilità.

MLE per una distribuzione Gaussiana

- Siano $\mathsf{X}_1, \mathsf{X}_2, \ldots, \mathsf{X}_n$ variabili aleatorie normali e indipendenti con media μ e deviazione standard σ, entrambe incognite
- La densità congiunta della likelihood è $f_{\mathsf{X}}(x_1, x_2, \ldots, x_n; \mu, \sigma)$
- Essendo variabili aleatorie indipendenti

$$f_{\mathsf{X}}(x_1, x_2, \ldots, x_n; \mu, \sigma) = \prod_{i=1}^{n} f_{\mathsf{X}_i}(x_i; \mu, \sigma)$$

- Considero la log-likelihood e ottengo

$$\ln(f_{\mathsf{X}}(x_1, x_2, \ldots, x_n; \mu, \sigma) =$$
$$-\frac{n}{2}\log(2\pi) - n\log\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2$$

- Procedo derivando sia rispetto a σ che rispetto a μ e ponendo a 0 le due derivate.

MLE per una distribuzione Gaussiana

Risolvendo il sistema di due equazioni in due incognite, trovo che gli stimatori per μ e σ sono

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$\hat{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \hat{\mu})^2}$$

Nota: lo stimatore per σ non coincide con la deviazione standard campionaria, che invece ha $n-1$ al denominatore. Il motivo è che $\mathbb{E}\left\{\hat{\sigma}^2\right\} = \frac{n-1}{n}\sigma^2$, e in questo caso si parla di stimatore non corretto

MLE per una distribuzione Uniforme

- Sia $x_1, x_2, \ldots, x_n$ un campione proveniente da una distribuzione uniforme sull'intervallo $[0, \theta]$, con θ incognita.
- La funzione di likelihood è

$$\Lambda(\theta) = \prod_{i=1}^{n} f_{\mathsf{x_i}}(x_i; \theta) = \begin{cases} \frac{1}{\theta^n} & x_i \in [0, \theta] \forall i \\ 0 & \text{altrimenti} \end{cases}$$

- La likelihood assume valore massimo per θ più piccolo possibile
- Abbiamo un vincolo: θ deve essere comunque maggiore di tutti i valori osservati per la definizione stessa di θ
- Segue che $\hat{\theta} = \max_i x_i$ con $x_i \leqslant \theta$
- Lo stimatore della media è $\left(\max_{i, x_i \leqslant \theta} x_i\right)/2$

Indice

Intervalli di Confidenza

- Come abbiamo visto e sottolineato, uno stimatore puntuale è una statistica, una variabile aleatoria.

Intervalli di Confidenza

- Come abbiamo visto e sottolineato, uno stimatore puntuale è una statistica, una variabile aleatoria.
- La stima del parametro difficilmente sarà esattamente uguale al parametro stesso, soprattutto quando si tratta di una variabile aleatoria continua.

Intervalli di Confidenza

- Come abbiamo visto e sottolineato, uno stimatore puntuale è una statistica, una variabile aleatoria.
- La stima del parametro difficilmente sarà esattamente uguale al parametro stesso, soprattutto quando si tratta di una variabile aleatoria continua.
- Può essere utile dunque definire un intervallo di valori che il parametro può assumere.

Intervalli di Confidenza

- Come abbiamo visto e sottolineato, uno stimatore puntuale è una statistica, una variabile aleatoria.
- La stima del parametro difficilmente sarà esattamente uguale al parametro stesso, soprattutto quando si tratta di una variabile aleatoria continua.
- Può essere utile dunque definire un intervallo di valori che il parametro può assumere.
- Abbiamo definito la credibilità della stima puntuale come verosimiglianza. Nel caso di un intervallo, analogamente, dobbiamo decidere con quale credibilità il nostro parametro appartiene o meno a tale intervallo.

Intervalli di Confidenza

- Come abbiamo visto e sottolineato, uno stimatore puntuale è una statistica, una variabile aleatoria.
- La stima del parametro difficilmente sarà esattamente uguale al parametro stesso, soprattutto quando si tratta di una variabile aleatoria continua.
- Può essere utile dunque definire un intervallo di valori che il parametro può assumere.
- Abbiamo definito la credibilità della stima puntuale come verosimiglianza. Nel caso di un intervallo, analogamente, dobbiamo decidere con quale credibilità il nostro parametro appartiene o meno a tale intervallo.
- Pertanto parliamo di intervalli di confidenza.

Intervalli di confidenza

- Informalmente, un intervallo di confidenza per θ è un intervallo di valori (l, u) calcolato a partire dai dati e che contiene il parametro θ con una certo livello di confidenza.

Intervalli di confidenza

- Informalmente, un intervallo di confidenza per θ è un intervallo di valori (l, u) calcolato a partire dai dati e che contiene il parametro θ con una certo livello di confidenza.
- Gli estremi di tale intervallo variano a seconda del campione dei dati che viene considerato. La randomicità del campione è riflessa sugli estremi dell'intervallo.

Intervalli di confidenza

- Informalmente, un intervallo di confidenza per θ è un intervallo di valori (l, u) calcolato a partire dai dati e che contiene il parametro θ con una certo livello di confidenza.
- Gli estremi di tale intervallo variano a seconda del campione dei dati che viene considerato. La randomicità del campione è riflessa sugli estremi dell'intervallo.
- Variando i dati, variamo l'intervallo. Quando diremo che θ è compreso nell'intervallo con il 95% di confidenza, intenderemo che che il calcolo di tale intervallo, basato su diversi campioni, nel 95% produce un intervallo che contiene il valore vero di θ

Intervalli di confidenza

- Informalmente, un intervallo di confidenza per θ è un intervallo di valori (l, u) calcolato a partire dai dati e che contiene il parametro θ con una certo livello di confidenza.
- Gli estremi di tale intervallo variano a seconda del campione dei dati che viene considerato. La randomicità del campione è riflessa sugli estremi dell'intervallo.
- Variando i dati, variamo l'intervallo. Quando diremo che θ è compreso nell'intervallo con il 95% di confidenza, intenderemo che che il calcolo di tale intervallo, basato su diversi campioni, nel 95% produce un intervallo che contiene il valore vero di θ
- Non è invece appropriato dire che θ appartiene all'intervallo con il 95% di probabilità.

Intervalli di confidenza

- Prima di osservare i dati possiamo dire che l'intervallo che otterremo conterrà θ con il 95% di probabilità.

Intervalli di confidenza

- Prima di osservare i dati possiamo dire che l'intervallo che otterremo conterrà θ con il 95% di probabilità.
- Dopo aver osservato i dati, possiamo dire che l'intervallo trovato contiene θ col 95% di confidenza.

Intervalli di confidenza

- Prima di osservare i dati possiamo dire che l'intervallo che otterremo conterrà θ con il 95% di probabilità.
- Dopo aver osservato i dati, possiamo dire che l'intervallo trovato contiene θ col 95% di confidenza.
- Sì ma come si calcolano questi intervalli?

Intervalli di confidenza

- Prima di osservare i dati possiamo dire che l'intervallo che otterremo conterrà θ con il 95% di probabilità.
- Dopo aver osservato i dati, possiamo dire che l'intervallo trovato contiene θ col 95% di confidenza.
- Sì ma come si calcolano questi intervalli?
- Naturalmente, a partire dalla distribuzione dello stimatore puntuale.

Esempio

Consideriamo un campione $x_1, x_2, \ldots x_n$ estratto da n variabili aleatorie normali indipendenti e identicamente distribuite $\mathsf{X}_i \sim \mathcal{N}(\mu, \sigma^2)$.

- Partiamo col costruire un intervallo di confidenza per μ, sapendo che il suo stimatore puntuale è $\overline{\mathsf{X}}_n \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$

Esempio

Consideriamo un campione $x_1, x_2, \dots x_n$ estratto da n variabili aleatorie normali indipendenti e identicamente distribuite $\mathsf{X}_i \sim \mathcal{N}(\mu, \sigma^2)$.

- Partiamo col costruire un intervallo di confidenza per μ, sapendo che il suo stimatore puntuale è $\overline{\mathsf{X}}_n \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$
- Sappiamo anche che $\frac{\overline{\mathsf{X}}_n - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)$

Esempio

Consideriamo un campione $x_1, x_2, \ldots x_n$ estratto da n variabili aleatorie normali indipendenti e identicamente distribuite $\mathsf{X}_i \sim \mathcal{N}(\mu, \sigma^2)$.

- Partiamo col costruire un intervallo di confidenza per μ, sapendo che il suo stimatore puntuale è $\overline{\mathsf{X}}_n \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$
- Sappiamo anche che $\frac{\overline{\mathsf{X}}_n - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)$
- Segue che $\mathbb{P}\left\{-1.96 \leqslant \frac{\overline{\mathsf{X}}_n - \mu}{\sigma/\sqrt{n}} \leqslant 1.96\right\} = 0.95$

Esempio

Consideriamo un campione $x_1, x_2, \ldots x_n$ estratto da n variabili aleatorie normali indipendenti e identicamente distribuite $\mathsf{X}_i \sim \mathcal{N}(\mu, \sigma^2)$.

- Partiamo col costruire un intervallo di confidenza per μ, sapendo che il suo stimatore puntuale è $\overline{\mathsf{X}}_n \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$
- Sappiamo anche che $\frac{\overline{\mathsf{X}}_n - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$
- Segue che $\mathbb{P}\left\{-1.96 \leqslant \frac{\overline{\mathsf{X}}_n - \mu}{\sigma/\sqrt{n}} \leqslant 1.96\right\} = 0.95$
- Da cui $\mathbb{P}\left\{1.96\sigma/\sqrt{n} + \overline{\mathsf{X}}_n \geqslant \mu \geqslant \overline{\mathsf{X}}_n - 1.96\sigma/\sqrt{n}\right\} = 0.95$

Esempio

- Il 95% delle volte μ sarà a una distanza non superiore a $1.96\sigma/\sqrt{n}$ dalla media aritmetica dei dati

$$\mathbb{P}\left\{\overline{\mathsf{X}}_n - 1.96\sigma/\sqrt{n} \leqslant \mu \leqslant \overline{\mathsf{X}}_n + 1.96\sigma/\sqrt{n}\right\} = 0.95$$

- Per un dato campione, μ sarà all'interno dell'intervallo con il 95% di confidenza
- L'intervallo $(\overline{\mathsf{X}}_n - 1.96\sigma/\sqrt{n}, \overline{\mathsf{X}}_n + 1.96\sigma/\sqrt{n})$ è l'intervallo di confidenza al 95% per μ

Intervalli di confidenza unilaterali

Alcune volte può essere utile considerare intervalli unilaterali. Ad esempio, μ è superiore ad l con il 95% di confidenza significa che μ appartiene ad un intervallo unilaterale $(l, +\infty)$ con il 95% di confidenza.

$$\begin{aligned} 0.95 &= \mathbb{P}\left\{\mathsf{Z} \leqslant 1.645\right\} \\ &= \mathbb{P}\left\{\frac{\overline{\mathsf{X}}_n - \mu}{\sigma/\sqrt{n}} \leqslant 1.645\right\} \\ &= \mathbb{P}\left\{\overline{\mathsf{X}}_n - 1.645\sigma/\sqrt{n} \leqslant \mu\right\} \end{aligned}$$

L'intervallo unilaterale destro al 95% per μ è $(\overline{\mathsf{X}}_n - 1.645\sigma/\sqrt{n}, \infty)$

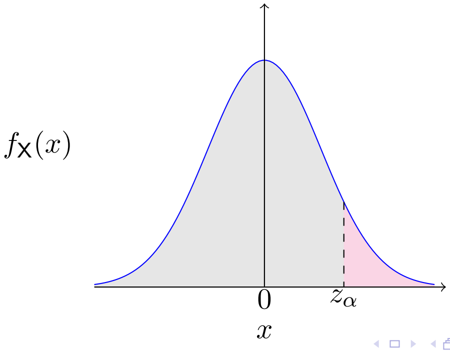
Generalizzazione per ogni livello di confidenza

Sia $\mathsf{Z} \sim \mathcal{N}(0,1)$ e $\alpha \in [0,1]$. Definiamo z_α il valore tale che

$$\mathbb{P}\{\mathsf{Z} \geqslant z_\alpha\} = \alpha$$

Allora $\mathbb{P}\{\mathsf{Z} \geqslant z_\alpha\} = 1 - \Phi(z_\alpha) = \alpha$. Possiamo dunque ottenere z_α come

$$z_\alpha = \Phi^{-1}(1-\alpha)$$



Quantile Gaussiano

Possiamo definire il $k$simo quantile gaussiano come il valore z_α tale che

$$\Phi(z_\alpha) = k/100$$

con $k = (1 - \alpha)100$.
Ad esempio per $\alpha = 0.05$, il calcolo di $z_{0.05}$ si calcola tramite qnorm con argomento 0.95, $\mathrm{qnorm}(0.95) = 1.645$.

Intervalli di confidenza unilaterali

A questo punto siamo in grado di generalizzare immediatamente l'intervallo unilaterale considerando che

$$\mathbb{P}\left\{\mathsf{Z} > z_\alpha\right\} = \alpha$$

e dunque gli intervalli unilaterali con un livello $1-\alpha$ per μ corrispondono a

$$\left(\overline{x}_n - z_\alpha \frac{\sigma}{\sqrt{n}}, \infty\right) \qquad \left(-\infty, \overline{x}_n + z_\alpha \frac{\sigma}{\sqrt{n}}\right)$$

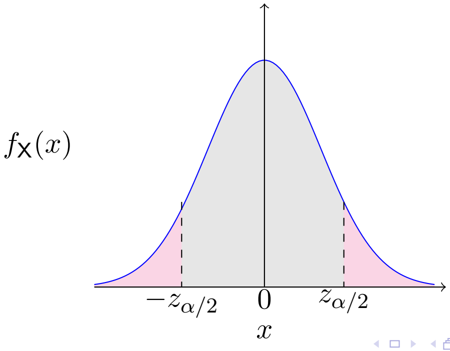
Intervalli di confidenza bilaterali

Considerando inoltre che

$$\mathbb{P}\left\{-z_{\alpha/2} \leqslant \mathsf{Z}_n \leqslant z_{\alpha/2}\right\} = 1 - \alpha$$

Si deduce che l'intervallo unilaterale di livello $1 - \alpha$ per μ corrisponde a

$$\left(\overline{x}_n - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{x}_n + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$



Distribuzione t di Student

Ricordiamo la definizione di variabile aleatoria chi-quadro con k gradi di libertà $\mathsf{c}_k \sim \chi^2_k$, che descrive la somma dei quadrati k variabili aleatorie normali standard e indipendenti $\mathsf{X}_i \sim \mathcal{N}(0,1)$

$$\mathsf{c}_k = \sum_{i=1}^{k} \mathsf{X}_i^2 \,.$$

Distribuzione t di Student

Ricordiamo la definizione di variabile aleatoria chi-quadro con k gradi di libertà $\mathsf{c}_k \sim \chi^2_k$, che descrive la somma dei quadrati k variabili aleatorie normali standard e indipendenti $\mathsf{X}_i \sim \mathcal{N}(0,1)$

$$\mathsf{c}_k = \sum_{i=1}^{k} \mathsf{X}_i^2 \,.$$

Se $\mathsf{Z} \sim \mathcal{N}(0,1)$ e $\mathsf{c}_k \sim \chi^2_k$ sono variabili aleatorie indipendenti, la prima normale standard e la seconda chi-quadro con k gradi di libertà, allorala variabile aleatoria

$$\mathsf{t}_n = \frac{\mathsf{Z}}{\sqrt{\mathsf{c}_k/k}}$$

ha distribuzione t con k gradi di libertà ovvero $\mathsf{t}_k \sim \mathrm{t}_k$, anche detta t di Student con k gradi di libertà.

Distribuzione t di Student

Come nel caso delle variabili aleatorie normali, anche in questo caso possiamo definire i quantili tramite

$$\mathbb{P}\left\{\mathsf{t}_n > t_{\alpha,n}\right\} = \alpha$$

e per simmetria

$$\alpha = \mathbb{P}\left\{-\mathsf{t}_n > t_{\alpha,n}\right\} = \mathbb{P}\left\{\mathsf{t}_n \leqslant -t_{\alpha,n}\right\} = 1 - \mathbb{P}\left\{\mathsf{t}_n > -t_{\alpha,n}\right\}$$

Segue che $\mathbb{P}\left\{-\mathsf{t}_n > t_{\alpha,n}\right\} = 1 - \alpha$

Distribuzione t di Student

Si tratta di una distribuzione importante in quanto si dimostra che

$$\frac{\overline{x}_n - \mu}{s_n/\sqrt{n}} \sim \mathrm{t}_{n-1}$$

dove $\overline{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$ è la media campionaria di n variabili aleatorie normali e indipendenti con media μ e per deviazione standard campionaria $s_n = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2}$.

Intervallo di confidenza per σ incognito

Analogamente a ciò che accade nel caso con σ noto,

$$\mathbb{P}\left\{-t_{\alpha/2} \leqslant \mathsf{t}_{n-1} \leqslant t_{\alpha/2}\right\} = 1 - \alpha$$

Si deduce che l'intervallo unilaterale di livello $1 - \alpha$ per μ corrisponde a

$$\left(\overline{x}_n - t_{\alpha/2}\frac{s_n}{\sqrt{n}}, \overline{x}_n + t_{\alpha/2}\frac{s_n}{\sqrt{n}}\right)$$

Intervallo di confidenza per la differenza di due medie

Consideriamo due campioni indipendenti estratti da due popolazioni diverse con valori attesi μ_1 e μ_2 e deviazioni standard σ_1 e σ_2.
La MLE $\boldsymbol{\mu}_{\mathsf{d}}$ per $\mu_{\mathsf{d}} = \mu_1 - \mu_2$ è data da $\boldsymbol{\mu}_{\mathsf{d}} = \overline{x} - \overline{y}$ Per ottenere un intervallo di confidenza possiamo considerare che

$$\frac{\overline{x} - \overline{y} - (\mu_1 - \mu_2)}{\sqrt{(\sigma_1^2/n + \sigma_2^2/m)}} \sim \mathcal{N}(0,1)$$

Da cui

$$\mu_1 - \mu_2 \in \left(\overline{x} - \overline{y} - z_{\alpha/2}\sqrt{(\sigma_1^2/n + \sigma_2^2/m)},\right.$$
$$\left.\overline{x} - \overline{y} + z_{\alpha/2}\sqrt{(\sigma_1^2/n + \sigma_2^2/m)}\right)$$

Intervallo di confidenza per il parametro di una Bernoulli

Consideriamo una popolazione in cui ogni elemento possiede una data caratteristica con probabilità p.
Testiamo n oggetti e vogliamo stimare p. Di fatto, la frequenza assoluta di oggetti che possiede tale caratteristica è X, una binomiale con parametri n e p. Considerando n abbastanza elevato, possiamo approssimare

$$\frac{\mathsf{X} - np}{\sqrt{np(1-p)}} \sim \mathcal{N}(0,1)$$

Se costruissimo l'intervallo di confidenza seguendo l'esempio della media per il caso Gaussiano, otterremmo

$$\mathbb{P}\left\{-z_{\alpha/2} \leqslant \frac{\mathsf{X} - np}{\sqrt{np(1-p)}} \leqslant z_{\alpha/2}\right\} = 1 - \alpha$$

Tuttavia resta molto difficile esplicitare p. Pertanto, si approssima il denominatore con una sua stima MLE, i.e. $\hat{\mathsf{p}} = \mathsf{X}/n$. Se ne ricava

$$\mathbb{P}\left\{-z_{\alpha/2} \leqslant \frac{\mathsf{X} - np}{\sqrt{n\hat{\mathsf{p}}(1-\hat{\mathsf{p}})}} \leqslant z_{\alpha/2}\right\} = 1 - \alpha$$

Pertanto

$$\mathbb{P}\left\{\mathsf{X} - z_{\alpha/2}\sqrt{n\hat{\mathsf{p}}(1-\hat{\mathsf{p}})} \leqslant p \leqslant \mathsf{X} + z_{\alpha/2}\sqrt{n\hat{\mathsf{p}}(1-\hat{\mathsf{p}})}\right\} = 1 - \alpha$$

$$\mathbb{P}\left\{\hat{\mathsf{p}} - z_{\alpha/2}\sqrt{\hat{\mathsf{p}}(1-\hat{\mathsf{p}})/n} \leqslant p \leqslant \hat{\mathsf{p}} + z_{\alpha/2}\sqrt{n\hat{\mathsf{p}}(1-\hat{\mathsf{p}})/n}\right\} = 1 - \alpha$$