Modelli Statistici - Sommario

X

Primer sulla statistica descrittiva e inferenziale.

X

A. STATISTICA DESCRITTIVA

A1. Nozioni di Statistica Descrittiva

Nozioni di Statistica Descrittiva

X

Presentazione e organizzazione dei dati. Nozioni fondamentali della statistica descrittiva: unità statistica, popolazione, campione, caratteri dei dati. Frequenza assoluta e relativa.

X

1. Presentazione dei Dati

Presentiamo delle nozioni sulla *statistica descrittiva*, utili per *raccogliere* e *organizzare* i *dati*. In particolare vedremo come *presentare* i dati.

#Definizione

Definizione (unità statistica, popolazione, campione e caratteri dei dati).

Le proprietà salienti dei dati vengono riassunte dai seguenti indicatori sisntetici:

Si dice l'*unità statistica* come la minima unità della quale si raccolgono i dati

Si dice la *popolazione* come l'insieme delle unità statistiche

Si dice *campione* una porzione della popolazione

Si dicono *caratteri* le proprietà che sono oggetto di rilevazione dei dati. Possono essere o *qualitativi* o *quantitativi*. I caratteri qualitativi vengono indicati attraverso le *espressioni verbali* e sono sempre suddivisi in *categorie*; quelli quantitativi con *numeri*, in particolare interi.

2. Organizzazione dei Dati

Adesso vedremo come si *organizza* i dati. In particolare ci focalizziamo sulla *tabella di distribuzione di frequenza*, ovvero l'idea è quella di dividere l'insieme dei dati in *classi* e di misurare ogni insieme.

#Definizione

Definizione (frequenza assoluta, relativa, percentuale e assoluta cumulativa).

Supponiamo di avere n dati suddivisi in N classi.

Allora si dice la *frequenza assoluta* come la misura di una classe k; la indichiamo con $f_a(k)$. Ovviamente si ha $\sum_k^N f_a(k) = n$.

Normalizzando la frequenza assoluta di una classe k per il quantitativo totale n, si ha la *frequenza relativa*: ovvero $f_r(k) := \frac{f_a(k)}{n}$. Come osservato prima, si ha $\sum_k^N f_r(k) = \sum_k^N \frac{f_a(k)}{n} = 1$.

Dopodiché moltiplicando la *frequenza relativa* per 100 si ottiene la quantità $f_p(k) := 100 \cdot f_r(k)$.

Infine si definisce la *frequenza assoluta cumulativa* $F_a(k)$ come la *misura totale* che ricadono nelle classi fino alla k-esima compresa. Ovvero

$$F_a(k) := \sum_{j=1}^{k} f_a(j)$$

Ovviamente si osserva che F_a è non-decrescente e $F(N) = n$.

#Osservazione

Osservazione (come si creano le classi?).

Adesso una domanda spontanea è quella di *chiedersi come si creano queste classi*.

Nel caso qualitativo, le *classi* coincidono con le *categorie*: ovvero ad ogni scatola assegniamo un numero.

Se invece il carattere è di tipo *quantitativo*, dobbiamo vedere più casi.

Se ho *numeri discreti* (ovvero in $\mathbb{Z}$), allora è sufficiente prendere $x_1, \ldots, x_n, \ldots$ i valori assumibili e dedicare ad ogni classe un insieme A_k: ovvero $A_k := \{\text{unità con valore } x_k\}$.

Se ho invece *numeri continui*, devo suddividere tutto in *intervalli*: di solito si prende il primo intervallo come chiuso da entrambi i lati (ovvero del tipo $[a, b]$) poi per prendere intervalli aperti-chiusi (ovvero del tipo $(a, b]$).
Una problematica è quella di scegliere bene l'*ampiezza* dell'intervallo, in quanto se ho un'ampiezza troppo piccola, i dati diventano *difficilmente leggibili*: se invece ho un'ampiezza troppo grande, allora perdo delle *informazioni* (ovvero diventa tutto troppo riassuntivo).

A2. Rappresentazione Grafica dei Dati

Rappresentazione Grafica dei Dati nella Statistica Descrittiva

X

Metodi di rappresentazione grafica dei dati: tabella delle frequenze, istogramma e ogiva.

X

0. Voci correlate

- Nozioni di Statistica Descrittiva

1. Tabella delle Frequenze

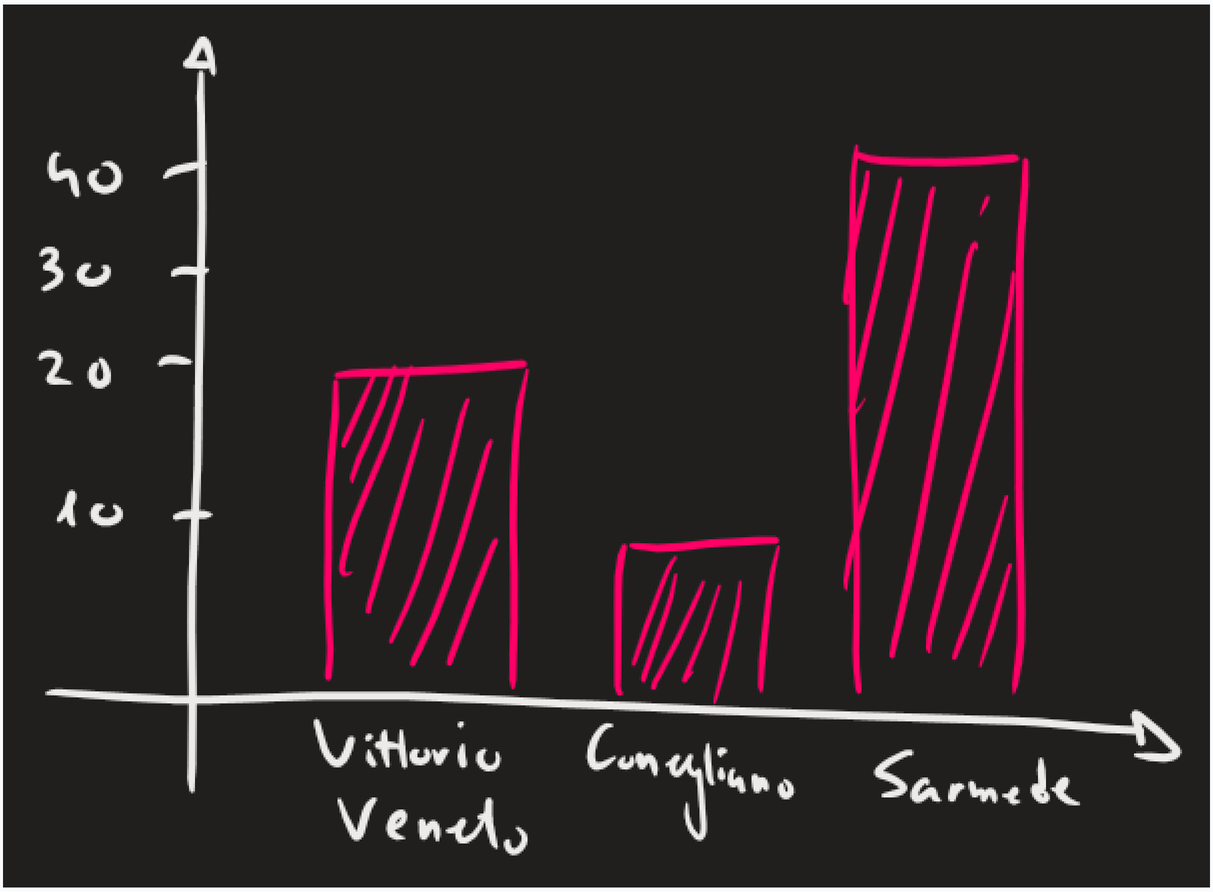
Adesso vediamo come *rappresentare in una maniera grafica* i dati. Partiamo dal modo più semplice: la tabella delle frequenze. In questo caso si tratta banalmente di mettere in una tabella le *classi* e le loro *frequenze relative* associate. Ad esempio, si ha

#Esempio

Voto	Frequenza relativa
18	5
19	6
20	1
21	0
22	0
...	...
30	2

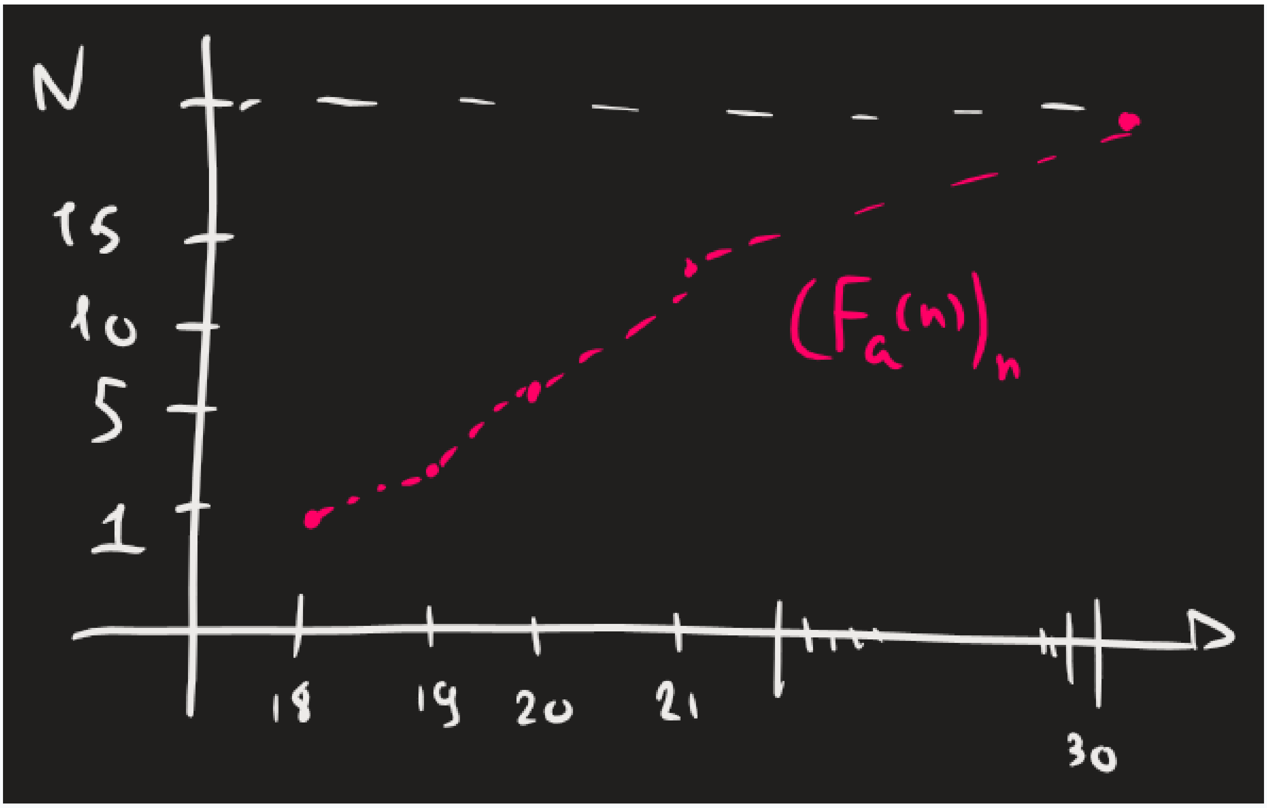
2. Istogramma

Alternativamente possono essere rappresentati attraverso un *istogramma*, una sorta di *grafico* costituito da *rettangoli*, in cui la *base* rappresenta la classe e l'*altezza* la *frequenza relativa* (o anche assoluta, è una scelta arbitraria)



3. Ogiva

Nel caso *quantitativo*, la *frequenza cumulativa* F_a può essere rappresentato con un vero e proprio *grafico* (o successione, dipende se ho numeri discreti o continui). Il grafico si dice *ogiva*: in ordinata si riportano le *frequenza cumulative*, nell'ascisse si pongono i *valori osservati* (o gli estremi degli intervalli).



A3. Indicatori Sintetici

Indici Sintetici per la Statistica Descrittiva

X

Indici sintetici per la statistica descrittiva: moda, mediana, media campionaria e varianza. Problema: mediana per dati raggruppati. Problema: media e varianza campionaria per dati raggruppati.

X

0. Voci correlate

- Nozioni di Statistica Descrittiva
- Rappresentazione Grafica dei Dati nella Statistica Descrittiva

1. Definizione di Moda, Mediana, Media Campionaria

Per riassumere i dati usiamo degli *indicatori sintetici*, che forniscono un'idea di massima della situazione. Adesso vedremo il *"dove"* (ovvero *dove* si stabilizzano dei dati): vedremo la *moda*, *mediana* e *media campionaria*.

#Definizione

Definizione (moda, mediana, media campionaria).

Si dice la *moda* come la *categoria* (classe) che ha la *massima frequenza*.

Si dice la *mediana* come il *valore* che occupa il *posto di mezzo*, ovvero disponendo i dati in ordine *crescente* e sono in numero dispari (altrimenti prendiamo la media aritmetica dei due valori centrali).

Si dice la *media campionaria* come la *media aritmetica dei dati*, ovvero $\bar{x} = \frac{1}{n} \sum_n x_n$.

2. Problemi di Posizione per Dati Raggruppati

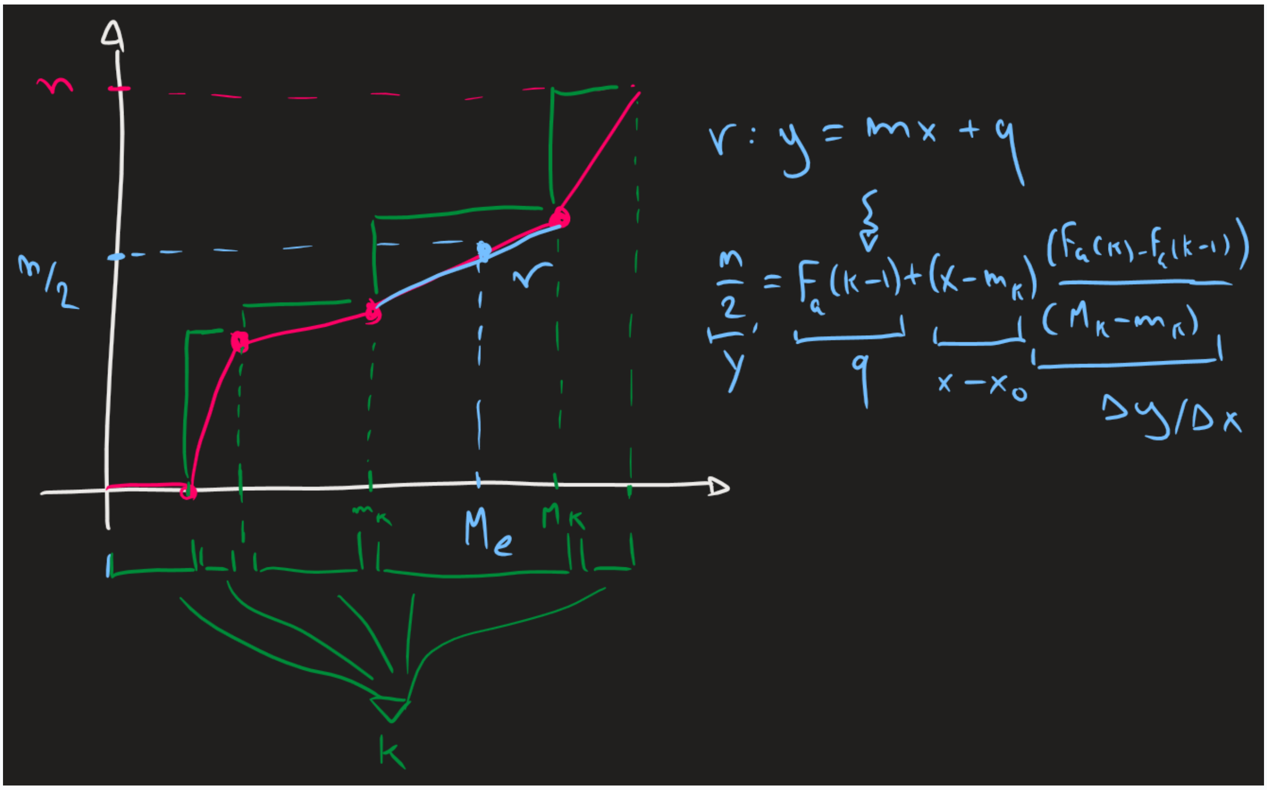
#Osservazione

PROBLEMA. (*Mediana per dati raggruppati*)
Supponiamo di avere solo dei *dati numerici continui già suddivisi in classi* e di conoscere le loro frequenze relative: allora possiamo definire la *mediana* come quel valore che divide l'insieme ei dati in due *gruppi ugualmente numerosi*: ovvero vogliamo trovare il valore k per cui si ha $F_a(k-1) < \frac{n}{2} \leq F_a(k)$. Chiamando $M_k = \sup A_k$ e $m_k = A_k$ e $m_k = \inf A_k$ (quindi $M_k - m_k$ fornisce la misura dell'intervallo), abbiamo che la mediana M_e è il valore per cui

$$\frac{n}{2} = F_a(k-1) + \frac{(M_e - m_k)}{(M_n - m_k)} \underbrace{(F_a(k) - F_a(k-1))}_{f_a(k)}$$

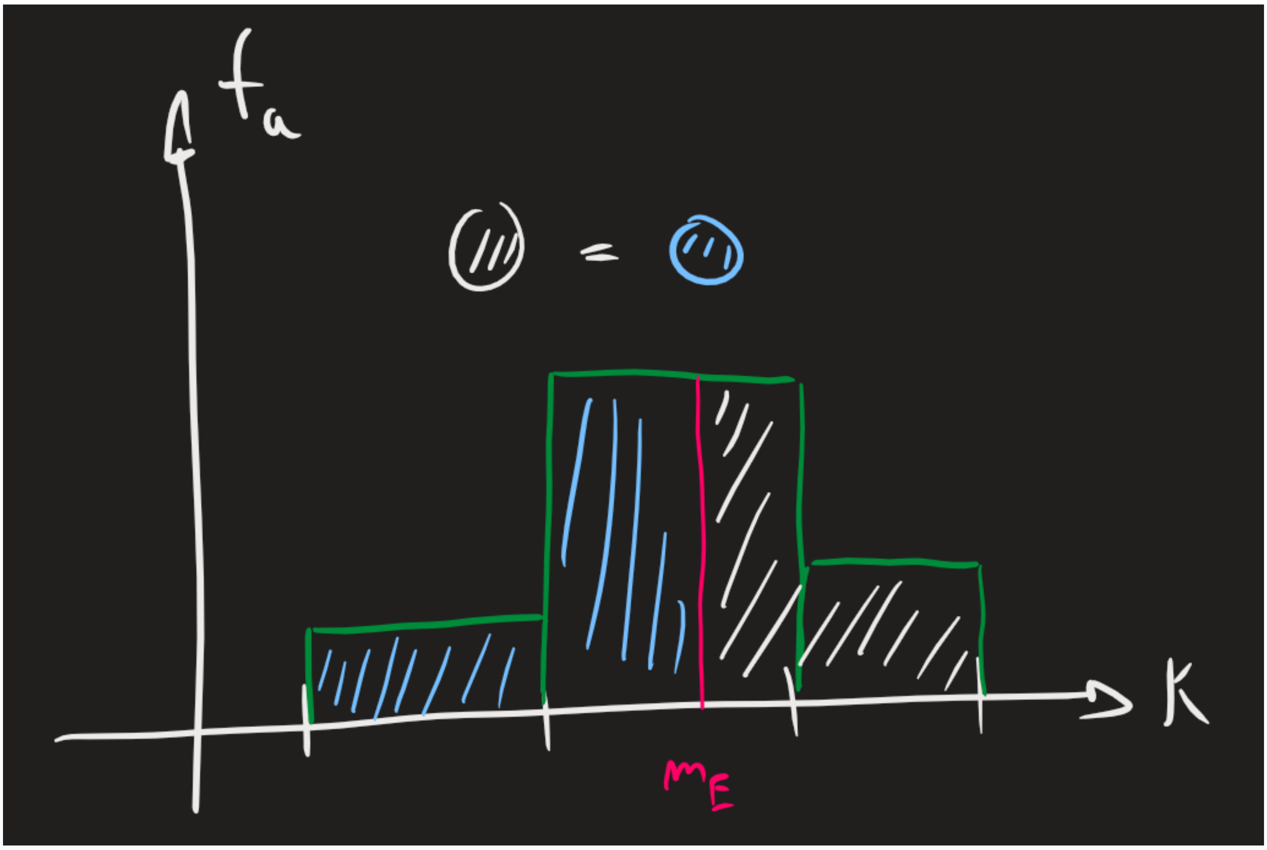
Geometricamente possiamo interpretare la formula in due modi.
La prima consiste nel considerare l'*ogiva* e l'*istogramma* dei dati: possiamo vedere l'*ogiva* come una *funzione lineare* (assumiamo l'uniformità dei dati), ne calcolo la retta r e impongo che $y(k) = \frac{n}{2}$, infine per ottenere la formula. Infatti si può vedere la formula appena data come

$$y = q + \frac{\Delta y}{\Delta x}(x - m_k)$$



Altrimenti, pensando direttamente con l'*istogramma*, possiamo pensare la mediana M_e come il punto per cui l'area di tutti i rettangoli prima di M_e e dopo di M_e siano uguali, con $A = \frac{n}{2}$. In questo caso lo vedo come un *parametro* al variare. Quindi si tratta di trovare un valore M_e tale che

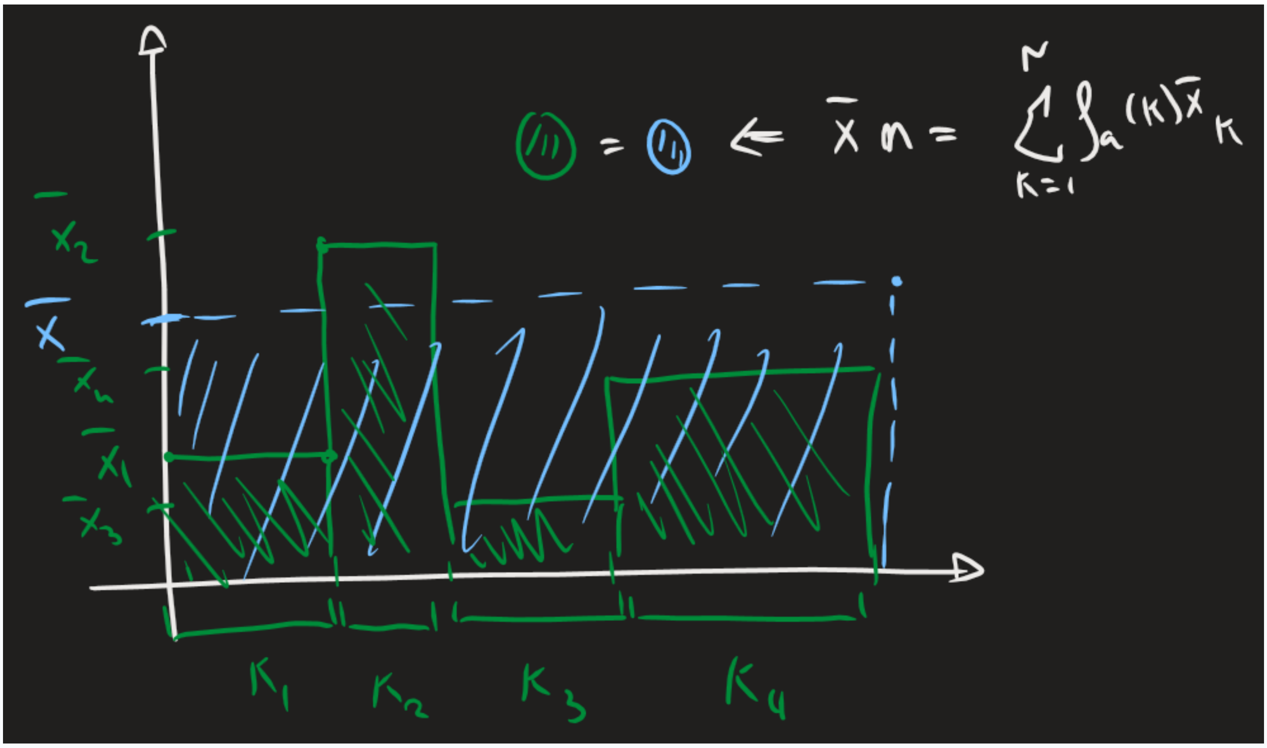
$$
\begin{aligned}
&\frac{n}{2} \\
&= \sum_{j<k} f_a(j)(M_j - m_j) + (M_e - m_k) f_a(k) \\
&= (M_k - M_e) f_a(k) + \sum_{k<j<n} f_a(j)(M_j - m_j)
\end{aligned}
$$



#Osservazione

PROBLEMA. (*Media per dati raggruppati*)
Come fatto prima, assumo *linearità* (ovvero l'uniformità dei *dati raccolti*). In questo modo il loro *valore medio* coincide con il *valore medio dell'intervallo che realizza la classe* A_k. Sia definita $\bar{x}_k$ questo tale valore per A_k (ovvero $\bar{x}_k := (M_k - m_k)(0.5)$). Assumendo di avere N classi, abbiamo che $\bar{x}$ è ottenuta con la *media ponderata* degli $\bar{x}_k$ con i pesi $f_a(k)$:

$$\bar{x} = \sum_{n=1}^{N} f_r(k)\bar{x}_k = \frac{1}{n}\sum_{n=1}^{N} f_a(k)\bar{x}_k$$



#Corollario

Corollario (proprietà della media per dati raggruppati).

Si ha che:
i. Applicando una qualsiasi trasformazione *lineare* ai dati del tipo $y_j = ax_j + b$, si ha che la sua media diventa $\bar{y} = a\bar{x} + b$
ii. La somma degli scarti della media è nulla (dovuta per l'uniformità):

$$\sum_{j=1}^{n}(x_j - \bar{x}) = 0$$

iii. La somma dei quadrati dagli scarti della media è minima per $x = \bar{x}$:

$$\sum_{j=1}^{n}(x_j - x)^2 = \min \iff x = \bar{x}$$

3. Mediana e Media

#Osservazione

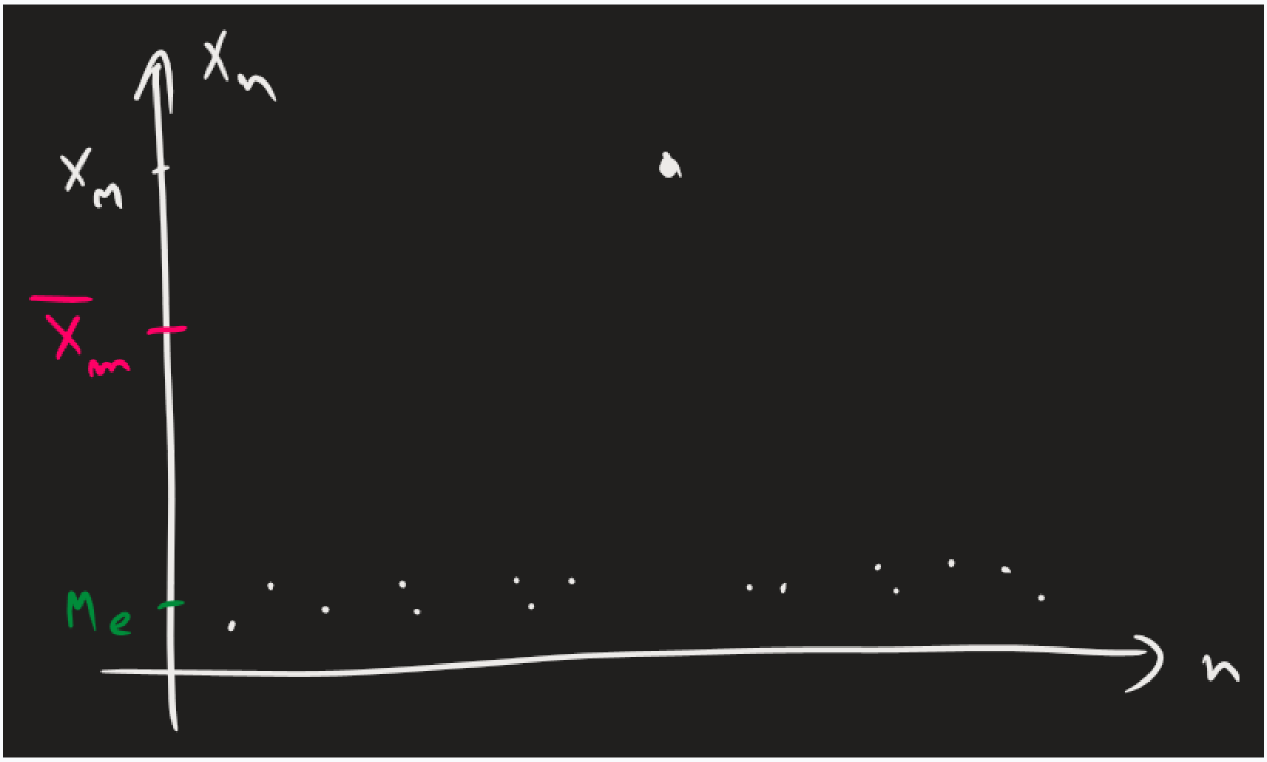
Osservazione (differenza tra media e mediana).

Osserviamo un po' di differenze tra la *media* e la *mediana*. Cosa succede aggiungiamo un dato x_M con ampiezza minima, a cui facciamo tendere il valore $x_M \to +\infty$ (in altre parole stiamo aggiungendo delle *singolarità*, o delle c.d.

outlier)? Nel caso della mediana M_e abbiamo che rimane invariata. Invece per la media $\bar{x}$ abbiamo che *cresce* di tanto!

In un certo senso, abbiamo che la *media* è un dato *molto sensibile*, la *mediana* è *meno sensibile*.

Inoltre, questi due *sono tanto più vicine* quanto più *i dati sono disposti uniformemente*.



3. Varianza

Con l'osservazione appena effettuata, vogliamo descrive *"come"* si comporta un set di dati rispetto alla sua media $\bar{x}$.

#Definizione

Definizione (varianza, scarto quadratico medio (o deviazione standard)).

Il principale indice di dispersione è la *varianza*. Dato $\{x_1, \dots, x_n\}$ dati numerici e $\bar{x}$ la sua media campionaria, ho

$$\sigma^2 := \frac{1}{n} \sum_{j=1}^{n} (x_j - \bar{x})^2$$

In particolare prendendo la sua *radice squadrata* ho la *deviazione standard*.

Ho che questa indica la *"dispersione"* da $\bar{x}$; infatti si annulla per $x_j = \bar{x}$ per tutti gli x_j, e si ha pure la formula equivalente

$$\sigma^2 = \left(\frac{1}{n}\sum_{j=1}^{n} x_j^2\right) - (\bar{x})^2$$

Adesso andiamo a vedere un problema.

#Osservazione

PROBLEMA. (*Varianza per dati raggruppati*)

Se i dati sono già stati suddivisi e si conosce solo la loro *frequenza*, possiamo trovare solo un'approssimazione dato che con tale rappresentazione si è perso una parte dell'informazione.

Allora, per *"recuperare"* l'informazione, seguiamo la seguente convenzione per approssimare: ai dati nella classe k-esima prendiamo $\tilde{x}_k := f_a(k) \cdot \bar{x}_k$ (ovvero moltiplichiamo la media relativa della classe per la sua frequenza assoluta).

Adesso possiamo prendere la varianza per $\tilde{x}_k$ e abbiamo

$$\sigma^2 = \frac{1}{n}\sum_{k=1}^{N} f_a(k)(\bar{x}_k - \bar{x})^2 = \frac{1}{n}\left(\sum_{k=1}^{N} f_a(k)\bar{x}_k^2\right) - (\bar{x})^2$$

Ovviamente è possibile fare anche *assumendo l'uniformità dei dati*, che ci porta allo stesso risultato.

Badiamo bene che dividiamo per n, dal momento che è come se trattassimo n dati, non N.

X

B. MODELLI STATISTICI

B1. Definizione di Modello Statistico

Definizione di Modello Statistico

X

Definizione di modello statistico: esempio euristico. Definizioni relative: campione, statistica, stimatore, stima corretta e consistente.

X

1. Preambolo

#Esempio

ESEMPIO. (*Il lancio del dado*)

Torniamo ad esaminare il lancio di un dado: l'abbiamo descritta come uno spazio di probabilità $(\Omega, \mathcal{A}, p)$ dove $\Omega = \{1, \ldots, 6\}$, $\mathcal{A} = \operatorname{Part} \Omega$. In particolare abbiamo scelto di assumere l'equiprobabilità, ovvero che valga

$$\forall \omega \in \Omega, p(\{\omega\}) = \frac{1}{6}$$

Tuttavia, questo vale solo se il dado è effettivamente *bilanciato*! Se fosse sbilanciato, potrebbe valere qualcosa del tipo

$$p(\{1\}) = \frac{1}{4}, p(\{2\}) = \ldots = p(\{6\}) = \frac{3}{20}$$

Adesso generalizziamo questo concetto ponendo $\theta \in [1, +\infty)$ e la *probabilità al variare in* θ come

$$p_\theta(\{1\}) = \frac{1}{\theta}, p(\{2\}) = \ldots = p(\{6\}) = \left(1 - \frac{1}{\theta}\right)\frac{1}{5}$$

Quindi a seconda della scelta di θ, abbiamo una specifica probabilità $p_\theta : \mathcal{A} \longrightarrow [0, 1]$ e un relativo spazio di probabilità $(\Omega, \mathcal{A}, p_\theta)$. A seconda della situazione, abbiamo il parametro θ che fornisce uno spazio probabilistico *"migliore"*. Tale valore viene determinato dalle *osservazioni*, che di solito sono dei *valori assunti* da una famiglia di variabili aleatorie.

2. Definizioni

#Definizione

Definizione (modello statistico, campione e statistica).

Un *modello statistico* è una *famiglia di spazi di probabilità* del tipo

$$(\Omega, \mathcal{A}, p_\theta)_{\theta \in \Theta}$$

dove Θ è un opportuno insieme per il parametro θ.

Si dice un *campione* per tale modello statistico come una *successione di variabili aleatorie i.i.d.* $(X_n)_n$, ciascuna con legge $\mathcal{L}_\theta$. Nel caso discreto ognuna avrà una densità $q(x, \theta)$ mentre nel caso continuo sarà $f(x, \theta)$.
La n-upla $(X_1, \ldots, X_n)$ è detta un *campione di ampiezza* n.

#Definizione

Definizione (nozioni per stimare il parametro).

Si dice *statistica* relativa al campione come una *variabile aleatoria* del tipo $H = h(X_1, \ldots, X_n)$ con $h : \mathbb{R}^n \longrightarrow \mathbb{R}$: ovvero una specie di *"elaborazione"* del campione.

Consideriamo una *successione di statistiche* $(H_n)_n$ del tipo $H_n = h_n(X_1, \ldots, X_n)$, con $h_n : \mathbb{R}^n \longrightarrow \mathbb{R}$ usate per stimare θ (o in più generale una funzione $\psi(\theta)$): questo si dice *stimatore* del parametro θ.
Se $\theta \in \mathbb{R}^M$ (ovvero ha più parametri), in generale si prendono le loro proiezioni singolari $\psi_1(\theta) = \theta_1, \ldots, \psi_M(\theta) = \theta_M$.

Una volta estratto un *particolare campione* $(x_1, \ldots, x_n)$ relativo alle v.a. $(X_1, \ldots, X_n)$, il valore numerico $h(x_1, \ldots, x_n)$ è detto *stima* per $\psi(\theta)$. Notiamo il fatto di osservare il campione $(x_1, \ldots, x_n)$ significa che si è verificato l'evento $\{X_k = x_k, \forall k = 1, \ldots, n\}$.

#Definizione

Definizione (proprietà degli stimatori).

Uno stimatore H_n di $\psi(\theta)$ è detto *corretto* se vale che

$$\forall \theta \in \Theta, \forall n \in \mathbb{N}, E_\theta[H_n] = \psi(\theta)$$

Inoltre è detto *consistente* se vale che

$$\forall \varepsilon > 0, \lim_n p_\theta(|H_n - \psi(\theta)| > \varepsilon) = 0$$

In altre parole uno stimatore è *corretto* se la media di H_n è *sempre* uguale a $\psi(\theta)$ (ovvero il *"target"*), è *consistente* se lo scarto dell'errore tende a 0.

#Esempio

Consideriamo un modello statistico in cui $q(x, \lambda)$ è la densità di tipo *Poisson* $P(\lambda)$ (Definizione 1 (densità di Poisson)). Prendiamo come parametro $\theta = \lambda$ e dunque $\Theta = (0, +\infty)$.
Adesso supponiamo di avere un campione $(x_1, \ldots, x_n)$ che realizza $\{X_1 = x_1, \ldots, X_n = x_n\}$.
Per stimare λ usiamo il fatto che $E_\theta[X] = \theta = \lambda$: ovvero la media è il parametro stesso.
Dunque poniamo il stimatore h_n come

$$h_n(\underline{x}) = \frac{1}{n} \sum_{x_i \in \underline{x}} x_i = \overline{x}_n$$

Abbiamo dunque che $H_n = h_n(X_1, \ldots, X_n)$, che coincide con la media campionaria $\overline{X}_n$.
Adesso, per verificare la correttezza di questo stimatore prendiamo il suo valor medio:

$$E_\theta[H_n] = \frac{1}{n} \sum_{k=1}^{n} E_\theta[X_k] = \frac{n\lambda}{n} = \lambda = \theta$$

che prova $(H_n)_n$ essere uno stimatore corretto per λ.

B2. Esempi Notevoli di Stimatori

Esempi Notevoli di Stimatori

X

Stimatori notevoli: media e varianza campionaria su v.a.

X

0. Voci correlate

- Definizione di Modello Statistico

1. Esempio Preliminare

#Esempio

Considerare un *modello statistico* in cui $f(x, \theta)$ è la densità del tipo gaussiano $\mathcal{N}(\mu, \sigma^2)$. Vogliamo stimare $\underline{\theta} = (\mu, \sigma^2)$ con $\underline{\Theta} = \mathbb{R} \times (0, +\infty)$. Dato un campione $(X_n)_n$ consideriamo le seguenti *statistiche campionarie*:

$$\overline{X}_n = \frac{1}{n} \sum_{k=1}^{n} X_n \mid s_n^2 = \frac{1}{n-1} \sum_{k=1}^{n} (X_k - \overline{X}_n)^2$$

Adesso consideriamo le seguenti funzioni sul stimatore target: la definiamo come

$$\psi(\underline{\theta} = (\mu, \sigma^2)) := \begin{pmatrix} \psi_1(\underline{\theta}) \\ \psi_2(\underline{\theta}) \end{pmatrix} = \begin{pmatrix} \mu \\ \sigma^2 \end{pmatrix}$$

ovvero scompone in suo componente tale parametro in μ e σ^2. Verificheremo che $\overline{X}_n$ è uno stimatore corretto per ψ_1 e s_n^2 per ψ_2.
i. Per il primo caso abbiamo

$$E_\theta[\overline{X}_n] = \frac{1}{n} \sum_{k=1}^{n} E[X_n] = \mu = \psi_1(\underline{\theta})$$

che dimostra essere corretto.
ii. Per il secondo caso abbiamo a che fare con conti analoghi:

$$
\begin{aligned}
(n-1)E_\theta[s_n^2] &= \sum_{k=1}^n E_\theta[(X_k - \overline{X}_n)^2] \\
&= \sum_{k=1}^n \left(E_\theta[X_k^2] + E_\theta[\overline{X}_n^2] - 2E[X_k\overline{X}_n]\right) \\
&= \sum_{k=1}^n E_\theta[X_k^2] + nE_\theta[\overline{X}_n^2] - 2\underbrace{E[\Sigma_{k=1}^n X_k\overline{X}_n]}_{nE[\overline{X}_n^2]} \\
&= n(\mu^2+\sigma^2) - n\left(\mu^2 + \frac{\sigma^2}{n}\right) = (n-1)\sigma^2
\end{aligned}
$$

che prova la correttezza di s_n^2. Nell'ultimo passaggio è stato usato il fatto che $E[X^2] = \operatorname{var} X + E[X]^2$ e che $\operatorname{var} \overline{X}_n = \frac{\sigma^2}{n}$.

Notiamo che nell'esempio non abbiamo usato il fatto che X_n sono del tipo gaussiane; bensì solo il fatto che hanno media μ e varianza σ^2. Possiamo infatti generalizzare come il seguente

2. Stimatori Notevoli

#Lemma

Lemma (stimatori notevoli).

Dato un *modello statistico* $(\Omega, \mathcal{A}, p_\theta)_{\theta\in\Theta}$ e un *campione* $(X_n)_n$ con X_n avente *momento secondo finito* rispetto ogni p_θ, definiamo dunque $\mu, \sigma^2 : \Theta \longrightarrow \mathbb{R}$ come

$$
\mu(\theta) = E_\theta[X_n] \ \wedge \ \sigma^2(\theta) = \operatorname{var}_\theta X_n
$$

Allora gli stimatori $\overline{X}_n$ e s_n^2 sono *ben posti* per $\mu = \mu(\theta)$ e $\sigma^2 = \sigma^2(\theta)$ rispettivamente.

#Dimostrazione

DIMOSTRAZIONE del Lemma 1 (stimatori notevoli)
Vedere l'esempio preliminare (ovvero (1) e (2)).

X

C. STIMA PUNTUALE E PER INTERVALLI DI CONFIDENZA

C1. Metodo della Massima Verosimiglianza

Metodo della Massima Verosimiglianza

X

Metodo empirico per la stima puntuale: il metodo della massima verosimiglianza. Principio base, definizione di funzione di verosimiglianza e stimatore di massima verosimiglianza. Equazione di verosimiglianza.

X

0. Voci correlate

- Definizione di Modello Statistico

1. Idea Base

#Osservazione

IDEA. (*Massima verosimiglianza*)
Sia $X = (X_1, \ldots, X_n)$ un campione di ampiezza n per il modello statistico $(\Omega, \mathcal{A}, p_\theta)_{\theta \in \Theta}$ (Definizione 1 (modello statistico, campione e statistica)). Supponiamo che siano discrete o assolutamente continue con densità congiunte $x \in \mathbb{R}^n \longrightarrow L(x, \theta)$ (ciò si può fare in quanto ho v.a. i.i.d.). Nel primo caso ho sommatorie, nel secondo caso ho integrali.
Supponiamo inoltre che si realizzi $\hat{x} = (x_1, \ldots, x_n)$.
Allora basandosi su tale $\hat{x}$ posso scegliere il *"miglior parametro"* θ scegliendo in corrispondenza la funzione $L_{|\hat{x}}(\theta) = L(\hat{x}, \theta)$ (ovvero prendiamo la restrizione su $\hat{x}$).
Adesso ciò che ci rimane è quello di *massimizzare* $L_{|\hat{x}}(\theta)$: il principio è che *"ciò che è stato osservato è ciò che aveva maggior probabilità di verificarsi"*.
Si sceglie dunque il θ nel modo che $p_\theta(X = \hat{x})$ sia massima in $x = \hat{x}$. In effetti nel caso discreto si ha

$$L(\hat{x}, \theta) = p_\theta(X = \hat{x}) = q(X = \hat{x}) \implies p_\theta = L$$

Per approssimazione lo si fa pure nel caso assolutamente continuo.

2. Definizioni

#Definizione

> Definizione (funzione di massima verosimiglianza e stimatore di massima verosimiglianza).
>
> Sia $(\Omega, \mathcal{A}, p_\theta)_{\theta \in \Theta}$ un modello statistico, X un campione di ampiezza n avente densità congiunta $L(\theta, x)$ e $\hat{x}$ una sua osservazione. Allora:
>
> Si dice una *funzione di massima verosimiglianza* come la restrizione $L_{|\hat{x}}(\theta)$.
> Si definisce uno *stimatore di massima verosimiglianza* $H = h(X)$ se per ogni

osservazione x ho la funzione di verosimiglianza $L_{|x}(\theta)$ che raggiunge il massimo per $\theta = h(x)$.

#Osservazione

Osservazione (note).

Si nota che:
Di solito si procede *"al contrario"* per gli stimatori di massima verosimiglianza, ovvero partiamo da $\theta = h(x)$ e poi ne deduciamo il valore.

Inoltre non è detto che uno *stimatore di massima verosimiglianza* possa esistere o dev'essere unico: ad esempio se ho $L_x(\theta)$ non raggiunge il massimo assoluto oppure se $\mathrm{d}L_x(\theta) = 0$ presenta più soluzioni.

3. Equazione della Massima Verosimiglianza

#Osservazione

Osservazione (trucco preliminare).

Prima di vedere la formuletta pratica per determinare il θ, vediamo un trucco tecnico.
Supponendo che Θ sia un aperto e che $L_x(\theta)$ sia regolare, andiamo a massimizzare L'_x (o in più generale ∇L_x).
Invece spesso è più comodo cercare i punti di massimo per $\log L_x(\theta)$: poiché il *logaritmo è crescente*, i punti di massimo sono gli stessi.

Una volta trovato un θ che massimizzi L'_x, occorre verificare che sia effettivamente il massimo: ovvero se

$$L_x(\theta(x)) = \max\{L_x(\theta) : \theta \in \Theta\}$$

(bisogna prestare attenzione particolare per la frontiera $\partial\Theta$!)

Poi quando ho finito, si pone $h(x) = \theta(x)$.

#Teorema

Teorema (formula della massima verosimiglianza).

Sia $(X_1, \dots, X_n)$ il campione di un modello statistico, sia $l(x, \theta)$ la densità di ognuna delle X_k e $L(x, \theta)$ la densità congiunta.
Supponiamo Θ aperto e l regolare.

> Allora si ha il θ massimizzante per
>
> $$\boxed{\sum_{k=1}^{n}\left(\frac{\mathrm{d}l(x_k,\theta)}{\mathrm{d}\theta}\cdot\frac{1}{l(x_k,\theta)}\right)=0}$$

#Dimostrazione

DIMOSTRAZIONE del Teorema 4 (formula della massima verosimiglianza)
Si tratta di usare il trucchetto appena osservato: ovvero consideriamo la la composizione $(\log\circ l_x)(\theta)$. Poiché il logaritmo è crescente, ho che i punti di massimo per la composizione e per l_x sono le stesse.
Adesso consideriamo la densità marginale: essendo le v.a. indipendenti, ho

$$L(x,\theta)=\prod_{k=1}^{n}l(x_k,\theta)$$

Adesso applicando il logaritmo ho una *banale somma* (per le proprietà del logaritmo), che ci semplifica di molto la nostra vita. Adesso supponendo che θ sia uno-dimensionale, imponiamo

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\ln L_x(\theta)=\frac{\mathrm{d}}{\mathrm{d}\theta}\sum_{k=1}^{n}\ln(l(x_k,\theta))=0$$

Usando le proprietà della derivata si ricava la tesi. ■

4. Esempi

Vediamo degli esempi.

#Esempio

ESEMPIO. (*Media campionaria della normale*)
Supponiamo di avere un campione $(X_1,\dots,X_n)$ con $X_k\sim\mathcal{N}(\mu,\sigma^2)$. Vogliamo stimare μ.
Essendo le X_k indipendenti, si ha la densità congiunta

$$\begin{aligned}L_x(\mu)&=\prod_{k=1}^{n}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x_k-\mu)^2}{2\sigma^2}\right)\\&=\frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}}\exp\left(-\sum_{k=1}^{n}\frac{(x_k-\mu)^2}{2\sigma^2}\right)\end{aligned}$$

In questo caso *non* ho bisogno di usare il logaritmo in alcun modo, dato che ho già fatto trasformare la produttoria in una sommatoria. Volendo possiamo derivarlo e ottenere l'equazione di massima verosimiglianza, ma per ora usiamo un altro modo.
Guardando l'equazione di verosimiglianza dobbiamo calcolare la derivata $l_\mu(x_k,\mu)$ che è

$$l_\mu(x_k,\mu) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \frac{2(x_k-\mu)}{2\sigma^2} \exp\left(-\frac{(x_k-\mu)^2}{2\sigma^2}\right) = \frac{x_k-\mu}{\sigma^2} l(x_k,\mu)$$

Adesso ponendo l'equazione di verosimiglianza ho

$$\sum_{k=1}^{n}\left(\frac{\mathrm{d}l(x_k,\mu)}{\mathrm{d}\mu}\cdot\frac{1}{l(x_k,\mu)}\right) = 0 \implies \sum_{k=1}^{n}\frac{x_k-\mu}{\sigma^2} = 0$$

da cui ricavo che

$$\mu(x) = \frac{1}{n}\sum_{k=1}^{n} x_k$$

che è proprio la *media campionaria*. ■

#Esempio

ESEMPIO. (*Massima verosimiglianza di una Poisson*)
Sia $(X_1,\ldots,X_n)$ con $X_k \sim P(\lambda)$ (Definizione 1 (densità di Poisson)). Stimare λ.
Sia fissato un $x=(x_1,\ldots,x_n)$ come uno dei possibili valori osservati. Allora si ha che la densità congiunta in λ è

$$L_{|x}(\lambda) = \prod_{k=1}^{n}\frac{\lambda^{x_n}}{x_n!}e^{-\lambda} = e^{-n\lambda}\frac{\lambda^{\sum_{k=1}^{n}x_n}}{\prod_{k=1}^{n}x_k!}$$

Prendendo il suo logaritmo ho

$$\ln(L_{|x}(x)) = -n\lambda + \left(\sum_{k=1}^{n}x_k\right)\ln\lambda + \cancel{\ln\left(\prod_{k=1}^{n}\frac{1}{x_k!}\right)}$$

(si cancella l'ultima parte dato che è una *costante*, che diventa trascurabile in termini di ottimizzazione)
Allora derivandolo rispetto a λ e ottimizzandolo ottengo

$$-n + \frac{1}{\lambda}\sum_{k=1}^{n}x_n = 0$$

Ne segue che ho lo *stimatore di massima confidenza* come

$$h(x) = \frac{1}{n}\sum_{k=1}^{n}x_n$$

che è proprio la *campionaria*: risultato non sorprendente, dato che $E[X_k]=\lambda$. ■

#Esempio

ESEMPIO. (*Distribuzione esponenziale*)

Adesso prendiamo $X_k \sim \mathcal{E}(\lambda)$ (distribuzione esponenziale, Definizione 1 (densità esponenziale)). Considerando la sua congiunta ho

$$L_{|x}(\lambda) = \prod_{k=1}^{n} \lambda e^{-\lambda x_n} = \lambda^n e^{-\lambda \sum_{k=1}^{n} x_k} = e^{n \ln(\lambda) - \lambda \sum(\dots)}$$

prendendo il logaritmo ho

$$\ln L_{|x}(\lambda) = n \ln \lambda - \lambda \sum_{k=1}^{n} x_k$$

Derivandolo ottengo l'equazione di massima verosimiglianza

$$\frac{n}{\lambda} - \sum_{k=1}^{n} x_k = 0$$

da cui ricavo

$$H(x) = \frac{1}{\frac{1}{n} \sum_{k=1}^{n} x_k} \sim \frac{1}{\bar{X}}$$

ovvero il reciproco della media campionaria. Notiamo che per la distribuzione geometrica (Definizione 1 (densità geometrica)) ho il risultato analogo. ■

5. Nota: questo metodo non garantisce la correttezza

Notiamo che questo metodo *non* garantisce la correttezza dello stimatore H. Vediamone un esempio.

#Esempio

ESEMPIO. (*Varianza della gaussiana*)

Torniamo al caso gaussiano, di cui vogliamo stimare σ^2. Allora:

$$\begin{aligned} L_x(\sigma^2) &= \prod_{k=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_k - \mu)^2}{2\sigma^2}\right) \\ &= \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left(-\sum_{k=1}^{n} \frac{(x_k - \mu)^2}{2\sigma^2}\right) \end{aligned}$$

Adesso prendendone il logaritmo ho

$$\ln(L_x(\sigma^2)) = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{k=1}^{n}(x_k-\mu)^2$$

Poi imponendo la sua derivata nulla, ottengo l'equazione della max. verosimiglianza che è

$$\ln(L_x(\sigma^2))' = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{k=1}^{n}(x_k-\mu) \equiv 0$$

Trovo che si annulla per

$$\sigma^2 = \frac{1}{n}\sum_{k=1}^{n}(x_k-\mu)^2$$

ovvero

$$H = \frac{1}{n}\sum_{k=1}^{n}(X_k-\mu)$$

Allora ho che è proprio la *varianza* con μ noto. E se invece μ non fosse noto? Allora in questo caso si dovrà considerare il vettore $\theta = (\mu, \sigma^2) \in \mathbb{R}^2$ e considerare invece il gradiente $\nabla \ln L_{|x}(\theta)$ e imporlo uguale a 0. Ovvero

$$\begin{cases} \dfrac{\partial L_{|x}(\theta)}{\partial \mu} = 0 \\ \dfrac{\partial L_{|x}(\theta)}{\partial \sigma^2} = 0 \end{cases}$$

Come visto prima, la prima equazione restituisce

$$\mu = \frac{1}{n}\sum_{k=1}^{n}x_k$$

Come visto prima, la seconda equazione restituisce

$$\sigma^2 = \frac{1}{n}\sum_{k=1}^{n}(x_k-\mu)^2$$

Quindi inserendo la prima nella seconda ho

$$\sigma^2 = \frac{1}{n}\sum_{k=1}^{n}\left(x_k - \frac{1}{n}\sum_{k=1}^{n}x_k\right)^2$$

Allora ho lo stimatore

$$H = \frac{1}{n}\sum_{k=1}^{n}\left(X_k - \bar{X}_n\right)$$

che purtroppo non è corretto. ■

C2. Intervallo di Confidenza

Intervallo di Confidenza

X

Stima per intervalli: idea preliminare (generalizzazione delle stime puntuali), definizione di intervallo di confidenza al livello $100\beta\%$.

X

0. Voci correlate

- Definizione di Modello Statistico
- Esempi Notevoli di Stimatori

1. Introduzione Preliminare

#Osservazione

IDEA. (*Stima per intervalli*)
Fino ad ora ci siamo occupati di *stime puntuali* per $\psi(\theta)$ con i vari metodi. In particolare abbiamo visto i casi in cui ψ è il valore medio μ o la varianza σ^2. Quindi adesso conosciamo uno stimatore $(H_n)_n$: tuttavia non ne conosciamo la *accuratezza*; se è corretto, sappiamo solo che sarà in grado di fornire una stima sempre più accurata al crescere di n, ma non la velocità.
Risulta quindi necessario estrarre dal campione anche una *stima dell'accuratezza della stima puntuale*. Tale stima si basa sulla costruzione di un intervallo, detto *intervallo di confidenza* che fa *"contenere"* il valore $\psi(\theta)$. Questa stima viene detta *"stima per intervalli"*. Vediamo il seguente esempio preliminare.

#Esempio

ESEMPIO. (*Stima per intervalli della media di una popolazione normale con varianza nota*)
Sia $(X_n)_n$ un campione di ampiezza n con $X_k \sim \mathcal{N}(\mu, \sigma^2)$. Supponiamo di conoscere σ^2 e di voler stimare $\mu \in [?, ?]$.
Allora per stimare μ abbiamo $\bar{X}_n$ come uno *stimatore corretto*. Ricordiamo che per n grandi abbiamo

$$\bar{X}_n \approx Y \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

Rinormalizzando la Y abbiamo

$$\tilde{Y} = \frac{Y - \mu}{\frac{\sigma}{\sqrt{n}}} \sim \mathcal{N}(0, 1)$$

Adesso vogliamo stimare il parametro $\mu \in \mathbb{R}$: dato un $\alpha \in \mathbb{R}$, abbiamo

$$p_\mu\{\tilde{Y} < \alpha\} = \phi(\alpha)$$

(ϕ è la funzione di ripartizione per la gaussiana normale)
Esplicitando tutto possiamo scrivere pure

$$p_\mu\left\{|\bar{X}_n - \mu| < \frac{\sigma}{\sqrt{n}}\alpha\right\} = p_\mu\left\{-\alpha < \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} < \alpha\right\} = \phi(\alpha) - \phi(-\alpha) = 2\phi($$

Adesso sia dato un parametro $\beta \in (0, 1)$ che avrà il ruolo di *"threshold per la confidenza"* tale che

$$\alpha(\beta) = 2\phi(\alpha) - 1 = \beta$$

Sia fissato il β, da cui ho un valore $\alpha(\beta)$ (che viene calcolato usando l'inversa ϕ^{-1}. Quindi abbiamo la relazione

$$p_\mu\left\{|\bar{X}_n - \mu| < \frac{\sigma}{\sqrt{n}}\alpha(\beta)\right\} = \beta$$

Questo significa che *prima di eseguire il campionamento*, valutiamo pari a β la probabilità p_μ che risulti

$$|\bar{X}_n - \mu| < \frac{\sigma}{\sqrt{n}}\alpha(\beta) \iff \mu \in \left(\bar{X}_n - \frac{\sigma}{\sqrt{n}}\alpha(\beta), \bar{X}_n + \frac{\sigma}{\sqrt{n}}\alpha(\beta)\right)$$

Abbiamo ottenuto solamente un *intervallo aleatorio*: adesso possiamo eseguire il campionamento (supponiamo che $\bar{x}$ sia la media campionaria realizzata), da cui ho

$$\mu \in \left(\bar{x} - \frac{\sigma}{\sqrt{n}}\alpha(\beta), \bar{x} + \frac{\sigma}{\sqrt{n}}\alpha(\beta)\right)$$

L'intervallo reale ottenuto si dice *intervallo di confidenza* per μ al livello del 95% calcolato *dal campione*.

2. Definizione di Intervallo di Confidenza

Adesso formalizziamo tutto.

#Definizione

Definizione (intervallo di confidenza e intervallo di confidenza calcolato dal campione).

Sia $(X_n)_n$ un campione con distribuzione avente densità $f(x, \theta)$. Fissata l'ampiezza $n \in \mathbb{N}$ del campione, siano $H_n = h_n(X_1, \ldots, X_n)$ e $G_n = g_n(X_1, \ldots, X_n)$ due *statistiche* e sia $\psi = \psi(\theta)$ una funzione del parametro che si vuole stimare.

Fissato un *numero* $\beta \in (0,1)$, l'intervallo aleatorio (H_n, G_n) è detto *intervallo di confidenza* al livello $100\beta\%$ per $\psi(\theta)$ se vale che

$$p_\theta(H_n < \psi(\theta) < G_n) = \beta$$

Sia adesso $\hat{x} = (x_1, \ldots, x_n)$ una realizzazione di X_n: l'intervallo numerico $(h_n(\hat{x}), g_n(\hat{x}))$ si dice *intervallo di confidenza calcolato dal campione*. Gli estremi dell'intervallo si dicono *limiti di confidenza*.

#Osservazione

Osservazione (la manipolazione dell'intervallo di confidenza).

Vediamo l'esempio preliminare: abbiamo che

$$\mu \in \left(\bar{X}_n - \frac{\sigma}{\sqrt{n}}\alpha(\beta), \bar{X}_n + \frac{\sigma}{\sqrt{n}}\alpha(\beta)\right)$$

Osserviamo le seguenti.
Prima di tutto vediamo che *fissando* l'ampiezza del campione X_n, abbiamo che se *aumentiamo* la confidenza β, allora crescerà anche α (dato che è sempre crescente) e quindi anche l'ampiezza dell'intervallo (e viceversa).

Per aumentare invece β senza dover aumentare l'ampiezza *dell'intervallo*, dobbiamo necessariamente aumentare l'ampiezza del campione X_n.

Similmente per diminuire l'ampiezza dell'intervallo, senza dover diminuire la confidenza, dobbiamo aumentare l'ampiezza del campione: infatti si ha il limite

$$\lim_n \left(\bar{X}_n - \frac{\sigma}{\sqrt{n}}\alpha(\beta), \bar{X}_n + \frac{\sigma}{\sqrt{n}}\alpha(\beta)\right) = \{\bar{X}_n\}$$

#Osservazione

Osservazione (regoletta pratica).

> Quando realizziamo l'intervallo aleatorio in intervallo reale, approssimiamo l'estremo sinistro in basso e l'estremo destro in alto.

C3. Stima per Intervallo della Media

Stima per Intervallo della Media

X

Stima per intervallo della media: varianza nota e incognita.

X

0. Voci correlate

- Intervallo di Confidenza

1. Stima per Intervallo con Varianza Nota

Adesso vediamo dei casi notevoli di *stima per intervallo*. Partiamo dal caso in cui vogliamo conoscere la *media* di un campione di gaussiane avente media μ e varianza σ^2.

#Teorema

> Teorema (stima per intervallo di media con varianza nota).
>
> Sia $(X_n)_n$ un *campione* di un modello statistico, con $X_k \sim \mathcal{N}(\mu, \sigma^2)$. Sia $\overline{X}_n$ la *media campionaria*. Sia $\phi(x)$ la funzione di ripartizione per la normale standard $\mathcal{N}(0, 1)$.
>
> Supponendo noto il valore σ^2, abbiamo che dato un $\beta \in (0, 1)$ l'intervallo di confidenza
>
> $$\boxed{\mu \in \left(\bar{X}_n - \frac{\sigma}{\sqrt{n}}\alpha(\beta), \bar{X}_n + \frac{\sigma}{\sqrt{n}}\alpha(\beta)\right)}$$
>
> dove $\alpha(\beta)$ viene calcolato come
>
> $$\alpha(\beta) = \phi^{-1}\left(\frac{\beta+1}{2}\right)$$

#Dimostrazione

DIMOSTRAZIONE del Teorema 1 (stima per intervallo di media con varianza nota)
Si tratta di usare il fatto che $\overline{X}_n \sim \mathcal{N}(\mu, \sigma^2/n)$ e poi di rinormalizzarlo verso $\mathcal{N}(0,1)$; dopodiché si prende $p_\mu(Y < \alpha) = \beta$, ed esplicitando tutto si ricava la tesi. Per ulteriori dettagli vedere la pagina sull'intervallo di confidenza (Intervallo di Confidenza > ^425062). ■

2. Stima della Media con Varianza Incognita

Adesso vediamo il caso in cui la varianza è *incognita*.

#Teorema

> Teorema (stima per intervallo di media con varianza incognita).
>
> Sia $(X_n)_n$ un *campione* di un modello statistico, con $X_k \sim \mathcal{N}(\mu, \sigma^2)$. Sia $\overline{X}_n$ la *media campionaria* e s_n^2 la sua *varianza campionaria*.
> Sia F la funzione di ripartizione della *t-student ad* $n-1$ *gradi di liberta* $t(n-1)$ (Definizione 1 (densità di Student)). Scelto un valore $\beta \in (0,1)$, abbiamo l'intervallo di confidenza per μ data come
>
> $$\boxed{\mu \in \left(\overline{X}_n - \alpha(\beta)\frac{\sqrt{s_n^2}}{\sqrt{n}}, \overline{X}_n + \alpha(\beta)\frac{\sqrt{s_n^2}}{\sqrt{n}}\right)}$$
>
> con $\alpha(\beta)$ data da
>
> $$\alpha(\beta) = F^{-1}\left(\frac{\beta+1}{2}\right)$$

#Dimostrazione

DIMOSTRAZIONE del Teorema 2 (stima per intervallo di media con varianza incognita)
Si tratta di usare la proprietà notevole della *densità di t-student*: ovvero che vale

$$t(n-1) \sim \sqrt{n}\frac{\overline{X}_n - \mu}{\sqrt{s_n^2}}$$

(Proposizione 4 (rappresentazione della varianza e media campionaria con la t-Student))
Usiamo questo fatto per effettuare la stima degli eventi

$$\left\{|\overline{X}_n - \mu| < \alpha\frac{\sqrt{s_n^2}}{\sqrt{n}}\right\}$$

Allora posto $\theta = (\mu, \sigma^2)$ abbiamo

$$p_\theta\left\{\bar{X}_n - \alpha\frac{\sqrt{s_n^2}}{\sqrt{n}} < \mu < \bar{X}_n + \alpha\frac{\sqrt{s_n^2}}{\sqrt{n}}\right\} \equiv \beta$$

Per trovare il valore α al variare di β si tratta di valutare

$$p_\theta\left\{|\overline{X}_n - \mu|\frac{\sqrt{n}}{\sqrt{s_n^2}} < \alpha\right\} = \beta \implies F(\alpha) - F(-\alpha) = \beta$$

da cui ho

$$\alpha(\beta) = F^{-1}\left(\frac{\beta+1}{2}\right)$$

per la simmetria della *t-student*. Per ottenere l'intervallo di confidenza in μ basta scegliere $\mu = \psi_1(\theta)$. ■

3. Esercizio

#Esercizio

Esercizio.

La ditta dei kiwi procede ad una misurazione a campione su 50 kiwi e trova che il diametro medio del campione è $2.04\ \mathrm{cm}$, con una varianza campionaria di $0.0225\ \mathrm{cm}^2$. Supponendo che il diametro dei frutti segue una legge normale, individuare un intervallo di confidenza al livello del 95% per la media calcolato dal campione.

C4. Stima per Intervallo della Varianza

Stima per Intervallo della Varianza

X

Stima per intervallo della varianza: caso media nota, media incognita.

X

0. Voci correlate

- Intervallo di Confidenza

1. Stima della Varianza con Media Nota

#Teorema

Teorema (stima della varianza con media nota).

Sia $(X_n)_n$ un campione di un modello statistico con $X_k \sim \mathcal{N}(\mu, \sigma^2)$ con μ nota. Sia H_n uno stimatore corretto per la varianza σ^2 (ovvero la *varianza campionaria*) e sia detta F la funzione di ripartizione associata a $\chi^2(n)$ (Definizione 1 (densità chi quadro χ^2)).

Allora, dato un valore $\beta \in (0,1)$ si ha il seguente *intervallo di confidenza* per σ^2:

$$\boxed{\sigma^2 \in \left(\frac{nH_n}{\alpha_2}, \frac{nH_n}{\alpha_1}\right)}$$

con α_1, α_2 dati come

$$\begin{cases} \alpha_1 = F^{-1}\left(\frac{1-\beta}{2}\right) \\ \alpha_2 = F^{-1}\left(\frac{1+\beta}{2}\right) \end{cases}$$

#Dimostrazione

DIMOSTRAZIONE del Teorema 1 (stima della varianza con media nota)

Dalla definizione di H_n abbiamo che

$$H_n = \frac{1}{n}\sum_{k=1}^{n}(X_k-\mu)^2$$

Allora calcolando $\frac{nH_n}{\sigma^2}$ ottengo

$$\frac{nH_n}{\sigma^2} = \sum_{k=1}^{n}\left(\frac{X_k-\mu}{\sigma}\right)^2 \sim: \chi^2(n)$$

Utilizziamo il punto appena dimostrato per stimare σ^2 dato in un intervallo aleatorio

$$\left\{\alpha_1 < \frac{nH_n}{\sigma^2} < \alpha_2\right\}$$

Per la correttezza di H_n avremo che la stima sarà buona. Adesso, per ottenere l'intervallo si tratta di calcolare p_{σ^2} nell'intervallo appena ottenuto, isolando la σ^2:

$$p_{\sigma^2}\left\{\frac{nH_n}{\alpha_2} < \sigma^2 < \frac{nH_n}{\alpha_1}\right\} \equiv \beta$$

Per calcolare i parametri α_1, α_2 si tratta di considerare

$$p_{\sigma^2}\left\{\alpha_1 < \frac{nH_n}{\sigma^2} < \alpha_2\right\} = F(\alpha_2) - F(\alpha_1) \equiv \beta$$

Dato che χ^2 non è simmetrica, non posso utilizzare nessun trucchetto: devo fare solo una scelta opportuna di $F(\alpha_2), F(\alpha_1)$ tali che la loro differenza viene β. In questo caso abbiamo $F(\alpha_2) = (1+\beta)/2$ e $F(\alpha_1) = (1-\beta)/2$, da cui segue la tesi. ■

2. Stima della Varianza con Media Incognita

#Teorema

> Teorema (stima della varianza con media incognita).
>
> Sia $(X_n)_n$ un campione di un modello statistico con $X_k \sim \mathcal{N}(\mu, \sigma^2)$. Sia H_n uno stimatore corretto per la varianza σ^2 (ovvero la *varianza campionaria*) e sia detta F la funzione di ripartizione associata a $\chi^2(n-1)$ (Definizione 1 (densità chi quadro χ^2)).
>
> Allora, dato un valore $\beta \in (0,1)$ si ha il seguente *intervallo di confidenza* per σ^2:
>
> $$\boxed{\sigma^2 \in \left(\frac{(n-1)H_n}{\alpha_2}, \frac{(n-1)H_n}{\alpha_1}\right)}$$
>
> con α_1, α_2 dati come
>
> $$\begin{cases} \alpha_1 = F^{-1}\left(\frac{1-\beta}{2}\right) \\ \alpha_2 = F^{-1}\left(\frac{1+\beta}{2}\right) \end{cases}$$

#Dimostrazione

DIMOSTRAZIONE del Teorema 1 (stima della varianza con media nota)
La dimostrazione è completamente analoga a quella, solo che si scende di un grado di libertà in quanto vale che

$$\frac{(n-1)s_n^2}{\sigma^2} \sim \chi^2(n-1)$$

(Proposizione 7 (proprietà fondamentale del chi quadro))
Dopodiché segue la tesi, considerando l'intervallo aleatorio

$$\left\{\alpha_1 < \frac{(n-1)s_n^2}{\sigma^2} < \alpha_2\right\}$$

■

3. Esercizio

#Esercizio

Esercizio.

Torniamo ancora una volta dalla nostra ditta di kiwi.
Su un campione di 40 kiwi è stata effettuata una varianza campionaria che restituisce $s_n^2 = 1.20 \text{ cm}^2$. Assumendo che i diametri dei kiwi seguano una legge normale, individuare un intervallo di confidenza del 98% per la varianza calcolato sul campione.