Laboatorio II: Statistica e probabilità

Marco Militello

Indice

Capitolo 1

Probabilità

1.1 Probabilità: definizioni e teoremi

Def (Evento casuale)
Associazione di una o più modalità ad un fenomeno casuale; risultato esperimento non prevedibile con certezza

- Ripetibile
- Diverse modalità mutualmente esclusive
- Singolarmente imprevedibili

Def (Variabile casuale)
Una variabile casuale è un numero reale da associare a una modalità di un fenomeno

- può essere discreta o continua
- Può assumere un numero finito o infinito

Def (Spazio campionario)
Insieme delle possibili modalità di un evento

Def (Popolazione)
Insieme dei possibili eventi; si può avere accesso a tutta la popolazione oppure solo ad una parte limitata (astrazione)

Def (Campione)
Insieme degli eventi casuali raccolti

Def (Sample-space)
Contiene tutti i possibili eventi; la probabilità associata è 1
Sottoinsieme $\rightarrow$ può contenere uno o più eventi; si usa notazione insiemistica

Teorema (Assiomi di Kolmogorov $\rightarrow$ definizione assiomatica probabilità)
Probabilità di una funzione

$$P:\ \Omega \rightarrow [0,1]$$

tale che:

- $p(A) \geq 0$

- $p(\Omega) = 1$
- $p(A \cup B) = p(A) + p(B)$ se $A \cap B = \emptyset$ (A e B disgiunti) $\rightarrow$ eventi mutualmente esclusivi

Proprietà:

- $p(A) = 1 - p(A^c)$ con A^c evento complementare
- $p(A) \leq 1$ con $p(A) = 1 \Rightarrow A$ evento certo e $p(A) = 0 \Rightarrow A$ evento impossibile
- $p(\emptyset) = 0$
- $B \subset A \Rightarrow p(A) \geq p(B)$
- $p(A \cup B) = p(A) + p(B) - p(A \cap B)$

Def (probabilità classica)
Rapporto casi favorevoli su casi possibili se i casi sono ugualmente probabili

Def (probabilità frequentista)

$$p(A) \sim \lim_{n \to \infty} \frac{n(A)}{N} \qquad N \text{ campione elementi, } n(A) \text{ numero successi}$$

Teorema (Bernoulli, legge dei grandi numeri)

$$\lim_{N \to \infty} \frac{n(A)}{N} = p(A)$$

Al crescere del numero di prove la frequenza relativa di qualunque evento converge alla probabilità dell'evento

- Statistica frequentista $\rightarrow$ approccio limitante: associa probabilità all'accadare di un evento
- Approccio Bayesiano $\rightarrow$ Concetto plausibilità di un evento; conoscenza a priori: informazioni a priori

Def (Probabilità condizionata)
Nuova funzione di probabilità $\quad P:\ \Omega \rightarrow [0,1]$, A e B non disgiunti tra loro

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \qquad \text{B: varibile indipendente; A: parametro}$$

Probabilità che si verifichi B sapendo che si è verificato A

Def (Eventi indipendenti)
N eventi sono statisticamente indipendenti se la probabilità di un evento non è influenzata da uno o più eventi già verificati

$$p(A \cap B) = p(A) \cdot p(B)$$
$$p(B|A) = p(B|\Omega) = p(B) \text{ e } p(A|B) = p(A|\Omega) = p(A) \Rightarrow \text{ statisticamente indipendenti}$$

Teorema (Bayes)

$$p(A|B) = \frac{p(B|A)}{p(B)} \cdot p(A) \rightarrow p(\text{teoria} \,|\, \text{risultato}) = \frac{p(\text{risultato} \,|\, \text{teoria})}{p(\text{risultato})} \cdot p(\text{teroria})$$
$$p(A_j|E) = \frac{p(A_j) \cdot p(E|A_j)}{\sum_j [p(A_j) \cdot p(E|A_j)]} \qquad A_j \text{ mutualmente esclusivi}$$

Def (Variabile aleatoria)
Numero associato all'evento
$P : \Omega \to [0, 1]$

- Sample-space sottoinsieme di $\mathbb{N}$ $\qquad p(k) > 0 \; \forall k$
- $p(k)$ =probabilità valore k $\qquad \sum P_k = 1$

Def (Random variable continua)
evento identificabile con numero reale x a cui viene associata una densità di probabilità

$$p(a < x < b) \qquad \text{intervallo}(a, b)$$

- il sample-space è Ω è un sottoinsieme di $\mathbb{R} \to \Omega \subseteq \mathbb{R}$
- la probabilità di uno specifico valore è infinitesima $\Rightarrow p(x) = 0$ (Ω contiene infiniti elementi)
- la probabilità che viene associata a un gruppo di Ω

Rappresentazione $\to$ ISTOGRAMMA (classi di frequenza $f = \frac{n}{N}$)

- L'area di un bin è la probabilità che misura cada in quel bin
- integrale istogramma è 1
- il bin-size può essere ridotto $\to$ se il bin-size diventa infinitesimo $\Rightarrow$ distibuzione continua pdf

Capitolo 2

Statistica

Def (Densità di probabilità)
funzione pdf: $\Omega \subseteq \mathbb{R} \to \mathbb{R}^+$ (non è una probabilità, ma una densità di probabilità)

$$P(a < x < b) = \int_a^b pdf(x)\, dx \leq 1$$

- $pdf(x) > 0\ \forall x$
- $\int_\Omega pdf(x)\, dx = 1$
- $P(a < x < b) = \int_a^b pdf(x)\, dx$

Def (Densità di probabilità cumulativa)
É una probabilità e indica la probabilità del punto x

$$cdf = \int_a^x pdf(x)\, dx$$

cdf è la primitiva della pdf $\to pdf(x) = \frac{d}{dx} cdf(x)$

Def (Valore di aspettazione)
Media della funzione usando come peso la $pdf(x)$

$$E(u(x)) = \int_\Omega u(x) \cdot pdf(x)\, dx$$

- $E(u(x) + v(x)) = E(u(x)) + E(v(x))$
- $E(k \cdot v(x)) = k \cdot E(u(x))$

2.1 Stime di tendenza centrale

2.1.1 Media aritmetica

Def (Media μ)
Valore di aspettazione di x

$$E(x) = \int x \cdot pdf(x)\, dx = \mu$$

1. Somma scarti della media è nulla

2. $\overline{x}$ rende minima $\sum_{i=1}^{N}(x_i - x)^2$

3. media $\mu \neq$ media campionaria $\left(\overline{x} = \frac{1}{N}\sum_{i=1}^{N} x_i\right)$

4. La media aritmetica di un insieme di dati numerici è quel valore che rende minima la somma dei quadrati degli scarti dalle x_i

La media aritmetica è la stima migliore, ha un errore inferiore di quello delle misure stesse

2.1.2 Moda

Def (Moda)
Tendenza centrale del campione; massimo della pdf: può essere più di uno $\rightarrow$ valore che si è presentato più volte

- Distribuzioni amodali (no massimo)
- Distibuzioni multimodali (più massimi)

Talvolta si dice che la distribuzione non ha moda anche se presenta un massimo, quando quest'ultimo si trova in uno degli estremi dell'intervallo che contiene le misure

2.1.3 Mediana

Def (Mediana)
Punto che divide area pdf(x) in due metà: è il valore centrale se il numero di misure è dispari, altrimenti è la semisomma dei valori centrali; la mediana esiste sempre

Media, moda e mediana sono stime di tendenza centrale

2.2 Momenti

Momento di ordine m della pdf $\rightarrow$ valore di aspettazione di x^m

$$E(x^m) = \int x^m \cdot pdf(x)\, dx$$

- momento di ordine 1 $\rightarrow$ media

Momenti centrali di ordine m della pdf $\rightarrow$ valore di aspettazione di $(x-\mu)^m$

- momento centrale di ordine 1 $\rightarrow E(x-\mu) = 0$
- momento centrale di ordine 2 $\rightarrow$ Varianza o scarto quadratico medio σ^2
  Valore di aspettazione di $(x-\mu)^2$

$$E((x-\mu)^2) = \int (x-\mu)^2 \cdot pdf(x)\, dx = \sigma^2$$

σ = deviazione standard ed è legato alla larghezza della pdf(x)

$$\sigma^2 = E(x^2) - \mu$$

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - x)^2 = \frac{1}{N}x_i^2 - \bar{x}^2$$

- momento centrale di ordine 3 $\rightarrow$ legato a parametro SKEWNESS (misura asimmetria curva)

$$\gamma_1 = \frac{E((x-\mu)^3)}{\sigma^3} \qquad \begin{cases} \gamma_1 = 0 & \text{simmetrica} \\ \gamma_1 < 0 & \text{destra} \\ \gamma_1 > 0 & \text{sinistra} \end{cases}$$

- momento centrale di ordine 4 $\rightarrow$ legato a parametro KURTOSI (misura quanto è piccata la curva)

$$\gamma_2 = \frac{E((x-\mu)^4)}{\sigma^4} - 3 \qquad \begin{cases} \gamma_2 = 0 & \text{gaussiana} \\ \gamma_2 > 0 & \text{+ piccata} \\ \gamma_2 < 0 & \text{- piccata} \end{cases}$$

Se esistono tutti i momenti centrali $\Rightarrow$ caratterizzano in modo univoco la pdf; se due variabili casuali hanno stessi momenti fino a qualsiasi ordine, allora la loro densità di probabilità è identica

Def (Riproduttività)
Siano x,y random variable distribuite secondo stessa pdf; se $z = x + y$ (z random variable) ha la stessa pdf $\Rightarrow$ pdf gode della proprietà riproduttiva
La $pdf(x; \alpha_1, \alpha_2, \dots)$ è caratterizzata da:

- una forma funzionale
- uno o più parametri

2.3 Distibuzioni

Distribuzione gaussiana

$$pdf(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Gode della proprietà riproduttiva
FWHM(largezza a mezza altezza) $\sim 2.35\sigma$
Distribuzione gaussiana standardizzata $\rightarrow \mu = 0$ $\sigma = 1$

$$G(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$$

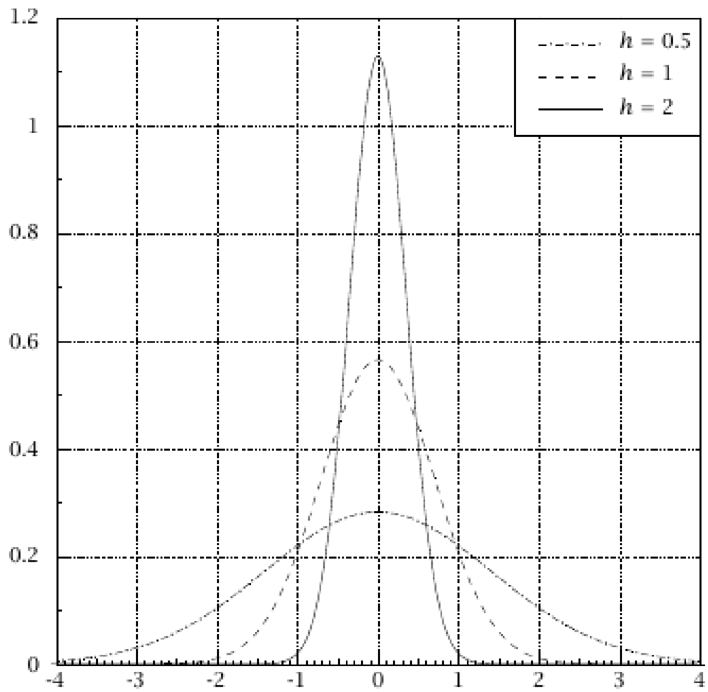
$$cdf(x) = \int_{-\infty}^{x} G(x) \rightarrow erf(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{x^2}{2}}$$

$erf(x)$ non ha una definizione univoca

- media: $\quad E(x) = \mu$

- varianza: $Var(x) = \sigma^2$
- skewness: $\gamma_1 = 0$
- kurtosi: $\gamma_2 = 0$

Le misure affette da errori casuali hanno una probabilità del 68% di cadere all'interno di un intervallo di semiampiezza σ centrato sul valore vero della grandezza; l'intervallo ha quindi il 68% di probabilità di contenere il valore vero (se gli errori sono casuali e normali)
L'errore quadratico medio σ può essere interpretato come valore assoluto delle ascisse dei due punti di flesso della Gaussiana

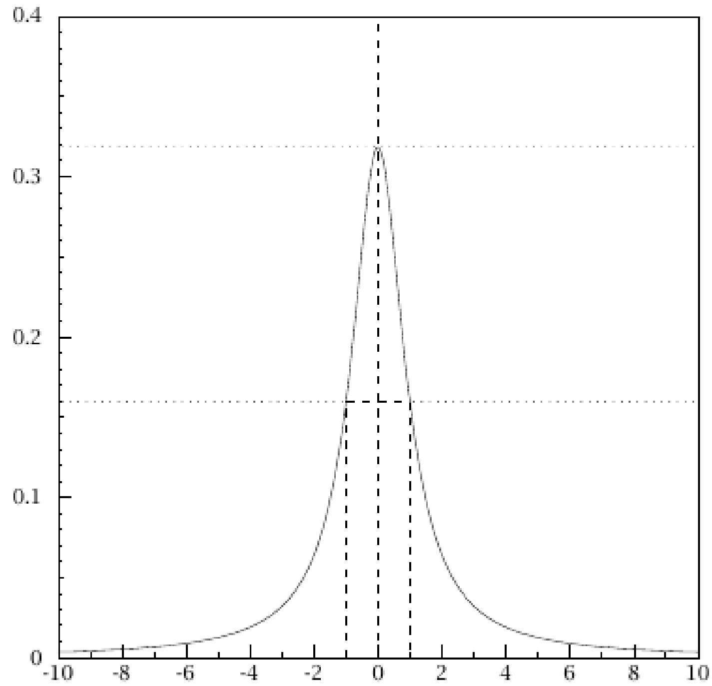


Distribuzione Breit-Wigner (o Cauchy)

$$BW(x; \alpha, M) = \frac{1}{\pi\alpha} \frac{1}{1 + \frac{(x-M)^2}{\alpha^2}}$$

es: fisica nucleare
Nessuno dei momenti centrali esistono, nemmeno la media; M è la mediana della distribuzione e α è la larghezza a metà altezza
Il valor medio di un campione proveniente da una popolazione che segua distribuzione di Cauchy è distribuito anch'esso secondo Cauchy

FIGURA 8d - L'andamento della distribuzione di Cauchy, per $\theta = 0$ e $d = 1$.



Def
Rappresentazione con distribuzione di probabilità; si estrae un campione dalla popolazione

Campionamento di una $pdf(x)$

- gli eventi che formano il campione sono numeri relativa
- la popolazione è completamente rappresentata dalla $pdf(x)$
- l'operazione di campionamento della pdf consiste nell'estarre numeri che siano rappresentativi della pdf stessa

Teorema (del limite centrale)
Il teorema stidia la pdf di una r.v. costruita in modo analogo $pdf_1(\mu_1, \sigma_1), \ldots, pdf_N(\mu_N, \sigma_N)$

- consideriamo N variabili aleatorie indipendenti x_i, ciascuna caretterizzata da pdf_i
- assummiamo che per ogni pdf_i esistano finite media μ_i e varianza σ_i
- definiamo nuova random variable $\bar{x}$ costruita come media aritmetica delle r.v. di ciascuna pdf_i

$$\bar{x} = \sum_{i=1}^{N} \frac{x_i}{N}$$

Allora se $N \to \infty \bar{x}$ è distibuita in modo gaussiano con media pari alla somma delle medie e varianza pari alla somma delle varianze (media pari a μ e varianza pari a $\frac{\sigma^2}{N}$)

$$E[\bar{x}] = \mu_{\bar{x}} = \sum_{i=1}^{N} \frac{\mu_i}{N}$$
$$Var[\bar{x}] = \sigma_{\bar{x}}^2 = \sum_{i=1}^{N} \frac{\sigma_i^2}{N^2}$$
$$pdf\left(\sum_{i=1}^{N} \frac{x_i}{N}\right) \to Gauss(\mu_{\bar{x}}, \sigma_{\bar{x}}) \quad N \to \infty$$

Misura come variabile aleatoria

Misura come grandezza fisica $\to$ affetta da errore casuale

- x_0 : valore vero della grandezza che misuriamo
- x : risultato della singola misura
- $\varepsilon = \sum \varepsilon_i$: errore associato ad ogni misura $\to$ somma dei ε_i casuali

Allora
$$x = x_0 + \varepsilon$$
x è una variabile aleatoria associata a $pdf(x)$
Ripetere misura equivale a campionare la pdf(x), che dipende da $pdf(\varepsilon)$
Se $\varepsilon = \sum \varepsilon_i$ sono casuali $\Rightarrow pdf(\varepsilon)$ è una gaussiana centrata in 0; se non è centrata in 0 allora sono presenti degli errori sistematici

Cambio di variabili nelle pdf

x è una r.v. descritta da $pdf_1(x) \to y = y(x) \Rightarrow pdf_2(y)$ con y biunivoca, monotona e derivabile

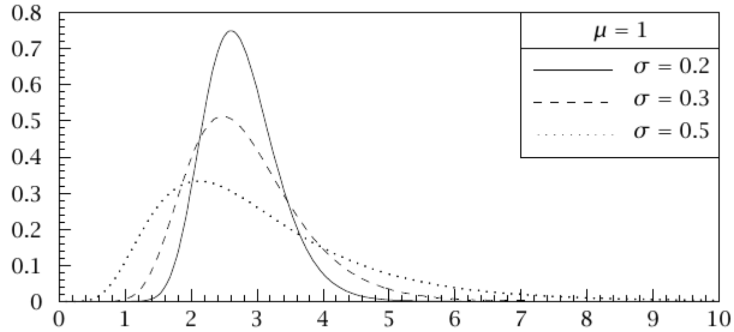
1. Calcolo $pdf_2(y)$ in forma analitica
2. determino μ_y σ_y della $pdf_2(y)$ con la propagazione degli errori

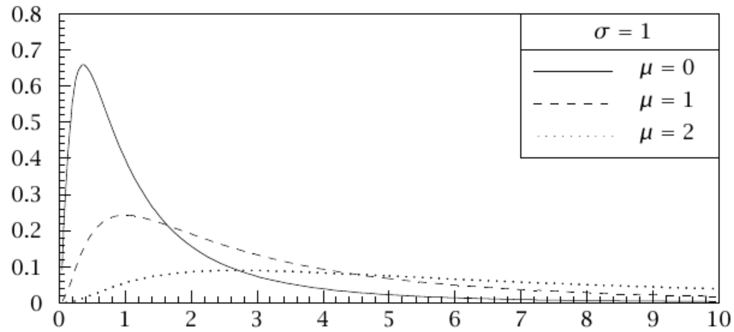
$$pdf_2(y) = pdf_1(x) \cdot |x'(y)| = odf_1(x) \cdot \frac{1}{|y'(x)|} = pdf_1(x(y)) \cdot |x'(y)|$$
$$pdf_2(x) = pdf_1(x) \cdot \left|\frac{dx}{dy}\right|$$

Distribuzione lognormale

x r.v. con distribuzione gaussiana
$y = e^x \Rightarrow x = \log y \; x' = \frac{1}{y}$
$$pdf(y; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\log y - \mu)^2}{2\sigma^2}}$$

$$E[x] = e^{\left(\mu + \frac{\sigma^2}{2}\right)}$$
$$E[x^2] = e^{(2\mu + 2\sigma^2)}$$
$$Var[x] = e^{(2\mu + 2\sigma^2)} \left(e^{\sigma^2} - 1\right)$$





2.4 Propagazione degli errori

- $y = ax + b$ Lineare $\mu_y = a\mu_x + b$ $\sigma_y^2 = a^2\sigma_x^2$ Formula esatta di propagazione degli errori
- $y = y(x)$ non lineare y(x) = y(μ_x) + $\left.\frac{dy}{dx}\right|_{\mu_x}(x - \mu_x) + \ldots$ Taylor

$$y(\mu_x) = \int (x - \mu_x) \cdot pdf(x)\, dx \quad \text{approssimazione primo ordine}$$

$$\mu_y \simeq y(\mu_y) + \frac{1}{2}\left.\frac{d^2y}{dx^2}\right|_{\mu_x} \sigma_x^2 \qquad \sigma_y^2 \simeq \left(\left.\frac{dy}{dx}\right|_{\mu_x}\right)^2 \sigma_x^2 \quad \text{approssimazione secondo ordine}$$

Questa si chiama formula approssimata di propagazione degli errori; per essere applicata gli errori non devono essere troppo grandi e le variabili stesse devono essere tra loro statisticamente indipendenti
Se l'evento casuale è identificabile con un numero intero k ⇒ r.v. discreta:

- il sample-space Ω è un sottoinsieme di $\mathbb{N}$
- $P(k)$ = probabilità che r.v. assuma valore k

- $P : (\Omega \subseteq \mathbb{N}) \mapsto [0,1]$

Allora gli assiomi di Kolmogorov diventano:

- $P(k) > 0 \, \forall k$
- $\sum_{\forall k \in \Omega} P(k) = 1$
- $P(k \,\&\, h) = P(k) + P(h) \quad \forall k, n \in \Omega \; h \neq k$

Bernoulli trial o distribuzione binomiale

Identifica la probabilità che un evento k si verifichi esattamente N volte, con ognuna delle prove statisticamente indipendente dalle altre Distribuzione Binomiale: discrive k successi in N prove

$$B(k, N, p) = \binom{N}{k} p^k (1-p)^{N-k}$$

media: $\mu = \sum_{k=0}^{N} k \cdot B(k, N, p) = Np$

varianza: $\sigma^2 = Var[k] = Np(1-p)$

Quando il numero di prove N è sufficientemente elevato e la probabilità p non è troppo vicina ai valori estremi di 0 e 1, allora di distribuzione binomiale è ben approssimata da una distribuzione binomiale; in generale si ritiene che l'approssimazione sia accettabile quando entrambi i prodotti Np e Nq hanno valore inferiore a 5
Vale proprietà riproduttività

$$\gamma_1 = \begin{cases} \frac{1-2p}{\sqrt{N_p(1-p)}} & p \neq 0 \\ 0 & p = \frac{1}{2} \end{cases} \qquad \gamma_1 \to 0 \; N \to \infty$$

$$\gamma_2 = \frac{1 - 6p(1-p)}{N} \qquad \gamma_2 \to 0 \; N \to \infty$$

Se $N \to \infty$ Binomiale asintotica ad una Gaussiana: $B(k; N, p) \simeq G(k; \mu, \sqrt{\mu})$
Se $p \to 0$ e $N \cdot p$ è finito $\Rightarrow$ Binomiale asintotica a Poissoni: $B(k; N, p) \simeq P(k; \lambda) \; \lambda = N \cdot p$

Distibuzione di Poisson

- La probabilità del verificarsi di un evento in un intervallo di tempo molto piccolo dt è proporzionale alla durata di tale intervallo
- Il verificarsi o meno di un evento in uncerto intervallo di tempo è indipendente dal verificarsi o meno di un evento prima o dopo di esso

Probabilità di contare k eventi in un intervallo Δx unitario; è il caso di un processo per cui la frequenza media è λ costante
L'accadere di un evento è indipendete dall'accadere di un evento successivo

$$Poiss(k; \lambda) = \frac{e^{-\lambda}}{k!} \lambda^k \qquad k \in [0, \infty) \quad \lambda = \frac{E[k]}{\Delta x = 1} \text{ media pdf}$$

- probabilità evento dx è proporzionale all'ampiezza dell'intervallo $p \cdot dx$
- p costante indipendente da x: FREQUENZA MEDIA COSTANTE
- la probabilità di avere più di un evento in dx è nulla: EVENTI PARI
- eventi indipendeti tra loro

* $p(1 \text{ evento in } dx) = p \cdot dx$
* $p(0 \text{ eventi in } dx) = 1 - p \cdot dx$
* $p(0 \text{ eventi in } [0, x]) = q(x)$
* $p(0 \text{ eventi in } [0, x + dx]) = q(0, x + dx)$

$$q(x + dx) = q(x)(1 - pdx)$$
$$\frac{q(x + dx) - q(x)}{dx} = -p \cdot q(x)$$
$$q(x) = q(0)e^{-px} \qquad q(0) \text{ da normalizzazione pdf: } \int_\Omega q(x) = 1$$

La probabilità di non avere eventi in $[0, x]$ è data da esponenziale; la probabilità di avere 0 eventi in $[0, x_1]$ e 1 eventi in $[x_1, x_1 + dx]$ è uguale a $e^{-px_1}(pdx_1)$
Probabilità di avere k eventi in $[0, x]$: $e^{-px}\frac{(px)^k}{k!}$

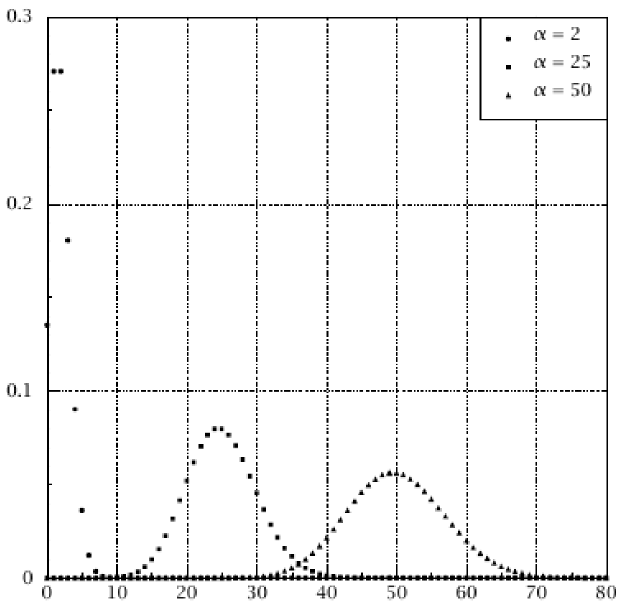
Media: $\mu = E[k] = \sum_{k=0}^{\infty} k \cdot Poiss(k; \lambda) = \lambda$

Varianza $\sigma^2 = \lambda = x$

$\gamma_1 = \frac{1}{\sqrt{x}} \qquad \gamma_2 = \frac{1}{\lambda}$

Se $\lambda \to \infty \Rightarrow$ Poiss$(k; \lambda) \to$ Gauss$(k; \lambda, \sqrt{\lambda})$
La distribuzione di Poisson è una approssimazione alla distribuzione binomiale che si può ritenere valida qualora si verifichino eventi casuali di probabilità estremamente piccola, e che ci è possibile vedere solo quando si compiono osservazioni su un numero elevato di esistano

$$p^2 << p \qquad p << Np << N$$



Distribuzione esponenziale

$$pdf(t) = \frac{1}{\tau} e^{-\frac{t}{\tau}} \qquad p = \frac{1}{\tau}$$

Media: τ

Varianza: τ^2

2.5 PDF IN N-DIMENSIONI

Un evento può essere identificato da un vettore $\rightarrow$ valgono gli assiomi di Kolmogorov

$$pdf(\vec{x}) : \mathbb{R}^n \rightarrow \mathbb{R}^+$$
$$cdf(\vec{x}) : \int_{-\infty}^{\vec{x}} pdf(\vec{x})\, d\vec{x}$$

Media: $\vec{\mu} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_N \end{pmatrix} = E[\vec{x}] = \int \vec{x} pdf(\vec{x}) d\vec{x} \quad E[x_i] = \int x_i pdf(\vec{x}) d\vec{x}$

Varianza: matrice di covarianza $n \times n \quad \sigma^2_{ij} = [(x_i - \mu_i)(x_j - \mu_j)]$

pdf congiunta tra r.v $\{x_1, \ldots, x_N\}$ $pdf(\vec{x}) \leftrightarrow pdf(x_1, \ldots, x_N)$
$pdf_n(x)$ marginale $\rightarrow$ distribuzione di x indipendente da y

$$pdf_n(x) = \int pdf(x, y)\, dy$$

Probabilità condizionata: $pdf(x \,|\, y = y_0) \rightarrow pdf$ associata ad x quando $y = y_0$

$$pdf(x \,|\, y = y_0) = \frac{pdf(x, y_0)}{pdf_{ny}(y_0)} \quad pdf_{ny}(y_0) \text{ normalizzazione}$$

x e y sono indipendenti $\iff pdf(x,y) = pdf_{nx}(x) \cdot pdf_{n_y}(y)$

$$E[\mu(x,y)] = \iint \mu(x,y) pdf(x,y)\, dx\, dy$$
$$\mu_x = E[x] = \int x\, dx \int pdf(x,y)\, dy$$
$$Var[x] = E[(x-\mu_x)^2] = \sigma_x^2$$
$$Cov[x,y] = E[(x-\mu_x)(y-\mu_y)] = E[(x,y)]\mu_x\mu_y$$

- x e y sono indipendenti $\Rightarrow Cov[x,y] = 0$, ma $Cov[x,y] = 0$ non implica x e y indipendenti
- x e y legati linearmente $\Rightarrow Cov[x,y] = 0$

2.5.1 Matrice covarianza

La matrice V di covarianza è simmetrica
Coefficiente di correlazione
$$\rho_{xy} = \frac{Cov[x,y]}{\sigma_x \sigma_y} \qquad \rho^2 \leq 1$$
$\rho^2 = 0$ se x e y sono legati linearmente
Il coefficiente di correlazione è ovviamente adimensionale ed è nullo quando le variabili stesse sono statisticamente indipendenti; il suo valore è compreso tra -1 e 1

Teorema
Due differenti combinazioni lineari delle stesse variabili sono sempre correlate

Cambio di variabili per rendere covarianza nulla $\Rightarrow$ variabili non correlate, ma non indipendenti

$$\vec{x} \mapsto \vec{y} = f(\vec{x})$$
$$pdf_x(\vec{x}) \mapsto pdf_y(\vec{y})$$

Caso bidimensionale $\rightarrow pdf_x(x_1, x_2)$: pdf congiunta

$$\begin{cases} x_1 \rightarrow y_1 = v_1(x_1, x_2) \\ x_2 \rightarrow y_1 = v_2(x_1, x_2) \end{cases}$$

$$pdf(y_1, y_2) dy_1 dy_2 = pfd(x_1, x_2) dx_1 dx_2$$

Funzioni inverse
$$\begin{cases} x_1 = w_1(y_1, y_2) \\ x_2 = w_2(y_1, y_2) \end{cases}$$

In generale
$$pdf_y(y_1, y_2) = pdf_x(w_1(y_1, y_2), w_2(y_1, y_2)) \det J \qquad \text{J: Jacobiano}$$

Le trasformazioni di variabili si usano per:

- ridefinire le variabili in modo che Cov[x,y]=0
- Generare numeri casuali: Box-Muller

2.5.2 Propagazione degli errori per variabili correlate

$$F = F(x_1, \ldots, x_N)$$

Introduco vettore F di dimensione N con componenti

$$F_i = \frac{\delta F}{\delta x_i}$$

e il suo trasposto F^T
Allora

$$Var(F) = F^T V F$$

2.5.3 Pdf normale multivariata

Estensione n-dimensionale della Gaussiana

$$E[\vec{x}] = \int \vec{x} pdf(\vec{x})\, d\vec{x}$$

Matrice di covarianza

$$\Sigma = E[(x_i - \mu_i)(x_j - \mu_j)] = \sigma_{ij}^2 = Cov[x_i, x_j]$$

$$N(\vec{x}; \vec{\mu}, \Sigma) = \frac{1}{\sqrt{(2\pi)^n \det(\Sigma)}} e^{\left(-\frac{1}{2}(\vec{x}-\vec{\mu})^T \Sigma^{-1}(\vec{x}-\vec{\mu})\right)}$$

In 2 dimensioni:

$$\Sigma = \begin{pmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix}$$

$$N(\vec{x}; \vec{\mu}, \Sigma) = \frac{1}{\sqrt{(2\pi)^2 \sigma_x^2 \sigma_y^2 (1-\rho)}} e^{-\left(\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right) + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]\right)}$$

$-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right) + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right] = k = cost \rightarrow$ Curve di equiprobabilità

2.5.4 Binomiale standardizzata

$\vec{\mu} = 0$ $\sigma_x = \sigma_y = 1$
Media e moda coincidono
Massimo per k=0 $\rightarrow$ origine

$$N(\vec{x}) = \frac{e^{-\left(-\frac{1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)\right)}}{2\pi\sqrt{1-\rho}}$$

Curve di equiprobabilità: ellissi centrati nell'origine
Forma a campionaria

- $\rho = 0 \rightarrow$ cerchi
- $\rho \rightarrow 1 \Rightarrow$ ellisse degenera nella retta

$k = -\frac{1}{2} \rightarrow$ riduce densità di $\frac{1}{\sqrt{e}}$ rispetto al massimo
Ellisse degli errori $(\mu_x \pm \sigma_x; \mu_y \pm \sigma_y) \rightarrow$ inegrale pdf su questo ellisse $\Rightarrow 39\%$
La matrice di covarianza è diagonalizzabile $\Rightarrow$ Esiste cambio di variabili con variabili non correlate

2.6 Stime parametri

Def (Statistica)
Una statistica è una funzione che dato un campione restituisce le proprietà e i parametri; variabile aleatoria: funzione degli N campionamenti

Def (Stimatore)
Uno stimatore è una statistica che usa N campionamenti per fornire una stima dei parametri; uno stimatore è dunque una funzione di variabili casuali e di conseguenza una variabile casuale

$$\bar{\theta} = \bar{\theta} = (x_1, x_2, \dots, x_N)$$

Def (Stimatore consistente)
Uno stimatore si dice consistente se converge probabilisticamente al valore vero del parametro

$$\lim_{N \to \infty} \bar{\theta}(x_1, x_2, \dots, x_n) = \theta^*$$

Def (Stimatore unbiased)
Uno stimatore si accurato (unbiased) se mediamente coincide con il valore vero del parametro

$$E[\bar{\theta}] = \int \bar{\theta} pdf_{\bar{\theta}}(\bar{\theta})\, d\bar{\theta} = \theta^*$$

Il bias non è sempre valutabile a priori

$$E[\bar{\theta}_N] - \theta^*$$

Def (Stimatore efficiente)
Uno stimatore si dice tanto più efficiente tanto è minore la sua varianza; l'efficienza determina la precisione dello stimatore
Teorema di Rao-Cramer

2.7 Funzione di verosimiglianza

Dato un campione di N eventi indipendenti x_i allora si può costruire una funzione di verosimiglianza che rappresenta la densità di probabilità da associare all'evento casuale consistente nell'essere un certo θ il valore vero del parametro

$$\mathcal{L}(x_1, x_2, \dots, x_N; \theta) = \prod_{i=1}^{N} f(x_i; \theta)$$

in cui la variabile indipendente è θ e il gli x_i sono costanti Il metodo di massima verosimiglianza consiste nello stimare il parametro θ come il valore che rende massima la funzione di verosimiglianza

$$\frac{d\mathcal{L}}{d\theta} = 0 \qquad \frac{d^2\mathcal{L}}{d\theta^2} < 0$$

Nel caso abbiano più soluzioni si sceglie il massimo assoluto
Dato che il logaritmo naturale è una funzione monotona strettamente crescente allora il massimo della $\ln \mathcal{L}$ corrisponde al massimo della $\mathcal{L}$

1. Lo stimatore di massima verosimiglianza è una stima asintoticamente consistente al crescere della dimensione del campione

2. Lo stimatore di massima verosimiglianza ha una densità di probabilità asintoticamente normale al crescere della dimensione del campione
3. Lo stimatore di massima verosimiglianza è asintoticamente anche lo stimatore più efficiente
4. Asintoticamente gli stimatori della maximum likelihood sono efficienti e privi di bias

La funzione di verosimiglianza permette:

- di misurare l'informazione dei campionamenti sul parametro; consente di studiare tecniche per ridurre i dati senza perdere informazioni
- di valutare la minima varianza raggiungibile con uno stimatore
- di realizzare un metodo per costruire uno stimatore

Teorema (Rao-Cramer)
Una qualsiasi stima imaparziale $\bar{\theta}$ del parametro ha una varianza che non può essere inferiore ad un valore limite

$$Var(\bar{\theta}) \geq \frac{\left[1 + \frac{\delta b_n}{\delta \theta}\right]}{E\left[-\frac{\delta^2 \ln(\mathcal{L}(\vec{x};\theta))}{\delta \theta^2}\right]}$$

Uno stimatore si dice efficiente se vale l'uguale nella relazione; in questo caso lo stimatore di minima varianza rende massima la funzione di verosimiglianza