CHAPTER 1

La probabilità e la distribuzione di probabilità

1.1 Assiomi di Kolgomorov

Come si definisce la probabilitá secondo Kolgomorov?

Consideriamo degli eventi $E_1, E_2 \subset \Omega$ spazio campione, vogliamo costruire definiamo probabilitá una un funzione $P : \Omega \to [0,1]$ che soddisfa le seguenti proprietá:

- $P(\Omega) = 1$
- $\forall A \subset \Omega$ si ha che $P(A) \geq 0$
- $\forall A, B \subset \Omega$, si ha che $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. Se gli eventi sono indipendenti $A \cap B = \varnothing$, si ha che $P(A \cup B) = P(A) + P(B)$.

1.1.1 Definizione di probabilitá frequentista

Come si definisce operativamente la probabilitá secondo la formulazione frequentista?

La probabilitá di un evento A é definita come il rapporto tra il numero di casi favorevoli (in cui avviene l'evento A) e il numero di casi possibili (la popolazione). Consideriamo di avere **N** eventi, e di contare il numero di volte in cui l'evento A, indicandolo come $\mathbf{n(A)}$. Definiamo la probabilitá di un evento come:

$$\lim_{N\to\infty}\frac{n(A)}{N} \tag{1.1}$$

dove il limite é inteso in senso probabilistico ovvero al crescere del numero degli eventi.

Probabilitá condizionata

Che cos'é la probabilitá condizionata ?

Consideriamo una coppia di eventi $A, B \subset \Omega$ definiamo **probabilitá condizionata** la probabilitá che si verifichi l'evento A al verificarsi previamente dell'evento B.

$$P(A|B) = \frac{P(A\cap B)}{P(B)} \tag{1.2}$$

si dimostra che la probabilitá condizionata soddisfa gli assiomi di Kolgomorov.
Nel caso di due eventi indipendenti, ovvero che il verificarsi di B non influenza la probabilitá che si verifichi A

$$P(A|B) = P(A|\Omega) = P(A)$$

per due eventi indipendenti possiamo riscrivere la probabilitá condizionata come:

$$P(A\cap B) = P(A)\cdot P(B) \tag{1.3}$$

1.1.2 Teorema di Bayes

Che cosa afferma il teorema di Bayes ?

Consideriamo due eventi $A, B \subset \Omega$ la probabilitá nel caso di due eventi non indipendenti possiamo riscrivere la probabilitá condizionata come:

$$P(A|B)\cdot P(B) = P(A\cap B) = P(B|A)\cdot P(A)$$

dunque l'equazione (1.2) puó essere riscritta come :

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)} \tag{1.4}$$

tale espressione definisce il **teorema di Bayes**. La probabilitá $P(A|B)$ viene detta " a posteriori " poiché permette di calcolare la probabilitá di ,sapendo che si é verificato (o si verificherá con certezza assoluta) B.
La probabilitá P(A) si dice invece "a priori" poiché non condizionata da alcun altro evento o da alcuna conoscenza che potremmo avere sul suo verificarsi.

1.2 Variabili aleatorie

Che cos'é una variabile aleatoria ?

Una **variabile aleatoria** é una funzione definita sullo spazio campione a cui viene associato un sottoinsieme misurabile.

$$X : \Omega \to E \tag{1.5}$$

dove Ω é lo spazio di tutti gli esiti possibili, mentre E é l'insieme degli esiti al verificarsi di un determinato evento.

Nel caso in cui $\Omega \subseteq \mathbb{N}$ si ha che x viene definita **variabile casuale discreta**.
Se $\Omega \subseteq \mathbb{R}$, x viene definita **variabile aleatoria continua**, in questo caso la probabilitá viene misurata su intervalli $P(a < x < b)$ e non piú sui singoli eventi come nel caso di quella discreta.

Esempio:

- Il conteggio del numero di volte in cui compare un certo valore di una misura é una variabile aleatoria discreta.

- La temperatura é una variabile aleatoria continua.

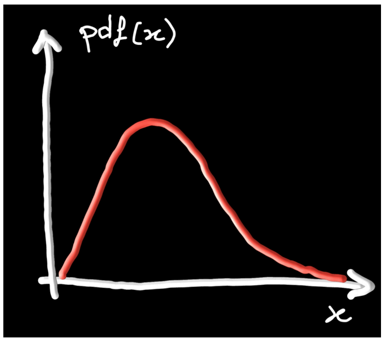
1.3 Probability Distribution Function (P.d.f)

Che cosa é una pdf e quali sono le sue proprietá ?

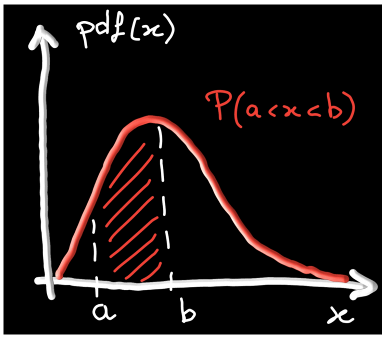
Si definisce densitá di probabilitá (p.d.f) una funzione che soddisfa le seguenti proprietá:

1. $pdf(x) : \Omega \subseteq \mathbb{R} \to \mathbb{R}^+$
2. $P(x < \hat{x} < x + dx) = pdf(x) \cdot dx$
3. $P(a < x < b) = \int_a^b pdf(x) \cdot dx \leq 1$
4. P(x) = 0, la probabilitá in una singola misura é nulla.

per la definizione di porbabilitá nei punti 2) e 3) valgono gli assiomi di Kolgomorov.



(a) pdf(x)



(b) $P(a < x < b)$

1.3.1 Cumulative distribution function (C.d.f)

Che cosé una c.d.f.?

Si definisce distribuzione cumulativa la primitiva dell'integrale della pdf di una variabile aleatoria continua **x**:

$$cdf(x) = \int_a^x pdf(x)dx \tag{1.6}$$

ovvero si ha che:

$$\frac{d}{dx}cdf(x) = pdf(x) \tag{1.7}$$

di conseguenza la cdf(x) descrive l'evolversi della probabilitá di una variabile casuale.

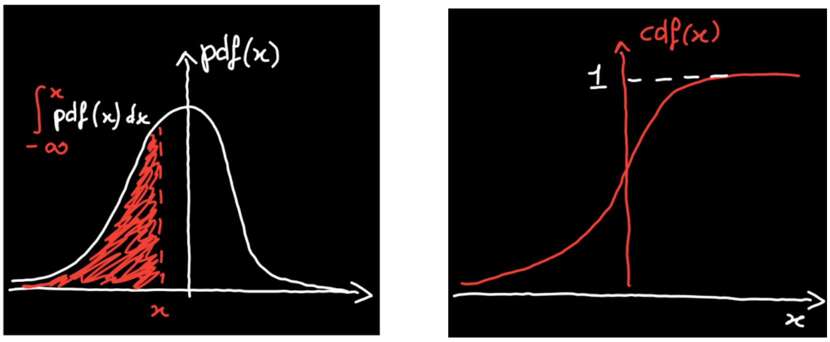
Proprietà di unc cdf

Una cdf gode delle seguenti proprietà:

- È una funzione monotona crescente e definita poisitiva
- cdf(x) $\in [0, 1]$
- $P(a \leq x \leq b) = cdf(a) - cdf(b)$

Corollario

Esiste una forma analitica della cdf(x) di una r.v. $\iff$ la pdf ammette primitiva.

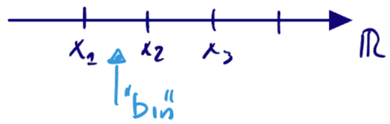


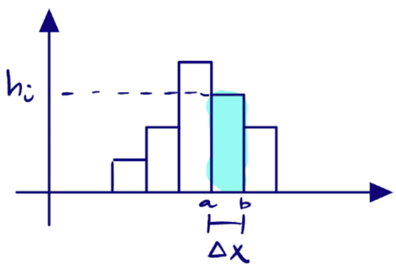
(a) $\int pdf(x)$ (b) cdf(x)

1.3.2 Istogrammi

Che cos'é un istogramma ?

Consideriamo una variabile aleatoria x, di cui conosciamo $\{x_i\}_i^N$ valori distinti, utilizziamo tali grandezze come estremi per definire degli intervalli $\omega_k = [x_i, x_{i+1})$ disgiunti tra loro $\omega_k \cap \omega_m = \varnothing$ per $k \neq m$, sull'asse reale. Tali intervalli vengono definiti **bin**.



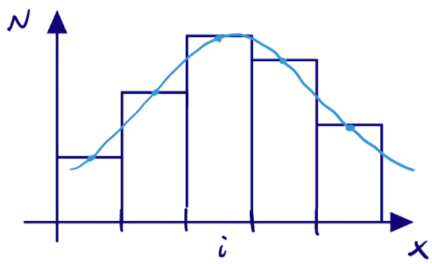


Ripetendo le misure contiamo il numero di volte n_l che un valore compare e cade all'interno di un'intervallo ω_k.

In questo modo posizionando sull'asse delle ordinate il numero di conteggi $n_l(x)$ si costruisce un istogramma.

Come si puó utilizzare un istogramma per rappresentare una pdf ?

All'aumentare del numero di elementi $N \to \infty$ e diminuendo l'ampiezza dei bin $\Delta x \to 0$, ci si riconduce alla nozione di probabilitá (1.1), in questo modo si passa a una pdf(x) continua. Per ogni bin è presente un punto x_i tale che



$$\int_{x^i_{min}}^{x^i_{max}} f(x)dx = f(x_i)\Delta x_i$$

si ha che

$$\lim_{\substack{N\to\infty \\ \Delta\to 0}} \int_{x^i_{min}}^{x^i_{max}} f(x)dx \to p_i \Rightarrow f(x_i) = \frac{p_i}{\Delta x_i}$$

1.3.3 Proprietà di una probability distribution function

Che cosa sono i momenti di una pdf?

Conoscere l'espressione analitica di una pdf è spesso poco significativo (soprattuto se la sua espressione non lascia intuire la forma della curva) in altri casi è impossibile definirla. Di conseguenza per studiare il comportamento di una pdf facciamo affidamento ad alcune grandezze che ne descrivono il comportamento, queste prendono il nome di **momenti**.

Tali momenti possiamo classificarli come:

- $E[x^m] = \int x^m \cdot pdf(x)dx$ prendono il nome di **momenti centrali di ordine m**.
- $E[(x-\mu)^m] = \int (x-\mu)^m \cdot pdf(x)dx$ prendono il nome di **momenti centrali**

il momento centrale di ordine 1 vale $E[(x-\mu)] = 0$.

1.3.4 Valore di aspettazione di una probability distribution function

Che cosa è il valore di aspettazione di una quantità rispetto ad una pdf ?

Data $x \in \Omega$ variabile aleatoria e $u(x) : \mathbb{R} \to \mathbb{R}$, il valore di aspettazione di u(x) è definito come:

$$E[u(x)] = \int u(x) \cdot pdf(x)dx \tag{1.8}$$

dove E[ ] è un operatore lineare:

- $E[x+y] = E[x] + E[y]$
- $E[\alpha x] = \alpha E[x] \;\; \forall \alpha \in \mathbb{R}$

Per una variabile aleatoria discreta il valore di aspettazione è dato da:

$$E[x] = \sum_{i=1}^{+\infty} p_i x_i \tag{1.9}$$

1.3.5 Momenti principali di una pdf per la popolazione e per un campione

Che cosa sono e che cosa rappresentano la media, la varianza, l'assimetria e la curtosi di un campione ? e di una popolazione?

Sia $x \in \Omega$ una variabile aleatoria continua definiamo valore di aspettazione **media della popolazione** la grandezza:

$$\mu \equiv E[x] = \int x \cdot pdf(x)dx \tag{1.10}$$

Essa rappresenta il centro della distribuzione di probabilità .
Per un campione di N misure **la media** viene definita come:

$$\overline{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{1.11}$$

Si definisce **varianza della popolazione**:

$$\sigma^2 \equiv V[x] = E[(x-\mu)^2] = \int (x-\mu)^2 \cdot pdf(x)dx \tag{1.12}$$

la sua radice prende il nome di **deviazione standard**e definisce l'ampiezza della pdf. La varianza può essere riscritta come:

$$V[x] = E[x^2] + E[x]^2 = E[x^2] + \mu^2 \tag{1.13}$$

Per un campione di N misure **la varianza** è data da:

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x-\mu)^2 \tag{1.14}$$

mentre **la deviazione standard** é di un campione é data da:

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x-\mu)^2} \tag{1.15}$$

Tra i momenti centrali sono anche presenti la **skewness** per un campione di N misure:

$$\gamma_1 = \frac{1}{N}\sum_{i=1}^{N}\frac{(x-\mu)^3}{\sigma^3} \tag{1.16}$$

e per la popolazione:

$$\gamma_1 = \frac{E[(x-\mu)^3]}{\sigma^3} \tag{1.17}$$

tale grandezza definisce la simmetria di una distribuzione di probabilitá rispetto al valore centrale dato dal valore di aspettazione μ per una popolazione. Se una distribuzione é simmetrica $\gamma_1 = 0$.

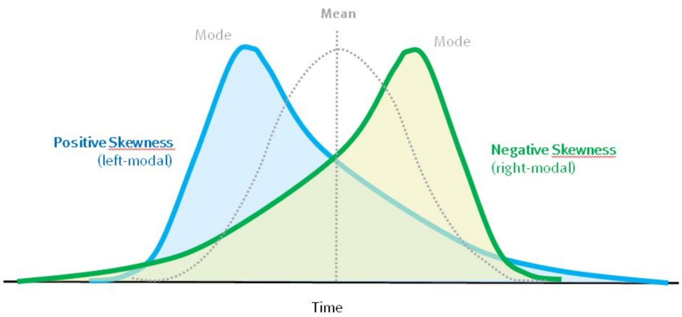


Figure 1.1: Simmetria di una distribuzione di probabilità

Si definisce indice di **kurtosi** di un campione di N elementi:

$$\gamma_2 = \frac{1}{N}\sum_{i=1}^{N}\frac{(x-\mu)^4}{\sigma^4} - 3 \tag{1.18}$$

mentre per la popolazione è data da:

$$\gamma_2 = \frac{E[(x-\mu)^4]}{\sigma^4} - 3 \tag{1.19}$$

tale stima fornisce informazioni sul peso delle code della distribuzione, nel senso che per code significative la distribuzione risulterà più schiacciata nell'intorno del valore μ, mentre per code meno incidenti sarà piú piccata.

Stime di tendenza centrale

Che cosa sono moda, media e mediana di un campione? e di una popolazione ?

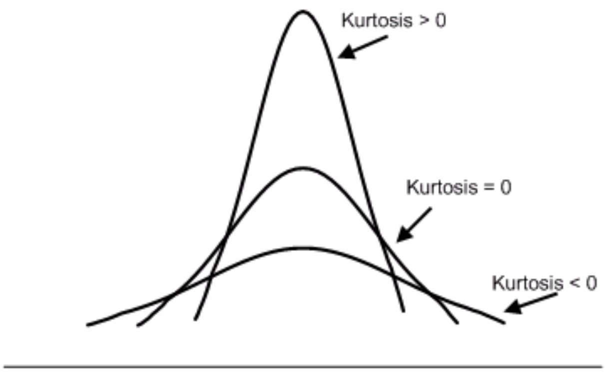
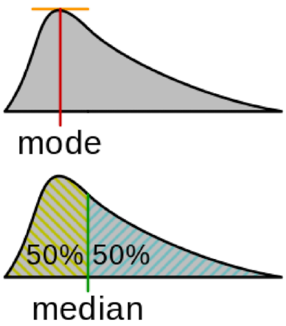


Figure 1.2: Kurtosi di una distribuzione di probabilità



Definiamo **moda** di una popolazione il valore che definisce il massimo della pdf. Mentre per un campione la moda è definita dalla misura che possiede la frequenza relativa maggiore.

Mentre si definisce **mediana** il valore che divide in due parti uguali la pdf. Nel caso dei campioni le grandezza diventano rispettivamente il valore che compare con maggior frequenza e la misura che divide in due parti eguali il campione.

Per media e mediana di una distribuzione sono unici mentre per una distribuzione può avere più mode, in questi casi si parla di distribuzione **multimodale**.

1.4 Trasformazioni di distribuzioni di probabilità

Sia x una variabile aleatori continua che segue una pdf(x) e ipotizziamo di voler costruire una nuova variabile aleatori y legandola ad x mediante una funzione f(x), come definiamo la pdf(y) ?

Sappiamo che la probabilità che una misura cada in un intervallo [a,b) è data da

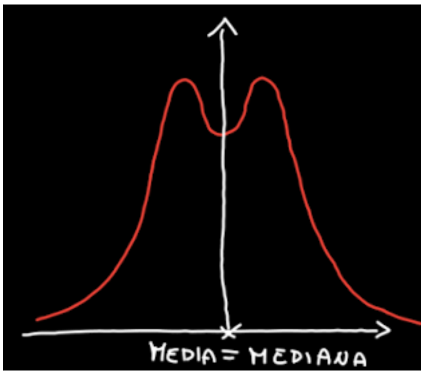


Figure 1.3: Distribuzione multimodale

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

Se effettuiamo un cambio di variabii y=f(x) vogliamo inanzitutto che la trasformazione sia biunivoca dunque f(x) deve essere monotona e continua. In questo modo applicando il teorema per il cambio di coordinate sotto segno d'integrale si ha che:

$$\frac{dy}{dx} = f'(x) \Rightarrow dx = f'(x)^{-1}dy \tag{1.20}$$

di conseguenza:

$$P(a < x < b) = \int_a^b pdf(x)dx = \int_{f(a)}^{f(b)} pdf(y)|f'(x)^{-1}|dy$$

per comodità si è posto un modulo affinchè la funzione sia sempre positiva. In conclusione si ha che:

$$pdf(y) = pdf(x)|f'(x)| \tag{1.21}$$

Nel caso in cui abbia più di una dimensione è necessario che la funzione sia differenziabile e monotona e il modulo della derivata prima della funzione di cambio delle variabili viene sostituita dal modulo del determinante dell'inversa della matrice Jacobiana.

$$pdf(\mathbf{y}) = pdf(\mathbf{x}) \cdot |detJ| \tag{1.22}$$

Come si osserva dalle equazioni (1.21) e (1.22) la nuova pdf(y) aumenta o diminuisce di un fattore rispetto alla pdf(x). Comunque sia la probabilità in un intervallo resta invariata.

Resta da discutere come cambiano i momenti della distribuzione, per farlo suddividiamo i risultati in due casi:

Caso in cui il cambio di variabile è lineare y = ax+b

Il valore di aspettazione per una pdf(y) dove y = ax+b diventa:

$$E[y] = \int y \cdot pdf(y)dy = \int (ax + b) \cdot pdf(x)dx = y(\mu_x) \tag{1.23}$$

analogamente la varianza sarà data da:

$$V[y] = E[y^2] - E[y]^2 = a^2 \cdot V[x] \tag{1.24}$$

Caso in cui il cambio di variabile é non lineare y = f(x)

il valore di aspettazione e la varianza di una pdf(y) per y = f(x) non lineare si stima approssimando con il polinomio di Taylor la funzione y(x) in un'intorno di μ_x al secondo ordine:

$$y(x) \approx y(\mu_x) + \frac{dy}{dx}\Big|_{x=\mu_x}(x - \mu_x) + \frac{1}{2}\frac{dy^2}{dx^2}\Big|_{x=\mu_x}(x - \mu_x)^2$$

di conseguenza il valore di aspettazione è:

$$E[y] = \int y \cdot pdf(y)dy = y(\mu_x) + \frac{dy}{dx}\Big|_{x=\mu_x}\int (x - \mu_x) \cdot pdf(x)dx+$$

$$+\frac{1}{2}\frac{dy^2}{dx^2}\Big|_{x=\mu_x}\int(x-\mu_x)^2\cdot pdf(x)dx =$$

$$= \boxed{y(\mu_x)+\frac{1}{2}\frac{dy^2}{dx^2}\Big|_{x=\mu_x}\cdot\sigma_x^2} \tag{1.25}$$

Mentre la varianza viene si ottiene dalla relazione:

$$V[y] = E[y^2]-E[y]^2 = \boxed{\left(\frac{dy}{dx}\Big|_{x=\mu_x}\right)^2\cdot\sigma_x^2} \tag{1.26}$$

L'espressione (1.26) è alla base della formula di propagazione degli errori.

1.5 Assenza di memoria per una distribuzione di probabilità

Gli eventi Poissoniani sono indipendenti l'uno dall'altro e la loro probabilità di decadimento è costante, dunque definita q(t) la probabilità che non si verifichi un evento in un intervallo di tempo [0,t] si ha che:

$$q(t+\delta t) = q(t)q(\Delta t) \tag{1.27}$$

la (1.27) definisce l'assenza di memoria di una distribuzione. Si può dimostrare che una distribuzione ha assenza di memoria $\iff$ è esponenziale.