UNIVERSITÀ DI PISA

Dipartimento di Informatica
Corso di Laurea Triennale in Informatica

Corso 2° anno - 6 CFU

Statistica

Professore:
Prof. Francesco Grotto

Autore:
Filippo Ghirardini

Anno Accademico 2023/2024

Contents

Statistica

Realizzato da: Filippo Ghirardini

A.A. 2023-2024

1 Statistica descrittiva

La statistica si occupa dello studio dei dati, ovvero della sua **raccolta**, **analisi** ed **interpretazione**. Le risposte dipendono dai dati e dalla conoscenza pregressa del problema, quindi da eventuali ipotesi ed assunzioni.

- Statistica **descrittiva**: quando i dati vengono analizzati senza fare assunzioni esterne per evidenziarne la struttura e rappresentarli in modo efficace
- **Inferenza statistica**: studia i dati utilizzando un modello probabilistico, ovvero supponendo che i dati siano valori assunti da *variabili aleatorie* con una certa *distribuzione di probabilità* dipendente da parametri non noti che devono essere stimati. Il modello potrà poi fare previsioni.

1.0.1 Campioni statistici

Definizione 1.0.1 (Popolazione)**.** *Insieme di oggetti o fenomeni che si vuole studiare su ognuno dei quali si può effettuare una stessa misura, ovvero un* ***carattere****. Può essere* ***ideale*** *o* ***reale****.*

Definizione 1.0.2 (Campione statistico)**.** *Un sottoinsieme della popolazione scelto per rappresentarla.*

Definizione 1.0.3 (Dati)**.** *Misure effettuate sul campione statistico.*

Definizione 1.0.4 (Frequenza)**.** *Può essere:*

- ***Assoluta****: il numero di volte in cui questo esito compare nei dati*
- ***Relativa****: frazione di volte in cui questo esito compare sul totale dei dati*

In generale dipendono dai dati e quindi non coincidono su tutta la popolazione.

Note 1.0.1. La scelta del campione in modo che sia rappresentativo è importante ma non verrà trattata.

1.0.2 Istogramma

Consiste in una serie di colonne ognuna delle quali ha per base un intervallo numerico e per area la frequenza relativa dei dati contenuti nell'intervallo.

Osservazione 1.0.1. La scelta delle ampiezze degli intervalli di base è cruciale. Un buon compromesso deve essere individuato sulla base della numerosità dei dati e sulla loro distribuzione.

Può avere varie forme:

- **Normale** se ha la forma di una *campana simmetrica*
- **Unimodale** se si concentra su una colonna più alta o **bimodale** se su due. Può essere asimmetrica a *destra* o a *sinistra* in base alla concentrazione dei dati in base al picco
- **Platicurtica** se i dati sono concentrati in un certo intervallo o **leptocurtica** se sono composti da un gruppo centrale e da molti *outliers*

1.0.3 Indici statistici

Dato un vettore $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ di dati numerici gli indici statistici sono quantità che riassumono alcune proprietà significative.

Definizione 1.0.5 (Media campionaria)**.** *La media aritmetica dei dati:*

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

Definizione 1.0.6 (Mediana)**.** *Il dato x_i tale che la metà degli altri valori è minore o uguale ad esso e l'altra metà maggiore o uguale.*

Osservazione 1.0.2. La **mediana** è utile nel caso di dati molto **asimmetrici** ed è robusta rispetto alle code delle distribuzione. Al contrario la **media campionaria** viene facilmente spostata da dati molto piccoli o grandi.

Definizione 1.0.7 (Varianza campionaria)**.** *Si usa per misurare la dispersione dei dati attorno alla media campionaria.*

$$var(x) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \tag{2}$$

È nulla se i dati sono tutti uguali. Possiamo mappare x diversamente:

- *$x \mapsto x^2$ misura la media dei punti della media campionaria*
- *$x \mapsto x^3$ misura la **sample skewness**, ovvero l'asimmetria della distribuzione*

$$b = \frac{1}{\sigma} \cdot \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3 \tag{3}$$

- *$x \mapsto x^4$ misura la piattezza della distribuzione dei dati, ovvero la **curtosi***

Definizione 1.0.8 (Scarto quadratico medio o deviazione standard)**.**

$$\sigma(x) = \sqrt{var(x)} \tag{4}$$

Proposizione 1.0.1. *Dato un campione di dati x ed un numero positivo d:*

$$\frac{\#\{x_i : |x_i - \bar{x}| > d\}}{n-1} \leq \frac{var(x)}{d^2} \tag{5}$$

Il termine a sinistra è la frazione di dati che differiscono dalla media campionaria più di d.

1.0.4 Quantili

Definizione 1.0.9 (Funzione di ripartizione empirica)**.** *Dato $x = (x_1, \dots, x_n) \in \mathbb{R}^n$:*

$$F_e(t) = \frac{\#\{i | x_i \leq t\}}{n} \tag{6}$$

*Per ogni $t \in \mathbb{R}$ restituisce la frequenza relativa dei dati minori o uguali a t. È sempre **non decrescente** e $F_e(-\infty) = 0$, $F(+\infty) = 1$.*

Definizione 1.0.10 (β-quantile)**.** *Il dato x_i tale che:*

- *almeno βn dati siano $\leq x_i$*
- *almeno $(1-\beta)n$ dati siano $\geq x_i$*

Inoltre:

- *Se βn non è intero vale $x_{(\lceil \beta n \rceil)}$*
- *Se βn è intero è la media aritmetica tra $x_{(\beta n)}$ e $x_{(\beta n+1)}$*

1.0.5 Dati multi-variati

Consideriamo coppie di dati **bivariati** del tipo

$$(x, y) = ((x_1, y_1), \dots, (x_n, y_n))$$

Definizione 1.0.11 (Covarianza campionaria)**.**

$$cov(x, y) = \sum_{i=1}^{n}\frac{(x_i - \bar{x})(y_i - \bar{y})}{n-1} \tag{7}$$

Definizione 1.0.12 (Coefficiente di correlazione)**.** *Dati $\sigma(x) \neq 0$ e $\sigma(y) \neq 0$:*

$$r(x,y) = \frac{cov(x,y)}{\sigma(x)\sigma(y)} \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{t})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{8}$$

Misura la presenza di una relazione lineare tra i dati x e y quantificata dalla ***retta di regressione****.*

Proposizione 1.0.2 (Disuguaglianza di Cauchy-Scwarz)**.**

$$\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) \leq \sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{9}$$

e quindi

$$|r(x,y)| \leq 1 \tag{10}$$

La **retta di regressione** è un'approssimazione dei dati con y_i con una combinazione lineare affine a $a + bx_i$, ottenuta cercando il minimo della distanza dai dati da questa retta con i quadrati degli scarti. L'obiettivo è quindi di cercare i parametri a e b calcolando

$$\inf_{a,b\in\mathbb{R}} \sum_{i=1}^{n}(y_i - a - bx_i)^2 \tag{11}$$

Teorema 1.0.1 (Retta di regressione)**.** Se $\sigma(x) \neq 0$ e $\sigma(y) \neq 0$, esiste un unico minimo al variare di $a, b \in \mathbb{R}$ della quantità 11, dato da:

$$b^\star = \frac{(n-1)cov(x,y)}{n \cdot var(x)} \qquad a^\star = -b^\star\bar{x} + \bar{y} \tag{12}$$

e vale

$$\min_{a,b\in\mathbb{R}} \sum_{i=1}^{n}(y_i - a - bx_i)^2 = (1 - r(x,y)^2)\sum_{i=1}^{n}(y_i - \bar{y})^2 \tag{13}$$

Quanto più $r(x,y)$ è vicino a 1, tanto più i valori tendono ad allinearsi con la retta. Se vale 1 vuol dire che i punti sono tutti sulla retta. Il segno di $r(x,y)$ corrisponde al segno del coefficiente angolare. Se è prossimo a zero allora non è una buona approssimazione.

2 Probabilità e indipendenza

La probabilità serve per quantificare l'incertezza misurando la fiducia che un evento possa accadere.

2.1 Spazi di probabilità

Definizione 2.1.1 (Spazio campionario)**.** *Lo spazio di probabilità Ω è l'insieme di tutti gli esiti possibili (**eventi elementari**) ω dell'esperimento. Ogni affermazione sulle misure corrisponde ad un sottoinsieme $A \subset \Omega$ degli esiti che la soddisfa. Ognuna delle affermazioni può essere combinata logicamente con le operazioni insiemistiche.*

Definizione 2.1.2 (Eventi incompatibili)**.**

$$A \cap B = \emptyset \tag{14}$$

Definizione 2.1.3 (Esperimento composto)**.** *Se un esperimento è composto da una successione ordinata di n sotto-esperimenti, il suo spazio campionario è*

$$\Omega = \{(\omega_1, \omega_2, \dots, \omega_n) | \omega_1 \in \Omega_1, \dots, \omega_n \in \Omega_n\} \tag{15}$$

dove Ω_i è l'insieme degli esiti dell'i-esimo sotto-esperimento.

Definizione 2.1.4 (σ-algebre)**.** *L'insieme di tutti i sottoinsiemi di Ω che sia chiuso per le operazioni logiche come **unione** e **intersezione**.*

Osservazione 2.1.1. Se due eventi sono incompatibili la probabilità che si realizzi uno qualsiasi dei due è la somma delle probabilità dei singoli eventi.

Definizione 2.1.5 (Probabilità)**.** *È il grado di fiducia che un evento si realizzi. È compreso tra 0 e 1. Più precisamente, dato Ω un insieme e F una σ-algebra di parti di Ω, è una funzione $\mathbb{P} : F \to [0, 1]$ tale che:*

- *l'evento certo ha probabilità $\mathbb{P}(\Omega) = 1$*
- *(**σ-addittività**) se $(A_n)_{n=1,2,...}$ è una successione di eventi a due a due **disgiunti**, vale*

$$\mathbb{P}\left(\bigcup_{n=1}^{+\infty} A_n\right) = \sum_{n=1}^{+\infty} \mathbb{P}(A_n) \tag{16}$$

  e nel caso di finiti sottoinsiemi disgiunti

$$\mathbb{P}\left(\bigcup_{n=1}^{N} A_n\right) = \sum_{n=1}^{+N} \mathbb{P}(A_n) \tag{17}$$

Note 2.1.1. Si dice **trascurabile** un evento A tale che $\mathbb{P}(A) = 0$ e **quasi certo** un evento A tale che $\mathbb{P}(A) = 1$.

Proposizione 2.1.1. *Proprietà della probabilità:*

- $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$ *e di conseguenza* $\mathbb{P}(\emptyset) = 0$
- $B \subset A \Longrightarrow \mathbb{P}(A \setminus B) = \mathbb{P}(A) - \mathbb{P}(B)$
- $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$
- $\mathbb{P}(A \cup B \cup C) = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(A \cap B) - \mathbb{P}(A \cap C) - \mathbb{P}(B \cap C) + \mathbb{P}(A \cap B \cap C)$

Proposizione 2.1.2 (Limite di una successione di eventi)**.** *Data una successione di eventi $A_1, \dots, A_n, \dots$, questa può essere:*

- ***Crescente**: $A_n \subseteq A_{n+1}$ e quindi $A = \bigcup_{n=1}^{+\infty} A_n = \lim_{n\to\infty A_n}$*
- ***Decrescente**: $A_n \supseteq A_{n+1}$ e quindi $A = \bigcap_{n=1}^{+\infty} A_n = \lim_{n\to\infty A_n}$*

In entrambi i casi vale:

$$\mathbb{P}(A) = \lim_{n\to\infty} \mathbb{P}(A_n) \tag{18}$$

2.2 Probabilità discreta

Definizione 2.2.1 (Probabilità discreta)**.** *Dato* Ω ***numerabile***

$$\Omega = (\omega_1, \omega_2, \dots, \omega_n, \dots)$$

per ogni evento $A \subset \Omega$*, la misura di probabilità è:*

$$\mathbb{P}(A) = \sum_{\omega_i \in A} p_i = \sum_{\omega_i \in A} \mathbb{P}(\{\omega_i\}) \tag{19}$$

2.2.1 Probabilità uniforme su un insieme finito

Un esempio di probabilità discreta è quella uniforme su un insieme finito Ω, ovvero dove

$$p_1 = p_2 = \dots = p_N$$

In questo caso vale:

$$\mathbb{P}(A) = \frac{\#A}{\#\Omega} = \frac{\text{"casi favorevoli"}}{\text{"casi possibili"}} \qquad A \subseteq \Omega \tag{20}$$

2.2.2 Calcolo combinatorio

Alcune formule notevoli:

- **Sequenze ordinate con ripetizione** di k numeri da 1 a n: n^k
- **Ordinamenti possibili** di $\{1, \dots, n\}$: $n!$
- **Sequenze ordinate senza ripetizione** di k numeri di $1, \dots, n$

$$\frac{n!}{(n-k)!} \qquad 0 \leq k \leq n$$

- **Sottoinsiemi** di $\{1, \dots, n\}$ formati da k elementi

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \qquad 0 \leq k \leq n$$

2.2.3 Funzione di massa

Definizione 2.2.2 (Funzione di massa)**.** *Dato*

$$\Omega = \{x_1, x_2, \dots\} \subset \mathbb{R}$$

un sottoinsieme numerabile in cui ogni punto x_i *può contenere successioni (che possono andare a* $\pm\infty$*), la funzione di massa è*

$$\Omega \ni x_i \mapsto p(x_u) = \mathbb{P}(\{x_i\}) \in [0, 1] \tag{21}$$

Se poniamo che la probabilità di ogni altro punto non appartenente al sottoinsieme vale 0

$$x \neq x_i \implies p(x) = \mathbb{P}(\{x\}) = 0$$

allora possiamo estendere la funzione a $\mathbb{R}$ e dire che

$$\mathbb{P}(A) = \sum_{i : x_i \in A} p(x_i) \qquad \forall A \subseteq \mathbb{R} \tag{22}$$

Proposizione 2.2.1. *Valgono:*

$$p(x_i) \geq 0 \tag{23}$$

$$\sum_{i=1,2,\dots} p(x_i) = 1 \tag{24}$$

2.3 Probabilità condizionata

Quando si è a conoscenza della realizzazione di un evento, cambia la valutazione di probabilità di ogni altro evento.

Definizione 2.3.1 (Probabilità condizionata)**.** *Dati due eventi A, B con B non trascurabile, la probabilità condizionata di A rispetto a B è*

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \tag{25}$$

Proposizione 2.3.1 (Condizionamento ripetuto)**.** *Se l'intersezione di eventi $A_1 \cap \ldots \cap A_{n-1}$ non è trascurabile vale*

$$\mathbb{P}(A_1 \cap \ldots \cap A_n) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2|A_1) \cdot \ldots \cdot \mathbb{P}(A_n|A_1 \cap \ldots \cap A_{n-1}) \tag{26}$$

Definizione 2.3.2 (Partizione)**.** *Una partizione di Ω è una collezione di n eventi $B_1, \ldots, B_n$ a due a due disgiunti tali che*

$$B_1 \cup \ldots \cup B_n = \Omega \tag{27}$$

Definizione 2.3.3 (Sistema di alternative)**.** *È una partizione di Ω in eventi non trascurabili.*

Teorema 2.3.1 (Formula della probabilità o della fattorizzazione)**.** Dato $B_1, \ldots, B_n$ un sistema di alternative, per un qualunque evento A vale

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A|B_i)\mathbb{P}(B_i) \tag{28}$$

Definizione 2.3.4 (Formula di Bayes)**.** *Dati A e B due eventi non trascurabili vale*

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A|B)\mathbb{P}(B)}{\mathbb{P}(A)} \tag{29}$$

Definizione 2.3.5 (Formula di Bayes - Alternative)**.** *Dati A un evento e $B_1, \ldots, B_n$ un sistema di alternative vale*

$$\mathbb{P}(B_i|A) = \frac{\mathbb{P}(A|B_i)\mathbb{P}(B_i)}{\sum_{j=1}^{n} \mathbb{P}(AB_j)\mathbb{P}(B_j)} \tag{30}$$

2.4 Indipendenza

L'idea è che la conoscenza che si è realizzato un certo evento non modifica la valutazione di probabilità di un altro evento.

Definizione 2.4.1. *Dati n eventi $A_1, \ldots, A_n$, questi sono indipendenti se per ogni k con $2 \leq k \leq n$ e per ogni scelta di interi $1 \leq i_1 < i_2 < \ldots < i_k \leq n$ vale*

$$\mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \cdot \ldots \cdot \mathbb{P}(A_{i_k}) \tag{31}$$

Osservazione 2.4.1 (Complessità)**.** Il numero di uguaglianze da verificare per n eventi è

$$2^n - n - 1$$

Proposizione 2.4.1 (Spazi prodotto)**.** *Si consideri*

$$\Omega = \{a = (a_1, \ldots, a_n) | a_i = 0, 1\} = \{0, 1\}^n$$

su cui definiamo per ogni a la probabilità

$$\mathbb{P}(\{a\}) = p^{\#\{i: a_i = 1\}}(a-p)^{\#\{i: a_i = 0\}} = p^{\sum_{i=1}^{n} a_i}(a-p)^{n - \sum_{i=1}^{n} a_i}$$

E gli eventi

$$A_i = \{a \in \Omega : a_i = 1\} \qquad i = 1, \ldots, n$$

sono indipendenti tra di loro, così come i complementari A_i^c.

Osservazione 2.4.2. Due eventi possono essere indipendenti anche in presenza di una relazione causale. Viceversa due eventi possono essere dipendenti anche in assenza di una relazione causale.

2.5 Entropia di Shannon

Una misura di probabilità può essere uno strumento per quantificare l'informazione.

Definizione 2.5.1 (Entropia)**.** *Data una misura di probabilità discreta* $\mathbb{P}$ *su* $\Omega = \{x_1, \dots, x_n\}$*, con* $p_i = \mathbb{P}(\{x_i\})$*, la sua entropia è data dalla funzione*

$$H^{(n)}(p_1, \dots, p_n) = -\sum_{i=1}^{n} p_i \log(p_i) \tag{32}$$

Proposizione 2.5.1. *Valgono:*

1. *La funzione dell'entropia è* ***simmetrica****: scambiando* p_i *e* p_j *non cambia*
2. $H^{(n)}(1, 0, \dots, 0) = 0$
3. *È coerente tra* n *diversi:* $H^{(n)}(p_1 = 0, p_2, \dots, p_n) = H^{(n-1)}(p_2, \dots, p_n)$
4. $h^{(n)}(p_1, \dots, p_n) \leq H^{(n)}\left(\frac{1}{n}, \dots, \frac{1}{n}\right)$*, ovvero la massima entropia è data dalla distribuzione uniforme di probabilità*
5. *Data una probabilità su* $n \times m$ *oggetti* $\Omega = \{x_{11}, \dots, x_{ij}, \dots, xnm\}$ *con* $\mathbb{P}(x_ij) = q_{ij}$*, considerando gli eventi* $A_i = \{x_{i,1}, \dots, x_{i,m}\}$ *con* $\mathbb{P}(A_i) = p_1$ *vale*

$$H^{nm}(q_{11}, \dots, q_{ij}, \dots, q_{nm}) = H^{(n)}(p_1, \dots, p_n) + \sum_{i=1}^{n} p_i H^{(m)}\left(\frac{q_{i1}}{p_1}, \dots, \frac{q_{im}}{p_i}\right)$$

   ovvero l'entropia è data da quella relative al sistema di alternative A_i *più la media pesata delle entropie relative nei blocchi* A_i*.*

Teorema 2.5.1 (Shannon)**.** Una funzione che soddisfa le 5 proprietà ha la forma

$$cH^{(n)} \qquad c > 0 \tag{33}$$

2.6 Densità di probabilità

Definizione 2.6.1 (Densità di probabilità)**.** *Una funzione non negativa* $f : \mathbb{R} \to [0, +\infty]$*, integrabile e tale che*

$$\int_{-\infty}^{+\infty} f(x)dx = 1$$

La sua probabilità è

$$\mathbb{P}(A) = \int_A f(x)dx \qquad A \subseteq \Omega \tag{34}$$

Osservazione 2.6.1. La probabilità di ogni singolo punto è nulla

$$\mathbb{P}(\{t\}) = \int_{\{t\}} f(x)dx = 0 \tag{35}$$

e in generale

$$\mathbb{P}(A) = 0 \qquad \forall A \subset \mathbb{R} \tag{36}$$

3 Variabili aleatorie

Le variabili aleatorie sono funzioni dello spazio di probabilità. Permettono di scrivere osservazioni diverse fatte su uno stesso spazio Ω.

Definizione 3.0.1 (Variabile aleatoria)**.** *È una funzione*

$$X : \Omega \to \mathbb{R} \tag{37}$$

definita su uno spazio di probabilità.

3.1 Legge di una variabile aleatoria

Ad una variabile aleatoria sono associati eventi del tipo "X prende valori in un insieme $A \subseteq \mathbb{R}$:

$$\{X \in A\} = X^{-1}(A) = \{\omega \in \Omega : X(\omega) \in A\}$$

$$\mathbb{P}_X(A) = \mathbb{P}(X \in A) = \mathbb{P}(X^{-1}(A))$$

Definizione 3.1.1 (Legge di probabilità di una v.a.)**.** *La funzione $\mathbb{P}_X$ è una probabilità su $\mathbb{R}$ ed è detta* ***legge di probabilità*** *di X.*

Note 3.1.1. Quando due variabili aleatorie hanno la stessa legge di probabilità sono dette **equi distribuite**.

3.2 Tipi di variabili aleatorie

3.2.1 Variabili discrete

Definizione 3.2.1 (Variabile aleatoria discreta)**.** *Una variabile aleatoria è discreta se la sua immagine $X(\Omega) \subset \mathbb{R}$ è un sottoinsieme al più numerabile di $\mathbb{R}$ o se la sua legge di probabilità è discreta. Se $A \subseteq \mathbb{R}$ vale*

$$p_X(A) = \mathbb{P}(x \in A) = \sum_{x_i \in A} p_X(x_i)$$

3.2.2 Variabili continue

Definizione 3.2.2 (Variabile aleatoria continua)**.** *Una variabile aleatoria è detta con densità o continua se la sua legge di probabilità è definita da una densità f, ovvero se esiste una f tale che*

$$\mathbb{P}_X(A) = \mathbb{P}\{X \in A\} = \int_A f(x)dx \tag{38}$$

Se $A = [a, b]$ è un segmento, vale

$$\mathbb{P}\{X \in A\} = \mathbb{P}(a \leq X \leq b) = \int_a^b f(x)dx \tag{39}$$

3.3 Funzione di ripartizione

Per studiare una legge di probabilità di una variabile aleatoria è conveniente usare una funzione su $\mathbb{R}$.

Definizione 3.3.1 (Funzione di ripartizione)**.** *La funzione di ripartizione (c.d.f.) su X è*

$$F_X : \mathbb{R} \to [0, 1] \qquad F_X(x) = \mathbb{P}\{X \leq x\} \tag{40}$$

Proposizione 3.3.1. *Data $F = F_X$ la funzione di ripartizione di una variabile aleatoria X, valgono:*

- *F è* ***non decrescente***
$$x < y \implies F(X) \leq F(y) \tag{41}$$
- $\lim_{x\to-\infty} = 0$, $\lim_{x\to+\infty} F(x) = 1$

- *F è **continua** a destra*

$$\forall x \in \mathbb{R} \quad F(x_n) \to F(x) \tag{42}$$

per ogni successione $x_n \to x \qquad x_n \geq x$

Proposizione 3.3.2. *La probabilità che X cada in un dato intervallo $[a, b]$ per $a < b$ è*

$$\mathbb{P}\{a < X \leq b\} = F(b) - F(a) \tag{43}$$

3.3.1 Funzioni di variabili discrete

Data una variabile aleatoria discreta X, la sua c.d.f. che assume valori $x_1, x_2, \dots$ è

$$F_X(t) = \sum_{x_i \leq t} p(x_i) \tag{44}$$

Questa è una funzione a **gradini** che esegue un salto in ogni punto x tale che $\mathbb{P}(X = x) > 0$ di ampiezza pari alla probabilità di quel punto. Vale quindi

$$\mathbb{P}\{X = x\} = F(x) - F_{-}(x) \tag{45}$$

3.3.2 Funzioni di variabili continue

Quando la variabile ha densità f la sua funzione di ripartizione (**continua**) è

$$F(x) = \int_{-\infty}^{x} dt \tag{46}$$

o nel caso in cui è **continua a tratti** si ottiene:

$$f(x) = \frac{dF(x)}{dx} \tag{47}$$

3.4 β-quantile

Definizione 3.4.1 (β-quantile)**.** *Data una variabile aleatoria X ed un numero $0 < \beta < 1$ il β-quantile è:*

$$r_\beta = \inf\{r \in \mathbb{R} : F(r) \geq \beta\} \qquad \beta \in (0, 1) \tag{48}$$

Definizione 3.4.2 (Inversa generalizzata)**.** *L'inversa generalizzata di F è*

$$F^{\leftarrow} : (0, 1) \to \mathbb{R} \qquad F^{\leftarrow}(t) = \inf\{r \in \mathbb{R} : F(r) \geq t\} \tag{49}$$

Proposizione 3.4.1. *Valgono:*

- *Se F è strettamente crescente $F^{\leftarrow} = F^{-1}$*
- *$F^{\leftarrow}$ è sempre **non decrescente***
- $F^{\leftarrow}(F(t)) \leq t \qquad \forall t \in \mathbb{R}$
- $F(F^{\leftarrow}(t)) \geq t \qquad \forall t \in \mathbb{R}$
- $F^{\leftarrow}(t) \leq s \iff F(s) \geq t$

3.5 Variabili discrete notevoli

3.5.1 Binomiali

$$B(n,p) \tag{50}$$

Date n prove ripetute di un esperimento con **due esiti**, chiamiamo uno di questi *successo* con probabilità $0 < p < 1$. Sia X la variabile che conta il numero di successi $(0, 1, \dots, n)$. Vale:

$$\mathbb{P}(X = h) = \binom{n}{h} p^h (1-p)^{n-h} \qquad 0 \leq h \leq n \tag{51}$$

Ovvero dati h successi e $n-h$ insuccessi, calcoliamo il numero di modi di disporre i successi.

Osservazione 3.5.1. Date due successioni $x_1, x_2, \dots \in \mathbb{R}$ e $p_1, p_2, \dots \in [0, \infty)$ tale che $\sum_{i=1}^{\infty} p_1 = 1$, possiamo definire una variabile discreta tramite

$$\Omega = \mathbb{N} \qquad \mathbb{P}(\{k\}) = p_k \qquad X(k) = x_k \tag{52}$$

ovvero dove

$$\mathbb{P}_X(k) = \mathbb{P}(X = x_k) = p_k$$

Un caso particolare delle variabili binomiali è quando $n = 1$, ovvero le variabili di **Bernoulli**.

3.5.2 Geometriche

$$G(p) \tag{53}$$

Consideriamo la stessa situazione delle variabili binomiali ma definiamo X come l'istante del primo successo, ovvero il numero h tale che alla prova h-esima si verifichi il primo successo. Vale:

$$P(X = h) = (1-p)^{h-1} p \qquad h \in \mathbb{N}_0 \tag{54}$$

Questo corrisponde a dire, dato l'evento A_i successo della prova i-esima,

$$\mathbb{P}(X = h) = \mathbb{P}(A_1^c \cap A_2^c \cap \ldots \cap A_{h-1}^c \cap A_h) = \mathbb{P}(A_1^c) \cdot \mathbb{P}(A_2^c) \cdot \ldots \cdot \mathbb{P}(A_{h-1}^c) \cdot \mathbb{P}(A_h) = (1-p)^{h-1} p$$

Osservazione 3.5.2 (Assenza di memoria)**.** Le variabili geometriche hanno assenza di memoria, ovvero

$$\mathbb{P}\{X = n + h | X > n\} = \mathbb{P}\{X = h\} \tag{55}$$

3.5.3 Ipergeometriche

$$I(n,h,r) \tag{56}$$

Prendiamo ad esempio un'urna con n biglie di cui $0 \leq h \leq n$ sono bianche e $n-h$ nere. Estraiamo $r \leq n$ biglie senza reinserirle. La variabile che conta quante biglie estratte k sono bianche ha funzione di massa

$$\mathbb{P}(X = k) = \frac{\binom{h}{k}\binom{n-h}{r-k}}{\binom{n}{r}} \qquad k = 0, \dots, h \tag{57}$$

Proposizione 3.5.1 (Identità di Vandermonde)**.** *Date k biglie bianche e $r-k$ nere, il numero di scelte possibili è*

$$\binom{h}{k}\binom{n-h}{r-k}$$

mentre il numero totale di scelte è

$$\binom{n}{r}$$

Otteniamo quindi

$$\sum_{k=0}^{h} \binom{h}{k}\binom{n-h}{r-k} = \binom{n}{r} \tag{58}$$

che mostra anche $\sum_{k=0}^{h} \mathbb{P}(X = k) = 1$

3.5.4 Poisson

$$P(\lambda) \tag{59}$$

Una variabile è di Poisson quando

$$\mathbb{P}(X = h) = e^{-\lambda}\frac{\lambda^h}{h!} \qquad h \in \mathbb{N}, \lambda > 0 \tag{60}$$

Dato che è una buona approssimazione di una distribuzione binomiale quando n è grande, p è piccolo np è circa λ, possiamo dire che conta il numero di successi quando il numero di prove è alto e la probabilità è bassa. Viene anche detta degli **eventi rari** (eruzioni vulcaniche, particelle α emesse da una sorgente radioattiva).

3.6 Variabili con densità notevoli

Consideriamo i casi in cui esiste una funzione di densità non negativa di integrale unitario su tutto $\mathbb{R}$ f_X tale che

$$\mathbb{P}(X \in [a, b]) = \mathbb{P}(X \in (a, b)) = \int_a^b f_X(t)dt \tag{61}$$

3.6.1 Uniformi su intervalli

Dati due numeri reali $a < b$, la densità uniforme sull'intervallo $[a, b]$ è

$$f(t) = \begin{cases} \frac{1}{b-a} & a < t < b \\ 0 & \text{altrove} \end{cases} \tag{62}$$

La c.d.f. è

$$F(t) = \begin{cases} 0 & t \leq a \\ \frac{t}{b-a} & 0 < t \leq b \\ 1 & t > b \end{cases} \tag{63}$$

Ad esempio un numero preso a caso tra 0 e 1.

3.6.2 Esponenziali

Dato il parametro $\lambda > 0$ la densità è

$$f(x) = \begin{cases} \lambda e^{-\lambda t} & t > 0 \\ 0 & t \leq 0 \end{cases} \tag{64}$$

La c.d.f. è

$$F(t) = \begin{cases} 1 - e^{-\lambda t} & t > 0 \\ 0 & t \leq 0 \end{cases} \tag{65}$$

Descrive ad esempio il tempo di attesa tra due eventi aleatori, come tra le chiamate di un call center.

Osservazione 3.6.1. Questa variabile prende solo valori positivi

$$\mathbb{P}\{X \leq 0\} = 0$$

3.6.3 Pareto

Dati $x_m, \alpha > 0$ la densità è

$$f(t) = \begin{cases} \alpha x_m^\alpha t^{-1-\alpha} & t > x_m \\ 0 & t \leq x_m \end{cases} \tag{66}$$

La densità è non nulla dopo la soglia x_m e al diminuire di α ha una coda sempre più pesante. La c.d.f. è

$$F(t) = \begin{cases} 1 & t < x_m \\ 1 - (\frac{x_m}{t})^\alpha & t \geq x_m \end{cases} \tag{67}$$

Chiamata anche **power law**, serve a descrivere fenomeni in cui eventi estremi hanno una buona probabilità di avvenire, come la distribuzione della ricchezza nella società.

3.6.4 Gaussiane standard

Viene indicata con $N(0,1)$ e ha densità

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} \tag{68}$$

e c.d.f.

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{-t^2}{2}} dt \tag{69}$$

Osservazione 3.6.2. Questa densità è una funzione pari ($\varphi(x) = \varphi(-x)$). Di conseguenza, dati $x \in \mathbb{R}$ e $0 < \alpha < 1$, si ha

$$\Phi(-x) = 1 - \Phi(x) \qquad q_{1-\alpha} = -q_\alpha \tag{70}$$

Di conseguenza, se X è una variabile aleatoria $N(0,1)$, valgono

$$\mathbb{P}\{-t \leq X \leq t\} = \Phi(t) - \Phi(-t) = 1\Phi(t) - 1 \tag{71}$$

$$\Phi(0) = \mathbb{P}\{X \geq 0\} = \mathbb{P}\{X \leq 0\} = \frac{1}{2} \tag{72}$$

3.6.5 Gaussiane non standard

$$N(m, \sigma^2) \tag{73}$$

Data X una variabile Gaussiana Standard, dati $\sigma > 0$ e $m \in \mathbb{R}$, consideriamo la variabile aleatoria $Y = \sigma X + m$. La sua densità è

$$f_Y(t) = \frac{1}{\sigma} f_X\left(\frac{t-m}{\sigma}\right) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-m)^2}{2\sigma^2}} \tag{74}$$

mentre la sua c.d.f. è

$$F_Y(t) = \mathbb{P}\{Y \leq t\} = \mathbb{P}\{\sigma X + m \leq t\} = \mathbb{P}\left(X \leq \frac{t-m}{\sigma}\right) = \Phi\left(\frac{t-m}{\sigma}\right) \tag{75}$$

Osservazione 3.6.3. Vale:

$$\mathbb{P}\{a < Y < b\} = \mathbb{P}\left\{\frac{a-m}{\sigma} < X < \frac{b-m}{\sigma}\right\} \tag{76}$$

3.7 Trasformazioni di variabili con densità

Data la variabile aleatoria $X : \Omega \to \mathbb{R}$ con densità f e una funzione $h : \mathbb{R} \to \mathbb{R}$, vogliamo la densità della variabile aleatoria composta

$$Y : \Omega \to \mathbb{R} \qquad Y = h \circ X$$

Se è possibile calcolare la c.d.f. di Y

$$F_Y(y) = \mathbb{P}\{Y \leq y\} = \mathbb{P}\{h(X) \leq y\}$$

ed è **continua** e differenziabile, allora è sufficiente derivarla per ottenere la densità di Y.

Proposizione 3.7.1 (Cambio di variabile)**.** *Data X una variabile aleatoria con densità f_X, supportata su un intervallo aperto A (f_X nulla su A^c). Data una funzione $h : A \to B$, con B un intervallo aperto,* ***biunivoca***, ***differenziabile*** *e con* ***inversa differenziabile****. Allora $Y = h \circ X$ ha densità*

$$f_Y(y) = \begin{cases} f_X(h^{-1}(y)) \cdot \left|\frac{dh^{-1}(y)}{dy}\right| & y \in B \\ 0 & y \notin B \end{cases} \tag{77}$$

3.8 Valore atteso

Applichiamo il concetto di *media campionaria* e di *varianza campionaria* anche alle variabili aleatorie.

Definizione 3.8.1 (Valore atteso)**.** *Data una variabile discreta X con funzione di massa p_X, si dice che questa ha valore atteso se*

$$\sum_i |x_i| p_X(x_i) < +\infty$$

e vale

$$\mathbb{E}[X] = \sum_i x_i p_X(x_i) \tag{78}$$

Se X è con densità e

$$\int_{-\infty}^{+\infty} |x| f_X(x) dx < +\infty$$

allora il valore atteso è

$$\mathbb{E}[X] = \int_{-\infty}^{+\infty} t f_X(t) dt \tag{79}$$

Note 3.8.1. Il valore atteso è anche chiamato **momento primo** o speranza matematica.

Osservazione 3.8.1. Dato che il valore atteso dipende solo dalla funzione di massa o dalla densità, ovvero solo dalla legge $\mathbb{P}_X$ di X, allora se due variabili sono **equi distribuite** hanno anche lo stesso valore atteso.

Osservazione 3.8.2. Se X prende solo valori positivi, possiamo ammettere che $E[X]$ possa assumere il valore $+\infty$.
Nel caso discreto vuol dire che $x_1, x_2, \ldots$ sono sempre positivi e quindi ha senso

$$\mathbb{E}[X] = \sum_{i=1}^{+\infty} x_i p(x_i)$$

Nel caso con densità significa che $f(x) = 0 \quad x < o$ e quindi ha senso

$$\mathbb{E}[X] = \int_0^{+\infty} x f(x) dx \in [0, +\infty]$$

In generale

$$\mathbb{E}[|X|] < +\infty \tag{80}$$

Proposizione 3.8.1. *Valgono:*

- $\forall a, b \in \mathbb{R}$ *valgono* $\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$ *e* $\mathbb{E}[b] = b$
- $|\mathbb{E}[X]| \leq |\mathbb{E}[|X|]$
- $\mathbb{P}(X \geq 0) = 1 \Longrightarrow \mathbb{E}[X] \geq 0$

3.8.1 Valore atteso di trasformazioni

Supponiamo di voler calcolare il valore atteso di trasformazioni di una variabile aleatoria X, ovvero $Y = g(x) \quad g : \mathbb{R} \to \mathbb{R}$.

Proposizione 3.8.2 (Valore atteso di trasformazioni discrete)**.** *Se X è discreta e*

$$\sum_i |g(x_i)| p(x_i) < +\infty$$

allora

$$\mathbb{E}[g(X)] = \sum_i g(x_i) p(x_i) \tag{81}$$

Proposizione 3.8.3 (Valore atteso di trasformazioni con densità). *Se X è con densità e*

$$\int_{-\infty}^{+\infty} |g(x)|f(x)dx < +\infty$$

allora

$$\mathbb{E}[g(X)] = \int_{-\infty}^{+\infty} g(x)f(x)dx \tag{82}$$

3.8.2 Momenti

Definizione 3.8.2 (Momento). *La variabile aleatoria X ammette momento di ordine $n = 1, 2, \ldots$ se*

$$\mathbb{E}[|X|] < +\infty$$

e in quel caso si chiama $\mathbb{E}[X^n]$ il momento di ordine n.

Osservazione 3.8.3. Se una variabile discreta assume solo valori finiti, tutti i momenti sono finiti. Se una variabile con densità è diversa da 0 solo su un intervallo limitato, tutti i momenti sono finiti.

Proposizione 3.8.4. *Siano $1 \leq m < n$*

$$\mathbb{E}[|X|^n] < +\infty \Longrightarrow \mathbb{E}[|X|^m] < +\infty \tag{83}$$

*Ovvero se una variabile aleatoria ammette momenti fino a n, ammetterà anche tutti i suoi precedenti. In particolare vale la **disuguaglianza di Jensen**:*

$$\mathbb{E}[|X|^m]^{\frac{1}{m}} \leq \mathbb{E}[|X|^n]^{\frac{1}{n}} \tag{84}$$

Proposizione 3.8.5 (Disuguaglianza di Markov). *Se X è una variabile aleatoria a valori positivi e $a > 0$ vale*

$$a\mathbb{P}\{X \geq a\} \leq \mathbb{E}[X] \tag{85}$$

3.8.3 Varianza di una variabile aleatoria

Definizione 3.8.3 (Varianza). *Se X ammette momento secondo, la sua varianza è*

$$Var(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 \tag{86}$$

*e lo **scarto quadratico medio** o **deviazione standard** è*

$$\sigma(X) = \sqrt{Var(X)} \tag{87}$$

Proposizione 3.8.6 (Disuguaglianza di Chebyshev). *Data X una variabile aleatoria e $d > 0$, vale*

$$\mathbb{P}\{|X - \mathbb{E}| > d\} \leq \frac{Var(X)}{d^2} \tag{88}$$

Osservazione 3.8.4. La varianza di una variabile X vale 0 solo se questa è costante tranne che per un insieme trascurabile

$$\mathbb{P}\{|X - \mathbb{E}[X]| \neq 0\} = 0$$

3.8.4 Momenti notevoli

Vediamo i momenti delle variabili notevoli.

3.8.4.1 Variabili binomiali

Per una variabile di Bernoulli vale

$$\mathbb{E}[X^k] = p \qquad Var(X) = p - p^2 = p(1-p) \qquad k \geq 1 \tag{89}$$

Dato che una variabile Binomiale può essere vista come somma di variabili di Bernoulli, vale

$$\mathbb{E}[X] = np \qquad Var(X) = np(1-p) \tag{90}$$

3.8.4.2 Variabili di Poisson

Dato che assumono solo valori positivi:

$$\mathbb{E}[X] = \sum_{h=0}^{+\infty} he^{-\lambda}\frac{\lambda^h}{h!} = \lambda e^{-\lambda}\sum_{h=1}^{+\infty}\frac{\lambda h-1}{(h-1)!} = \lambda e^{-\lambda}\sum_{k=0}^{+\infty}\frac{\lambda^k}{k!} = \lambda \tag{91}$$

$$\mathbb{E}[X^2] = \lambda + \lambda^2 \qquad Var(X) = \lambda \tag{92}$$

3.8.4.3 Variabili uniformi su intervalli finiti

Data una variabile X con densità uniforme su $[a, b]$, vale:

$$\mathbb{E}[X] = \int_a^b \frac{x}{b-a}dx = \left.\frac{x^2}{2(b-a)}\right|_a^b = \frac{b^2-a^2}{2(b-a)} = \frac{a+b}{2} \tag{93}$$

$$\mathbb{E}[X^2] = \int_a^b \frac{x^2}{b-a}dx = \left.\frac{x^3}{3(b-a)}\right|_a^b = \frac{b^3-a^3}{3(b-a)} = \frac{a^2+ab+b^2}{3} \qquad Var(X) = \frac{(b-a)^2}{12} \tag{94}$$

3.8.4.4 Variabili esponenziali

Vale:

$$\mathbb{E}[X] = \int_0^{+\infty} \wedge x e^{-\lambda x}dx = \frac{1}{\lambda} \tag{95}$$

e più in generale

$$\mathbb{E}[X^n] = \frac{n!}{\lambda^n} \tag{96}$$

Quindi anche

$$Var(X) = \frac{1}{\lambda^2} \tag{97}$$

3.8.4.5 Variabili Gaussiane standard

Se X è Gaussiana Standard, notiamo che possiede tutti i momenti

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} |x|^n e^{-\frac{x^2}{2}}dx < +\infty \tag{98}$$

I momenti dispari valgono:

$$\mathbb{E}[X^{2h+1}] = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} x^n e^{-\frac{x^2}{2}}dx = \lim_{M\to\infty}\frac{1}{\sqrt{2\pi}}\int_{-M}^{+M} x^n e^{-\frac{x^2}{2}}dx = 0 \tag{99}$$

mentre quelli pari, guardando ad esempio il momento secondo

$$\mathbb{E}[X^2] = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} x^2 e^{-\frac{x^2}{2}}dx = \left.\frac{-xe^{-\frac{x^2}{2}}}{\sqrt{2\pi}}\right|_{-\infty}^{+\infty} + \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}}dx = 1 \qquad Var(X) = 1 \tag{100}$$

e più in generale

$$\mathbb{E}[X^{2h+2}] = (2h+1)\mathbb{E}[X^{2h}] \tag{101}$$

3.8.4.6 Variabili Gaussiane

Data $Y = \sigma X + m$, per linearità del valore atteso

$$\mathbb{E}[Y] = \mathbb{E}[\sigma X + m] = m \qquad Var(Y) = Var(\sigma X + m) = \sigma^2 Var(X) = \sigma^2 \tag{102}$$

3.9 Variabili doppie

Dato Ω uno spazio di probabilità e $X, Y : \Omega \to \mathbb{R}$, il vettore (X, Y) può essere visto come una funzione

$$\Omega \ni \omega \mapsto (X(\omega), Y(\omega)) \in \mathbb{R}^2 \tag{103}$$

La sua legge è una probabilità sui sottoinsiemi $A \subseteq \mathbb{R}^2$

$$\mathbb{P}_{X,Y)}(A) = \mathbb{P}((X, Y) \in A) = \mathbb{P}\{\omega \in \Omega : (X(\omega), Y(\omega)) \in A\} \tag{104}$$

Osservazione 3.9.1 (Insieme rettangolare)**.** Se $A = A_1 \times A_2$ è un sottoinsieme rettangolare vale

$$\{(X, Y) \in A\} = \{X \in A_1, Y \in A_2\} \tag{105}$$

Note 3.9.1. Con la virgola indichiamo l'intersezione di due condizioni

$$\{X \in A_1, Y \in A_2\} = \{X \in A_1\} \cap \{Y \in A_2\} = X^{-1}(A_1) \cap Y^{-1}(A_2) = (X, Y)^{-1}(A_1 \times A_2)$$

3.9.1 Distribuzioni marginali

Data una variabile doppia (X, Y) possiamo considerare separatamente le leggi delle due componenti $\mathbb{P}_X$ e $\mathbb{P}_Y$. Queste sono dette **leggi marginali**.
Se $I \subseteq \mathbb{R}$, valgono

$$\mathbb{P}_X(I) = \mathbb{P}_{(X,Y)}(I \times \mathbb{R}) \qquad \mathbb{P}_Y(I) = \mathbb{P}_{(X,Y)}(\mathbb{R} \times I) \tag{106}$$

Le distribuzioni marginali non contengono tutta l'informazione della legge $\mathbb{P}_{(X,Y)}$ e di conseguenza non si può ricostruire univocamente dalle prime. L'idea è che la legge totale codifica anche le relazioni tra le due leggi, cosa che le marginali non fanno

3.9.2 Variabili doppie discrete

Una variabile doppia (X, Y) è discreta se la sua immagine è concentrata in un insieme finito o numerabile di punti (x_i, y_j). La sua **distribuzione di probabilità** è

$$p(x_i, y_j) = \mathbb{P}(X = x_i, Y = y_j) \tag{107}$$

e se $A \subseteq \mathbb{R}^2$

$$\mathbb{P}_{(X,Y)}(A) = \mathbb{P}\{(X, Y) \in A\} = \sum_{(x_i,y_j)\in A} p(x_i, y_j) \tag{108}$$

Proposizione 3.9.1. *Se una variabile doppia è discreta con funzione di massa, lo sono anche le sue componenti*

$$p_X(x_i) = \sum_{j=1}^{+\infty} p(x_i, y_j) \qquad p_X(y_j) = \sum_{i=1}^{+\infty} p(x_i, y_j) \tag{109}$$

3.9.3 Variabili doppie con densità

Una variabile doppia (X, Y) è con densità se esiste una funzione $f : \mathbb{R}^2 \to [0, \infty)$ integrabile e con $\int\int_{\mathbb{R}^2} f(x, y)dxdy = 1$ tale che valga

$$\mathbb{P}_{(X,Y)}(A) = \mathbb{P}\{(X, Y) \in A\} = \int\int_A f(x, y)dxdy \qquad A \subseteq \mathbb{R}^2 \tag{110}$$

Teorema 3.9.1 (Teorema di Fubini-Tonelli)**.** Dato un insieme rettangolare $A = A_1 \times A_2$ vale

$$\int\int_A f(x, y)dxdy = \int_{A_1} \left(\int_{A_2} f(x, y)dy \right) dx = \int_{A_2} \left(\int_{A_1} f(x, y)dx \right) dy \tag{111}$$

Proposizione 3.9.2. *Se una variabile doppia ha densità, anche le sue componenti la hanno*

$$f_X(x) = \int_{-\infty}^{+\infty} f(x, y)dy \qquad f_Y(y) = \int_{-\infty}^{+\infty} f(x, y)dx \tag{112}$$

Osservazione 3.9.2. A differenza del caso discreto se X e Y sono con densità non è detto che anche $X, Y)$ la abbia. Ad esempio (X, X).

3.10 Indipendenza di variabili aleatorie

Definizione 3.10.1 (Variabili aleatorie indipendenti)**.** *Le variabili aleatorie $X_1, \dots, X_n : \Omega \to \mathbb{R}$ si dicono indipendenti se, presi comunque $A_1, \dots, A_n \subseteq \mathbb{R}$ vale*

$$\mathbb{P}(X_1 \in A_1, \dots, X_n \in A_n) = \mathbb{P}(X_1 \in A_1) \cdot \ldots \cdot \mathbb{P}(X_n \in A_n) \tag{113}$$

3.10.1 Indipendenza di variabili doppie

Proposizione 3.10.1 (Indipendenza di variabili doppie discrete)**.** *Date due variabili discrete X e Y con immagine nei punti x_i e y_j, sono indipendenti se e solo se vale*

$$p(x_i, y_j) = p_X(x_i) \cdot p_Y(y_j) \qquad \forall (x_i, y_j) \tag{114}$$

Proposizione 3.10.2 (Indipendenza di variabili doppie con densità)**.** *Date due variabili X e Y tale che (X, Y) abbia densità, sono indipendenti se e solo se vale*

$$f(x, y) = f_X(x) \cdot f_Y(y) \qquad \forall (x, y) \tag{115}$$

Osservazione 3.10.1. Due variabili aleatorie doppie possono avere le stesse distribuzioni marginali ma essere diverse, ad esempio perché in un caso le componenti sono indipendenti e nell'altro no.

3.10.2 Indipendenza di funzioni di variabili indipendenti

Funzioni di più variabili indipendenti sono indipendenti se la stessa variabile non compare in due funzioni diverse.

Proposizione 3.10.3. *Se $X, Y : \Omega \to \mathbb{N}$ sono variabili discrete a valori naturali e indipendenti e sia $Z = X + Y$ si ha*

$$p_Z(n) = \sum_{h=0}^{n} p_X(h) \cdot p_Y(n - h) \tag{116}$$

In particolare se X e Y sono binomiali $B(n, p)$ e $B(m, p)$, allora $Z = X + Y$ è binomiale $B(n + m, p)$.

Proposizione 3.10.4. *Se X e Y sono variabili con densità e indipendenti e sia $Z = X + Y$ si ha*

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(x) f_Y(z - x) dx = \int_{-\infty}^{+\infty} f_Y(y) f_X(z - y) dy \tag{117}$$

In particolare se X e Y sono Gaussiane $N(m_1, \sigma_1^2)$ e $N(m_2, \sigma_2^2)$, allora $Z = X + Y$ è Gaussiana $N(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$.

Proposizione 3.10.5. *Se X e Y sono variabili aleatorie indipendenti, allora per tutte le funzioni $h, k : \mathbb{R} \to \mathbb{R}$, anche $h(X)$ e $k(Y)$ lo sono.*

3.11 Correlazione

Proposizione 3.11.1. *Date due variabili aleatorie X e Y con valore atteso, allora $X + Y$ ha valore atteso e valgono:*

- $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$
- $X \geq Y \implies \mathbb{E}[X] \geq \mathbb{E}[Y]$

Proposizione 3.11.2. *Date due variabili aleatorie X e Y con valore atteso e* ***indipendenti****, allora XY ha valore atteso e vale:*

$$\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y] \tag{118}$$

Proposizione 3.11.3. *Se X e Y sono variabili aleatorie con valore atteso e* ***indipendenti****, allora per tutte le funzioni $h, k : \mathbb{R} \to \mathbb{R}$, vale*

$$\mathbb{E}[h(X)k(Y)] = \mathbb{E}[h(X)] \cdot \mathbb{E}[k(Y)] \tag{119}$$

Proposizione 3.11.4 (Disuguaglianza di Schwartz)**.** *Se X e Y hanno valore atteso, non è detto che il loro prodotto XY lo abbia ma se hanno momento secondo allora il loro prodotto ha valore atteso. Questo deriva da*

$$\mathbb{E}[|XY|] \leq \sqrt{\mathbb{E}[X^2]} \cdot \sqrt{\mathbb{E}[Y^2]} \tag{120}$$

3.12 Covarianza

Definizione 3.12.1 (Covarianza)**.** *La covarianza tra due variabili aleatorie X e Y con momento secondo finito è una misura della presenza di una relazione lineare tra le due e vale*

$$Cov(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \tag{121}$$

Quando vale 0 *le variabili sono* ***scorrelate****.*

3.13 Teoremi limite

Definizione 3.13.1. *Data una famiglia di variabili aleatorie $X_1, \dots, X_n, \dots$, sono i.i.d. se sono indipendenti ed equidistribuite.*
Equivalentemente lo sono se hanno tutte la stessa funzione di ripartizione

$$\mathbb{P}_{X_n}(t) = \mathbb{P}(X_n \leq t) = F(t) \qquad t \in \mathbb{R}, n = 1, 2, \dots \tag{122}$$

e se vale

$$\mathbb{P}(X_{k_1} \leq t_1, \dots, X_{k_n} \leq t_n) = F_{X_1}(t_1) \cdot \dots \cdot F_{X_n}(t_n) \qquad \forall k_1, \dots, k_n \in \mathbb{N}, t_1, \dots, t_n \in \mathbb{R} \tag{123}$$