

STATISTICA

marcowber

June 2024

Indice

1 STATISTICA	3
1.1 Probabilità	3
1.2 Teorema di Bayes	4
1.3 Concetto di limite	4
2 TEOREMA CENTRALE DEL LIMITE	4
3 Distribuzioni di probabilità	5
3.1 Momenti	6
3.2 Distribuzioni notevoli	6
3.3 r.v. continue	6
3.4 r.v. discrete	9
3.5 Joint-pdf	11
4 PROPAGAZIONE DEGLI ERRORI	11
4.1 Cambio di variabili 1-D	11
4.2 Propagazione 1-D	11
4.3 Cambio di variabili n-D	12
4.4 Propagazione n-D	12
5 STIMA DI PARAMETRI	12
5.1 Campionamenti IID	12
5.2 Joint-pdf e likelihood	12
5.3 Statistica	13
5.4 Stimatore	13
5.5 Stime di μ e σ	14
5.6 Confidenza	14
5.7 Distribuzione χ^2	15
5.8 Errore quadratico medio	15

6	MAXIMUM LIKELIHOOD	16
6.1	Informazione	16
6.2	Max likelihood	16
6.3	Likelihood e gaussiana	17
7	LEAST SQUARES	17
7.1	LQ vs ML	17
7.2	Caso gaussiano	17
7.3	Teorema Gauss-Markov	17
7.4	Propagazione dell'errore delle x su y	18
7.5	Estrapolazione	18
8	FIT DI ISTOGRAMMI	18
8.1	Binned data	19
8.2	Fit LS	19
8.3	Fit ML	20
9	TEST DI IPOTESI	20
9.1	Null hypothesis	20
9.2	Alternative hypothesis	21
9.3	Test del chi-2	21
9.4	Test di Kolmogorov	22
9.5	Osservazione	22

1 STATISTICA

EVENTO CASUALE: risultato di un esperimento che non può essere previsto con certezza.

- ripetibile
- Si presenta in diverse modalità mutualmente esclusive

È rappresentato da numeri detti random variables.

Esse possono essere discrete o continue.

SPAZIO CAMPIONARIO (Ω): insieme di tutte le possibili modalità dell'evento.

POPOLAZIONE: Insieme di tutti i possibili eventi (insieme astratto contenente ∞ eventi)

CAMPIONE (sample): insieme degli eventi casuali raccolti.

Tramite il campione si possono stimare proprietà della popolazione (sampling).

1.1 Probabilità

DEFINIZIONE MATEMATICA - funzione $P : \Omega \rightarrow [0,1]$ t.c.

- $P(A) \geq 0$
- $P(\Omega) = 1$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

DEFINIZIONE CLASSICA - Rapporto tra numero di casi favorevoli e casi possibili.

DEFINIZIONE FREQUENTISTA - Frazione dei casi per cui un evento avviene calcolata per $N \rightarrow \infty$: $P(A) = \lim_{n \rightarrow \infty} \frac{n(A)}{N}$.

PROBABILITÀ CONDIZIONATA - probabilità che si verifichi B assumendo che si sia verificato A:

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \quad (1)$$

EVENTI INDIPENDENTI:

$$P(B|A) = P(B|\Omega) = P(B) \Rightarrow P(A \cap B) = P(A) \cdot P(B) \quad (2)$$

1.2 Teorema di Bayes

$$P(A|B) = \frac{P(B|A)}{P(B)} \cdot P(A) \quad (3)$$

Spesso cerchiamo di associare una probabilità a uno statement logico, si parla di probabilità soggettiva/Bayesiana (probabilità della plausibilità di una teoria).

$$P(\text{teoria}|\text{misura}) = \frac{P(\text{misura}|\text{teoria})}{P(\text{misura})} \cdot P(\text{teoria}) \quad (4)$$

$$\Rightarrow \text{Posterior}(\text{teoria}) = \frac{P(\text{misura}|\text{teoria})}{P(\text{misura})} \cdot \text{Prior}(\text{teoria}) \quad (5)$$

Misuro come la probabilità della teoria è modificata grazie all'informazione aggiuntiva.

1.3 Concetto di limite

Il concetto di limite adottato in statistica è differente dal classico: quando parliamo di eventi casuali non possiamo avere la certezza che lo scarto sia sempre inferiore a un determinato valore γ , ma solo che la probabilità che lo scarto sia superiore a ϵ sia via via più bassa.

$$NON : \lim_{x \rightarrow \infty} f(x) = L \Rightarrow |L - f(x)| \rightarrow 0 \quad (6)$$

$$MA : \lim_{x \rightarrow \infty} f(x) = L \Rightarrow P(|L - f(x)| > \epsilon) \rightarrow 0 \quad (7)$$

2 TEOREMA CENTRALE DEL LIMITE

Come è distribuita una r.v. ottenuta come somma di N r.v.?

Sappiamo che se estraiamo da pdf identiche riproducibili la r.v. segue la pdf di partenza. Cosa accade se sono differenti fra loro o non riproducibili?

Consideriamo N variabili aleatorie indipendenti x_i ciascuna caratterizzata da pdf_i , e per ciascuna pdf esistono finite media e varianza. Definiamo una nuova r.v. \bar{x} costituita dalla media delle r.v. di ciascuna pdf_i : $\bar{x} = \frac{\sum_{i=1}^N x_i}{N}$. Il TCL afferma:

- media di $pdf_{\bar{x}}$ è la somma delle medie $\cdot \frac{1}{N}$: $E[\bar{x}] = \mu_{\bar{x}} = \frac{\sum_{i=1}^N \mu_i}{N}$
- la varianza è la somma delle varianze $\cdot \frac{1}{N^2}$: $Var[\bar{x}] = \sigma_{\bar{x}}^2 = \frac{\sum_{i=1}^N \sigma_i^2}{N^2}$
- se $N \rightarrow \infty$ la r.v. \bar{x} è distribuita in modo gaussiano

Il TCL fornisce una spiegazione per l'osservazione: ripetendo la misura di una grandezza X tante volte, le misure raccolte si distribuiscono in modo gaussiano.

- la singola misura è affetta da un errore che la sposta dal valore vero, e che cambia ogni volta che è ripetuta; misura = valore vero + ϵ
- le misure sono distribuite come l'errore casuale ϵ
- esso è solitamente la somma di tanti contributi dovuti a sorgenti differenti, quindi per il TCL la $pdf(\epsilon)$ è gaussiana.

3 Distribuzioni di probabilità

PDF: descrive la popolazione dicendo quanto è la densità di frequenza associata ad ogni valore di x .

$$P(a < x < b) = \left(\int_a^b pdf(x) dx \right) \leq 1 \quad (8)$$

- CDF: è la primitiva della pdf, e restituisce una probabilità

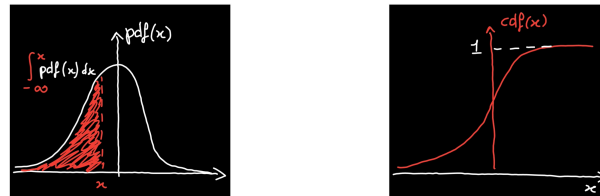


Figura 1: Relazione tra pdf e cdf

- Media (μ): valore di aspettazione di $x \Rightarrow E[x] = \mu$
- Varianza (σ^2): valore di aspettazione di $(x - \mu)^2$, è lo scarto quadratico medio sulla popolazione $\sigma^2 = E[x^2] - \mu^2$
- deviazione standard (σ): misura la larghezza della pdf
- Moda: massimo della pdf (potrebbero essercene più di una \Rightarrow pdf multimodale)
- Mediana: punto che divide a metà l'area della pdf

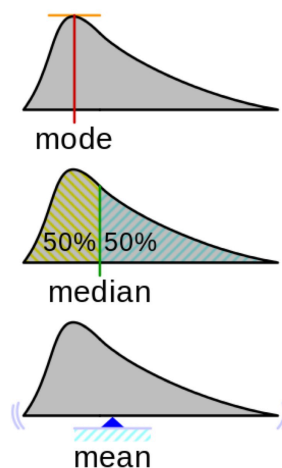


Figura 2: Stime di tendenza centrale

3.1 Momenti

$$E[x^m] = \int_a^b x^m pdf(x) dx$$

Momento di ordine 1: $E[x] = \mu$

Momento centrale: valore di aspettazione di $(x - \mu)^m$

- Ordine 1 \Rightarrow nullo
- Ordine 2 \Rightarrow varianza \Rightarrow larghezza
- Ordine 3 \Rightarrow legato al parametro γ_1 skewness (obliquità): $\gamma_1 = \frac{E[(x-\mu)^3]}{\sigma^3} \Rightarrow$ asimmetria
- Ordine 4 \Rightarrow legato al parametro γ_2 kurtosi: $\gamma_2 = \frac{E[(x-\mu)^4]}{\sigma^4} - 3 \Rightarrow$ quanto è piccata la curva



Figura 3: Skewness e kurtosi

RIPRODUTTIVITÀ: siano x e y due r.v. distribuite secondo stessa pdf, se la r.v. somma è distribuita secondo la pdf di partenza, gode di proprietà riproduttiva

3.2 Distribuzioni notevoli

Pdf caratterizzata da una forma che comprende:

- variabile indipendente x che assume valori nel dominio
- parametri α_i che possono essere espressi in termini dei momenti della pdf
- $\Rightarrow pdf(x; \alpha_1, \alpha_2, \dots)$

3.3 r.v. continue

DISTRIBUZIONE UNIFORME

Estraggo un numero casuale x compreso fra a e b . Ogni numero ha la stessa probabilità di essere estratto: $pdf(x) = k$.

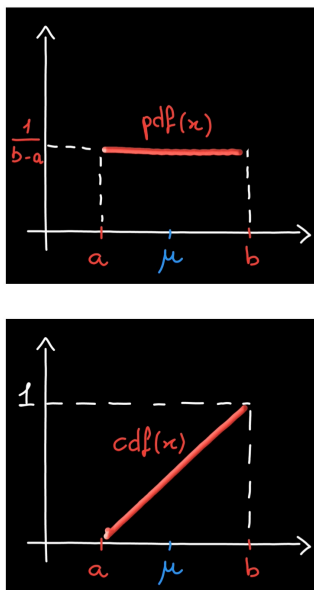


Figura 4: Rappresentazione

- **pdf:** $U(x; a, b) = \frac{1}{b-a}$
- **media:** $E[x] = \frac{a+b}{2}$
- **varianza:** $\text{Var}[x] = \frac{(b-a)^2}{12}$
- **cdf:** $\int_a^x \frac{dx}{b-a} = \frac{1}{b-a} (x-a)$
- **riproduttività:** non vale

DISTRIBUZIONE NORMALE (Gauss)

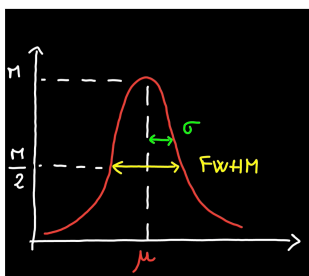


Figura 5: Rappresentazione

- **pdf:** $\text{Gauss}(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
- **media:** $E[x] = \mu$

- **varianza:** $\text{Var}[x] = \sigma^2$
- **skewness:** $\gamma_1 = 0$
- **kurtosis:** $\gamma_2 = 0$
- **cdf:** \nexists
- **riproduttività:** vale

La probabilità di un intervallo centrato in μ e largo $\pm 1\sigma$ è 68.27%.

La probabilità associata a 1 FWHM (full width at half maximum), parametro che vale 2.35σ , è di 98%.

GAUSSIANA STANDARDIZZATA (ricavata da gauss per cambio variabili, con $y = \frac{x-\mu}{\sigma}$)

- **pdf:** $\text{Gauss}(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{(x)^2}{2}}$
- **media:** $\mu = E[x] = 0$
- **varianza:** $\sigma^2 = \text{Var}[x] = 1$
- **skewness:** $\gamma_1 = 0$
- **kurtosis:** $\gamma_2 = 0$
- **cdf:** $\text{Erf}(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(x)^2}{2}} dx$
- **riproduttività:** vale

CHAUCHY

Riconducibile alla creazione e decadimento veloce di una particella elementare, essa è definita 'patologica': i suoi momenti non esistono.

- **pdf:** $\text{BW}(x; \alpha, x_0) = \frac{1}{\pi\alpha} \cdot \frac{\alpha^2}{(x+x_0)^2+\alpha^2}$
- **momenti:** \nexists
- **moda, mediana:** x_0

LOGNORMALE (ricavata da gaussiana per cambio di variabili, con $y = e^x$)

- **pdf:** $\text{pdf}(y; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \cdot \frac{1}{y} e^{-\frac{(\log(y)-\mu)^2}{2\sigma^2}}$
- **media:** $E[y] = e^{\mu+\frac{\sigma^2}{2}}$
- **varianza:** $\text{Var}[y] = e^{2\mu+2\sigma^2} \cdot (e^{\sigma^2} - 1)$

3.4 r.v. discrete

BINOMIALE

È il caso più semplice, descrive la probabilità di k successi in N prove, dove le prove sono Bernoulli trials (possibili solo due eventi complementari: successo o insuccesso) con probabilità di successo p

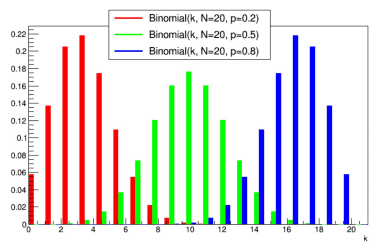


Figura 6: Rappresentazione

- **pdf:** $B(k, N, p) = \binom{N}{k} \cdot p^k (1-p)^{(N-k)}$
- **media:** $\mu = E[k] = Np$
- **varianza:** $\sigma^2 = \text{Var}[k] = Np(1-p)$
- **skewness:** $\gamma_1 \rightarrow 0$ quando $N \rightarrow \infty$
- **kurtosis:** $\gamma_2 \rightarrow 0$ quando $N \rightarrow \infty$

POISSONIANA

Descrive eventi indipendenti che avvengono in maniera casuale nel tempo (o spazio, o...) con frequenza media costante (indipendente da tempo, o spazio, o...). Utilizzata per descrivere eventi 'rari' come i decadimenti radioattivi all'interno di un campione.

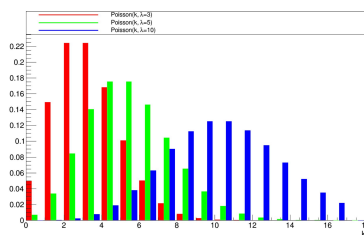


Figura 7: Rappresentazione

- **pdf:** $\text{Pois}(k; \lambda) = \frac{e^{-\lambda} \cdot \lambda^k}{k!}$ (probabilità di contare k eventi in un intervallo di tempo Δt unitario)

- **media:** $\mu = E[k] = \lambda$
- **varianza:** $\sigma^2 = \text{Var}[k] = \lambda$ (all'aumentare della media aumenta anche la varianza!)
- **skewness:** $\gamma_1 = \frac{1}{\sqrt{\lambda}}$
- **kurtosis:** $\gamma_2 = \frac{1}{\lambda}$
- **riproduttività:** vale

Si può ricavare la distribuzione in due modi: o partendo da una binomiale dove $p \rightarrow 0$ con $N \cdot p$ finito, o dimostrando che l'intervallo di tempo tra due eventi segue distribuzione esponenziale.

ESPONENZIALE (τ = tempo medio tra due eventi poissoniani)

- **pdf:** $\text{pdf}(t; \lambda) = \lambda \cdot e^{-\lambda t}$
- **pdf:** $\text{pdf}(t; \tau) = \frac{1}{\tau} \cdot e^{-\frac{t}{\tau}}$
- **media:** $\mu = E[t] = \tau = \frac{1}{\lambda}$
- **varianza:** $\sigma^2 = \text{Var}[t] = \tau^2 = \frac{1}{\lambda^2}$
- **skewness:** $\gamma_1 = 2$
- **kurtosis:** $\gamma_2 = 6$

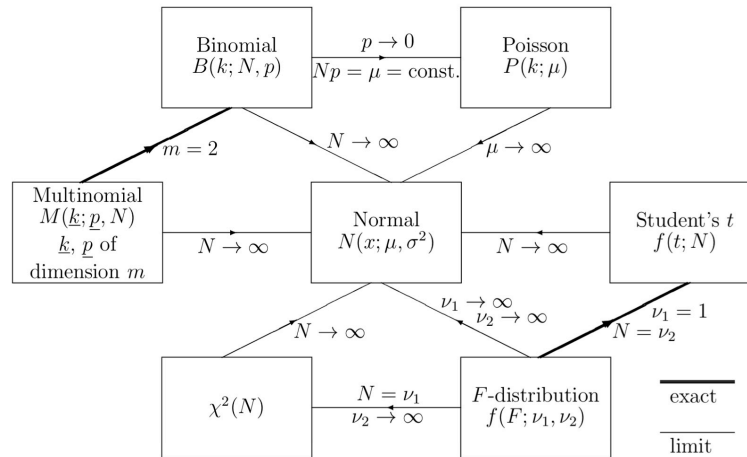


Figura 8: Comportamenti asintotici

3.5 Joint-pdf

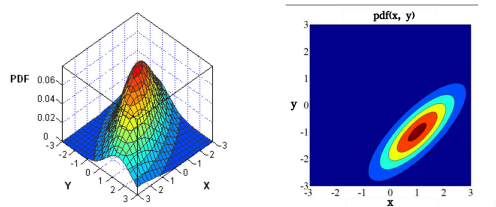


Figura 9: Probabilità congiunta (joint-pdf)

Quando un evento è identificato da un vettore $\vec{x} = x_1 \dots x_N$ parliamo di joint-pdf. Essa è definita per estensione di quanto fatto nel caso 1dimensionale, tuttavia la varianza è sostituita dalla covarianza.

COVARIANZA: matrice $n \times n$ simmetrica definita $\sigma_{ij}^2 = E[(x_i - \mu_i)(x_j - \mu_j)]$; i termini sulle diagonali sono l'equivalente della varianza: $\sigma_{ii}^2 = E[(x_i - \mu_i)^2]$.

Per le 2-dimensioni (x, y) si definiscono le probabilità:

- congiunta (o joint) ($pdf(x, y)$)
- marginale ($pdf_M(x), pdf_M(y)$), indipendente dai valori assunti dall'altra variabile
- condizionata ($pdf(x|y = y_0)$), pdf associata a x quando y ha un valore specifico.

Se due eventi x, y sono indipendenti la loro covarianza è nulla, $Cov(x, y) = 0$, tuttavia la covarianza nulla non è sufficiente a garantire l'indipendenza (ossia $pdf(x, y) = pdf(x) \cdot pdf(y)$)

4 PROPAGAZIONE DEGLI ERRORI

4.1 Cambio di variabili 1-D

x r.v. descritta da $pdf_1(x)$.

$y = y(x)$ monotona $\Rightarrow pdf_2(y) = pdf_1(x) \cdot |x'(y)|$.

Motivo: sia dx un'intervallo infinitesimo per x e dy per y , ai due intervallo deve essere associata la stessa probabilità: $pdf_1(x) \cdot dx = pdf_2(y) \cdot dy$

4.2 Propagazione 1-D

Misuro la grandezza X , ma mi interessa Y che è una sua funzione. Voglio passare dalla stima di X a Y .

- **Caso lineare:** $y(x) = ax + b \Rightarrow \mu_y = a \cdot \mu_x + b; \sigma_y^2 = a^2 \cdot \sigma_x^2$

- **Caso non lineare:** $y(x) = f(x)$ uso lo sviluppo di Taylor della funzione intorno a μ_x :

$$y(x) = y(\mu_x) + \left. \frac{dy}{dx} \right|_{\mu_x} (x - \mu_x) + \frac{1}{2} \left. \frac{d^2y}{dx^2} \right|_{\mu_x} (x - \mu_x)^2 + \dots \quad (9)$$

$$\text{MEDIA: } \mu_y \simeq y(\mu_x) + \left. \frac{1}{2} \frac{d^2y}{dx^2} \right|_{\mu_x} \cdot \sigma_x^2$$

$$\text{VARIANZA: } \sigma_y^2 = E[y^2] - \mu_y^2 \simeq \left(\left. \frac{dy}{dx} \right|_{\mu_x} \right)^2 \cdot \sigma_x^2$$

4.3 Cambio di variabili n-D

Si procede in modo analogo: $pdf_x(\vec{x}) \cdot d\vec{x} = pdf_y(\vec{y}) \cdot d\vec{y}$, da cui:

$$pdf_y(\vec{y}) = pdf_x(\vec{x}) \cdot |J(\vec{w})| \quad (10)$$

dove J è il determinante della matrice jacobiana della funzione $w(\vec{y}) = \vec{x}$, data da: $J_{i,j} = \frac{\partial w_i}{\partial y_j}$

4.4 Propagazione n-D

- MEDIA: $\mu_y = E[y(\vec{x})] \simeq y(E[\vec{x}] = f(E[x_1], \dots, E[x_N]))$
- VARIANZA: $\sigma_y^2 = Var(y) \simeq \sum_{i,j=1}^n \left. \frac{\partial y}{\partial x_i} \frac{\partial y}{\partial x_j} \right|_{E[\vec{x}]} \cdot \sigma_{ij}^2$

5 STIMA DI PARAMETRI

Ho a disposizione dei dati che sono r.v. provenienti dal campionamento di una pdf_x che dipende dal parametro θ che si vuole stimare. Costruisco una funzione dei campionamenti che possa stimare il parametro θ , la chiamiamo stimatore: $\hat{\theta} = f(x_1, \dots, x_n)$ a cui si associa un'incertezza $\delta_{\hat{\theta}}$.

5.1 Campionamenti IID

N campionamenti x_1, \dots, x_n che sono:

- indipendenti (non condizionati dai precedenti)
- identicamente distribuiti (estratti dalla stessa pdf)

5.2 Joint-pdf e likelihood

La joint-pdf di N campionamenti IID è:

$$pdf_{set}(x_1, \dots, x_n, \theta) = \prod_{i=1}^n pdf_x(x_i, \theta) \quad (11)$$

Questa misura la probabilità di estrarre uno specifico set di dati. Se vista come una funzione del parametro θ , si chiama likelihood. $L(\theta) = pdf_{set}(x_1 \dots x_N, \theta)$

5.3 Statistica

Una funzione di N campionamenti IID che contiene solo parametri noti: $f(x_1 \dots x_N)$ si chiama statistica. Essa è una variabile aleatoria, e come tale ha una sua pdf.

Abbiamo a che fare con 3 pdf:

- $pdf_x(x, \theta)$ campionata IID
- $pdf_{set}(x_1 \dots x_N, \theta)$ dei campionamenti
- pdf_f della statistica dei campionamenti

5.4 Stimatore

Statistica scelta in modo da poter usare N campionamenti IID per stimare il valore dei parametri della pdf_x . Notazione:

- $pdf_x(x, \theta)$ è la pdf della quale stimare il parametro
- $x_1 \dots x_N$ sono i campionamenti
- $\hat{\theta}(x_1 \dots x_N)$ è lo stimatore
- $pdf_{\hat{\theta}}(\hat{\theta})$ è la pdf dello stimatore
- $\hat{\theta}^*$ è il valore dello stimatore ottenuto per uno specifico campionamento (quindi per uno specifico set di dati)

CONSISTENZA: uno stimatore è consistente se per $N \rightarrow \infty$ restituisce il valore vero del parametro. Quindi: $\lim_{N \rightarrow \infty} \hat{\theta}_N = \theta_V$.

Uno stimatore, in quanto r.v. è caratterizzato da un valore medio $E[\hat{\theta}]$ e varianza $Var(\hat{\theta})$.

BIAS: misura quanto la stima sia capace di restituire un risultato prossimo a quello vero. Per ogni set di campionamenti ho stime diverse, quindi devo richiedere che la media delle stime sia il valore vero. Il bias è $b_N = E[\hat{\theta}_N] - \theta_{Vero}$. Se $b_N = 0$ lo stimatore è unbiased. Il bias ha a che fare con l'ACCURATEZZA.

VARIANZA: ripetendo i campionamenti vogliamo che le stime siano vicine fra loro. L'efficienza rappresenta questa vicinanza, vogliamo che sia minima e ha a che fare con la PRECISIONE.

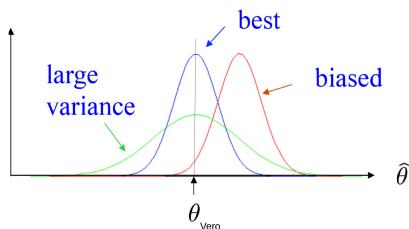


Figura 10: Varianza e Bias

5.5 Stime di μ e σ

Lo stimatore per $\mu = E[x]$ è la media campionaria \bar{x} .

$$\hat{\mu}_N = \bar{x} = \frac{1}{N} \sum_{i=1}^N x_i.$$

- è unbiased: $E[\bar{x}] = \mu$
- è efficiente (la varianza decresce all'aumentare di N): $Var[\bar{x}] = \frac{\sigma^2}{N}$
- se la pdf soddisfa le ipotesi del TCL, la $pdf_{\bar{x}}$ per $N \rightarrow \infty$ tende a una gaussiana con media μ e varianza σ/N .

Lo stimatore per σ^2 è la sample variance. Il bias di $s_{\bar{x}}^2$ è diverso da zero; applico quindi la correzione di Bessel, e ottengo: $s^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2$

- la varianza generalmente non può essere determinata, a meno che $pdf_x(x)$ sia una gaussiana, in tal caso si introduce una variabile ausiliaria chiamata χ^2 .
- $Var[\chi_N^2] = 2N \Rightarrow Var[s^2] = \frac{2\sigma^4}{(N-1)}$

5.6 Confidenza

Dato che μ si stima con \bar{x} , che ha una varianza $Var[\bar{x}] = \frac{\sigma^2}{N}$, asintoticamente \bar{x} ha una pdf gaussiana centrata su μ e larga $\frac{\sigma}{\sqrt{N}}$. Possiamo associare alla stima un intervallo $\bar{x} \pm \frac{\sigma}{\sqrt{N}}$.

La probabilità di trovare un \bar{x} che stia nell'intervallo $\mu \pm \frac{\sigma}{\sqrt{N}}$ è 68%. Il problema si complica quando σ non è nota, e devo utilizzare la sua stima s . In tal caso, la probabilità dipende non solo da $pdf(\bar{x})$, ma anche da $pdf(s^2)$.

Ricorro a una variabile ausiliaria: $t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{N}}}$.

L'intervallo $\mu \pm \frac{s}{\sqrt{N}}$ corrisponde all'intervallo per $t \pm 1$. Quando s^2 è distribuito come un χ^2 ($pdf_x(x)$ gaussiana) la distribuzione t è nota ed è la distribuzione di student.

5.7 Distribuzione χ^2

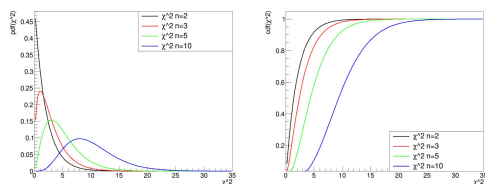


Figura 11: Rappresentazione

La pdf chi-quadro è descritta da un solo parametro chiamato gradi di libertà. Una r.v. χ^2 si costruisce estraendo N valori IID ciascuno da una pdf_i gaussiana di media μ_i e varianza σ_i^2 .

$$\chi^2 = \sum \frac{(x_i - \mu_i)^2}{\sigma_i^2} \quad (12)$$

La forma analitica della pdf della r.v. χ_N^2 è:

$$pdf(\chi_N^2) = (\chi^2)^{\frac{N}{2}-1} \cdot \frac{e^{-\chi^2/2}}{\Gamma(N/2)} \cdot 2^{N/2} \quad (13)$$

dove $\Gamma(N/2)$ è un'estensione del fattoriale per valori non interi.

- **media:** $E[k] = N$
- **varianza:** $Var[k] = 2N$
- **moda:** $N-2$
- **mediana:** è il punto dove la cdf vale 0.5, e vale 1.4 per $N=2$, 4.3 per $N=5$, 9.4 per $N=10$...
- **riproduttività:** vale

È spesso utile definire χ^2 ridotto, ossia χ_N^2/N , di media 1.

5.8 Errore quadratico medio

In generale:

- la stima è il valore assunto da $\hat{\theta}$ per uno specifico campionamento $\hat{\theta}^*$
- l'incertezza della stima è legata a $Var[\hat{\theta}]$
- la distanza tra stima e valor vero è il bias $b = E[\hat{\theta}] - \theta_V$
- si definisce errore quadratico medio la combinazione di errore statistico (varianza) con errore sistematico (bias): $E[(\hat{\theta} - \theta_V)^2] = Var[\hat{\theta}] + b^2$

6 MAXIMUM LIKELIHOOD

La funzione likelihood è definita come la probabilità di osservare il campione di dati IID \vec{x} , condizionata al valore assunto dal parametro θ oggetto di stima. Essa contiene sia il modello (quindi la pdf) che i dati raccolti.

$$L(\theta) = L(x_1 \dots x_N | \theta) = \prod_{i=1}^N pdf_x(x_i, \theta) \quad (14)$$

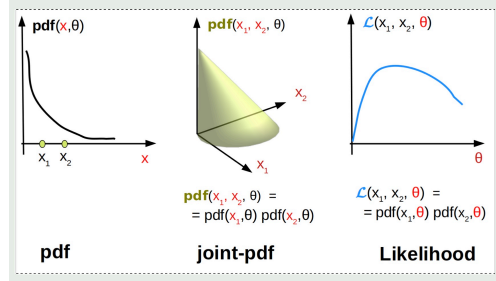


Figura 12: Differenza tra Joint e likelihood

6.1 Informazione

I campionamenti possono fornire o meno informazioni su θ : se la likelihood è piatta vorrà dire che avrò circa la stessa probabilità di ottenerli a prescindere dal valore di θ , se è a campana invece trovo delle differenze (larghezza della campana legata all'incertezza con cui posso determinare valore vero di θ). L'informazione consente di valutare la MINIMA VARIANZA raggiungibile da uno stimatore di $\hat{\theta}$.

Informazione di Fischer:

$$I_{\hat{x}}(\theta) = E \left[- \frac{\partial^2 \ln L(\vec{x}; \theta)}{\partial \theta^2} \right]_{\theta=\theta_v} \quad (15)$$

DISUGUAGLIANZA DI RAO-CRAMER o MVB:

$$Var[\hat{\theta}] \geq \frac{[1 + \frac{\partial b}{\partial \theta}]^2}{I(\theta)} \quad (16)$$

Se uno stimatore è asintoticamente unbiased, $Var[\hat{\theta}] \geq \frac{1}{I(\theta)}$

6.2 Max likelihood

Identifica lo stimatore per θ come $\hat{\theta}_{ML}$ in corrispondenza del quale la probabilità associata a un campionamento è quella maggiore possibile. Poichè L è un prodotto di N fattori, considero il logaritmo che diventa: $\sum \ln[pdf_x(x_i, \dots, \theta)]$. Ne devo annullare la derivata prima e assicurarmi che la derivata seconda, pari a $1/\sigma_{ML}^2 = I(\theta)$, sia negativa.

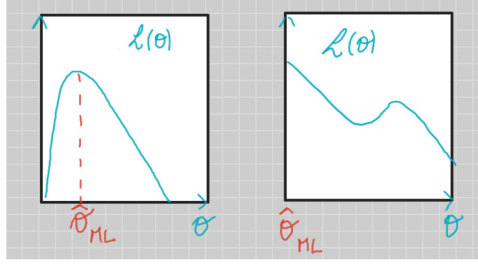


Figura 13: Stimatore massima verosimiglianza

6.3 Likelihood e gaussiana

Per $N \rightarrow \infty$, la likelihood è asintotica a una funzione gaussiana, e può essere scritta come

$$L(\theta) = L_{Max} \cdot e^{\left(\frac{\theta - \hat{\theta}_{ML}}{2\sigma_{ML}^2}\right)}.$$

7 LEAST SQUARES

Sceglie come stima di $\hat{\theta}$ il valore per cui è minimo:

$$Q^2(\hat{\theta}) = \sum_{i=1}^N \frac{[y_i - f(\hat{\theta}, x_i)]^2}{\sigma_i^2} \quad (17)$$

Q^2 è la somma dei quadrati delle distanze fra i punti campionati e la funzione $Y = f(\hat{\theta}, X)$. In questo caso si cerca un valore $\hat{\theta}_{LQ}$ per cui Q^2 ammetta un minimo assoluto.

7.1 LQ vs ML

LQ: cerco i valori dei parametri che rendano minima la distanza fra dati campionati e modello.
ML: cerco i valori dei parametri che rendano massima la probabilità, dato un modello, di osservare i dati campionati.

7.2 Caso gaussiano

Quando le $pdf_i(y_i)$ sono gaussiane, non solo i due stimatori coincidono, ma sappiamo anche che $f(x_i, \hat{\theta}_{LQ}) \simeq E[y_i] = \mu_i \Rightarrow Q_{min}^2 = \sum_{i=1}^N \frac{(y_i - \mu_i)^2}{\sigma_i^2}$, che è l'espressione di un chi quadro a N-K gradi di libertà.

Di conseguenza $E[Q_{min}^2] = N-K$, e sfruttando MVB, $\sigma_{\hat{\theta}_{LQ}}^2 = \frac{2}{-\left[\frac{\partial^2 Q^2(\vec{x}; \theta)}{\partial \theta^2}\right]_{\theta = \hat{\theta}_{LQ}}}$

7.3 Teorema Gauss-Markov

Due grandezze X e Y sono legate da una relazione lineare nei K parametri descritti dal vettore θ : $Y = \sum_{j=1}^K \theta_j \cdot h_j(X)$. i dati sono N misure indipendenti (x_i, y_i) , dove x_i sono privi di errore, e y_i hanno incertezza σ_i nota: $E[y_i] = \sum_{j=1}^K \theta_j \cdot h_j(x_i)$.

Definisco delle variabili ausiliarie ϵ_i con stessa pdf delle y_i ma traslate di $E[y_i]$, per cui con $E[\epsilon_i] = 0$. Queste rappresentano l'errore statistico, posso dunque scrivere: $y_i = \sum_{j=1}^k \theta_j \cdot h_j(x_i) + \epsilon_i$.

Scegliendo una rappresentazione matriciale, $\vec{y} = H(\vec{x}) \cdot \vec{\theta} + \vec{\epsilon}$.

In sostanza, ho scritto y_i come la somma tra il valore vero restituito dal modello in corrispondenza di x_i e l'errore di misura ϵ_i .

La pdf (N-dimensionale) di \vec{y} è associata a una matrice di covarianza diagonale V (composta da σ_i^2): $Cov[\vec{y}] = Cov[\vec{\epsilon}] = V$. In questo modo, $Q^2(\vec{\theta}) = \vec{\epsilon}^T \cdot V^{-1} \cdot \vec{\epsilon} = \sum_{i=1}^N \frac{\epsilon_i^2}{\sigma_i^2}$. Il valore ricavato per la stima è $\hat{\theta}_{LS} = (H^T V^{-1} H)^{-1} \cdot H^T V^{-1} \vec{y}$, che ha:

- $E[\hat{\theta}_{LS}] = \vec{\theta}$ (unbiased)
- $Var[\hat{\theta}_{LS}] = (H^T V^{-1} H)^{-1}$

Il teorema dice che, quando il modello è lineare nei parametri, e i campionamenti si possono scrivere come $y_i = f(x_i) + \epsilon_i$ (dove $E[\epsilon_i] = 0$ e $Var[\epsilon_i]$ finita indipendente dai parametri), allora lo stimatore LS è unbiased ed è quello con varianza minima.

7.4 Propagazione dell'errore delle x su y

Se un modello è una retta $Y = a + bX$, dove sia le x_i che le y_i sono provviste di errore, posso propagare sulle y l'errore delle x: $\sigma_i^2 = \sigma_{yi}^2 + b^2 \sigma_{xi}^2$.

7.5 Estrapolazione

Un problema che ci si pone spesso è quello di valutare il modello in un punto differente da quelli misurati sperimentalmente. Se abbiamo un modello $Y = f(\vec{\theta}, X)$, mediante i valori campionati ($\theta_1 \dots \theta_K$) si stimano i parametri che descrivono meglio il modello $\Rightarrow \hat{\theta}_{LS}$.

Si valuta la funzione in un nuovo punto x_0 : $y_0 = f(\hat{\theta}_{LS}, x_0)$.

Come valuto l'errore su y_0 ? propago su $\hat{\theta}_{LS}$ mediante la matrice di covarianza! nel caso lineare:

$$Var[y_0] = H(x_0) \cdot Cov[\hat{\theta}_{LS}] \cdot H(x_0)^T \quad (18)$$

dove H è la matrice $N \times K$ composta dai $h_i(x_j)$, Cov è la matrice covarianza.

8 FIT DI ISTOGRAMMI

Abbiamo N campionamenti IID di una pdf(x, θ) e vogliamo usarli per stimare θ . Esistono due percorsi:

- unbinned data (uso i dati campionati), quello che abbiamo praticamente visto finora
- binned data (raccolgo i dati in un istogramma, perdendo parte dell'informazione)

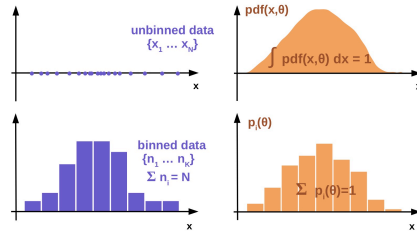


Figura 14: Differenza tra dati binned e unbinned

8.1 Binned data

Scelgo di lavorare su dati raggruppati in classi di frequenza (bin).

DATI: K interi che identificano i conteggi in ciascun bin n_1, \dots, n_k

MODELLO: pdf trasformata in un istogramma p_1, \dots, p_N

JOINT-PDF: multinomiale con parametri N, p_i : $\text{PDF}(n_1 \dots n_K, N; p_1(\theta) \dots p_k(\theta))$.

Perchè multinomiale? la probabilità che una misura cada in un bin è un bernoulli trial, quindi per un solo bin la distribuzione è binomiale (e ho molteplici bin).

- $E[n_i] = \mu_i = N \cdot P_i$
- $\text{Var}[n_i] = \sigma_i^2 = N \cdot P_i \cdot (1 - p_i) \simeq N \cdot p_i$
- $\text{Cov}(n_i, n_j) = -N \cdot p_i \cdot p_j \simeq 0$
- \simeq valgono se p_i è piccola, rendendo stretti i bin

Quando p_i piccola vale l'approssimazione per Poissoniana, di conseguenza la joint-pdf si scrive:

$$\prod \text{pdf}_i(n_i, p_i(\theta)) = \prod \frac{e^{-\mu_i} \mu_i^{n_i}}{n_i!} \quad (19)$$

dove $\mu_i = N p_i(\theta)$ è il valore di aspettazione di n_i .

Quindi conosco la pdf, il loro valore di aspettazione e la loro varianza \Rightarrow posso usare sia LS che ML.

8.2 Fit LS

Confronto l'istogramma dei dati con quello atteso.

$$Q^2 = \sum_{i=1}^k \frac{(y_i - E[y_i])^2}{\text{Var}[y_i]} \Rightarrow Q^2 = \sum_{i=1}^k \frac{(n_i - \mu_i(\theta))^2}{\mu_i(\theta)} \quad (20)$$

Se μ_i (quindi n_i) è grande posso approssimare poisson con gauss. Di conseguenza Q^2 è una r.v. χ^2 con K -s gradi di libertà (s = numero di parametri da stimare). Posso poi sostituire

all'errore atteso σ_i di n_i la sua stima $\sqrt{n_i}$. Il chi-quadro di Neyman è:

$$\chi_{Neyman}^2 = \sum_{i=1}^K K \frac{[O_i - E_i(\theta)]^2}{O_i} \quad (21)$$

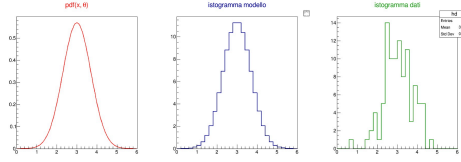


Figura 15: I due istogrammi confrontati

8.3 Fit ML

Se l'istogramma non può essere approssimato a gaussiana (ho pochi conteggi) devo usare la binned ML. Massimizzo $L(n_1 \dots n_k, \theta) = \prod \text{Bin}(N, n_i, p_i(\theta)) = \prod \frac{\mu_i^{n_i} e^{-\mu_i}}{n_i!}$

9 TEST DI IPOTESI

Confronto modelli con dati sperimentali per capire quale spiega meglio le osservazioni.

9.1 Null hypothesis

Voglio verificare la compatibilità dei dati con un'ipotesi H_0 , detta null hypothesis:

- ho un certo numero di dati sperimentali appartenenti a un sample space Ω
- sulla base di H_0 , identifico due sottoinsiemi di Ω : w è la regione critica (rigetto), $w^* = \Omega - w$ è la regione di accettazione.

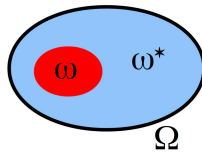


Figura 16: Sample space

C'è la possibilità che l'ipotesi sia vera ma la rigetto (falsi negativi): $P(\vec{x} \in w | H_0) = \alpha$. Quando i dati sono tanti, si opta per costruire una statistica $t(\vec{x})$ definendo la regione di accettazione a partire dal valore di α desiderato. Si hanno due possibili test: one sided (una coda), two sided (entrambe le code).

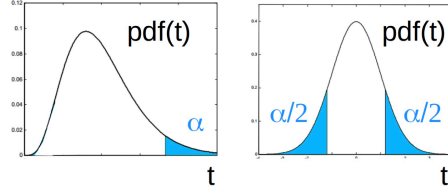


Figura 17: Sinistra: one sided, Destra: two sided

9.2 Alternative hypothesis

Se considero anche un'ipotesi alternativa, H_1 alternative, definita la statistica t avrò due distribuzioni: $pdf(t|H_0) \Rightarrow$ ottengo t quando vale H_0 , $pdf(t|H_1) \Rightarrow$ ottengo t quando vale H_1 . In questo caso possono anche esserci, oltre a falsi negativi, falsi positivi (probabilità β). È il caso in cui accetto H_0 quando è vera H_1 .

$$\alpha = \int_w pdf(t|H_0) dt \quad (22)$$

$$\beta = \int_{w^*} pdf(t|H_1) dt \quad (23)$$

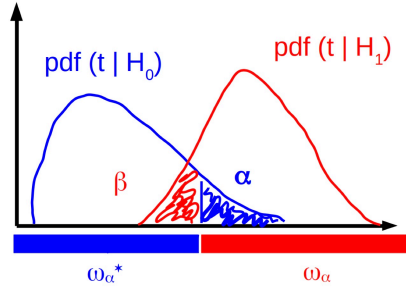


Figura 18: Falsi positivi e falsi negativi

9.3 Test del chi-2

Il valore del χ^2_{min} può essere usato per testare la validità del modello. Esso deve essere prossimo a N-K (o chi ridotto prossimo a 1).

P-VALUE: scelgo un criterio di rigettare l'ipotesi che corrisponde a un range di valori del chi-2 che hanno bassa probabilità di essere estratti; solitamente si sceglie la coda destra della distribuzione che corrisponde al 5%. Mi riduco quindi a chiedere:

$$p - value = \int_{\chi^2_{min}}^{\infty} pdf(\chi^2) d\chi^2 > \alpha = 0.05 \quad (24)$$

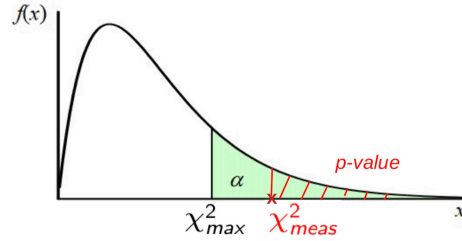


Figura 19: P-value

Attenzione: un valore di χ^2 vicino a 0 soddisfa il test! uso il test one sided (ossia solo una delle due code) perché le regioni a basso valore di chi2 corrispondono a scarto molto piccolo, quindi a un sovra-accordo (altamente improbabile).

9.4 Test di Kolmogorov

Abbiamo N misure ripetute di X e voglio testare l'ipotesi H_0 che siano campionamenti di una determinata pdf modello. Posso usare il test del chi-2 applicando a istogrammi costruiti con le misure raccolte, tuttavia non utilizzo tutta l'informazione perché binno i dati. Alternativa: confrontare le distribuzioni cumulative dei dati e della pdf modello.

Se chiamo $F(x)$ la distribuzione cumulative (discreta) dei dati, e $\Phi(x)$ la distribuzione cumulativa della pdf, posso confrontare con una quantità di riferimento δ_0 lo scarto massimo:

$$\delta = \text{Max}|F(x_i) - \Phi(x_i)| \quad (25)$$

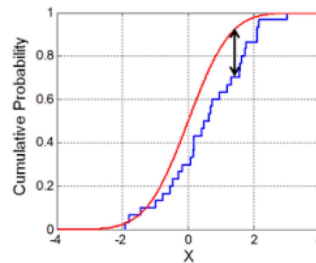


Figura 20: se $\delta > \delta_0$ l'ipotesi viene rigettata

9.5 Osservazione

Noto: quando si esegue un test di ipotesi, quindi quando si cerca di accettare l'ipotesi H_0 che i dati non siano in contraddizione col modello, non si ha mai una prova che il modello sia giusto, ma solo che non è contraddetto dai dati.