Probabilità e Statistica per l'Informatica

UniShare

Davide Cozzi
@dlcgold

Gabriele De Rosa
@derogab

Federica Di Lauro
@f_dila

Indice

Capitolo 1

Introduzione

Questi appunti sono presi a lezione. Per quanto sia stata fatta una revisione è altamente probabile (praticamente certo) che possano contenere errori, sia di stampa che di vero e proprio contenuto. Per eventuali proposte di correzione effettuare una pull request. Link: `https://github.com/dlcgold/Appunti`.
Grazie mille e buono studio!

Capitolo 2

Breve Introduzione

La statistica è una discliplina, basata sulla matematica, con il fine lo studio quantitativo e qualitativo di un particolare fenomeno collettivo in condizioni di incertezza o non determinismo ed è usata in molti ambiti come ad esempio l'intelligenza artificiale, data science, robotica, domotica e tutte le analisi per poter ottenere ricavare delle informazioni sui dati.

Ormai i dati sono pervasivi e un loro studio è diventato necessario ed inoltre si parla spesso di target marketing, con una selezione dei possibili clienti infatti è usata in maniera massiccia nel mondo dello shopping online. Si ha l'*A-B testing*, per decidere tra due scelte la migliore e per la decisione si analizzano i dati presi da campioni di popolazione, utilizzando il *tasso di conversione*, ossia la percentuale di visitatori unici che hanno effettuato la azione su cui si sta effettuando il test.

In codesto corso vengono effettuati i seguenti argomenti:

1. statistica descrittiva
2. calcolo delle probabilità
3. distribuzioni notevoli
4. teoremi di convergenza
5. stima dei parametri
6. test di ipotesi parametrici
7. test di ipotesi non parametrici
8. regressione lineare

Capitolo 3

Statistica Descrittiva

La statistica descrittiva è una raccolta di metodi e strumenti matematici usati per organizzare una o più serie di dati al fine di trovarne delle simmetrie, periodicità o delle eventuali leggi, ossia si effettua una descrizione delle informazioni implicite ai dati.

Solitamente i dati disponibili non rappresentano tutta la popolazione ma un numero limitato di osservazioni effettuato su un *campione*, sottoinsieme selezionato della popolazione su cui si effettua l'analisi statistica, la cui efficacia rispetto alla popolazione dipende da quale sottoinsieme è stato scelto infatti non esiste un solo campione ma vi sono diversi modi per sceglierli, più o meno efficaci, per l'analisi statistica.

Quando si effettua un analisi statistica si vuole affermare qualcosa riguardo i **caratteri** della popolazione, ossia gli elementi su cui effettuo l'analisi statistica, che possono essere:

- **caratteri qualitativi**, indicanti qualità (colori, stili, materiali etc...) e anche dati non numerici in cui solitamente non è definita una *relazione d'ordine*
- **caratteri quantitativi**, maggiormente studiati dal corso, dati numerici in cui vengono definite *relazioni d'ordine*:
  - **discreti**, come i lanci di un dado, rappresentanti valori in $\mathbb{Z}$
  - **continui**, che assumono valori reali, come la temperatura, in $\mathbb{R}$

Supponiamo di considerare n elementi della popolazione e di rilevare, per ognuno di essi, il dato relativo al carattere quantitativo da esaminare, ossia definiamo l'insieme di dati $E = \{x_1, x_2, \dots, x_n\}$ con la numerosità, il numero di elementi considerati, pari a n.

In caso il carattere è quantitativo discreto è comodo raggruppare i dati considerando l'insieme di tutti i valori assumibili, detta **modalità del carattere** ed associare ad ognuno di esso il numero di volte che esso compare in E.
Si ha quindi N il numero di totalità del carattere e si definisce l'insieme di modalita $S = \{s_1, ..., s_N\}$ su cui si definiscono i seguenti valori statistici:

frequenza assoluta f_j numero di volte che si presenta un elemento di un campione

frequenza cumulata assoluta F_j somma delle frequenze assolute di tutte le modalità:

$$F_j = \sum_{k:s_k \leq s_j} f_k$$

frequenza relativa p_j rapporto tra la frequenza assoluta e il numero di elementi

$$p_j = \frac{f_j}{n}$$

frequenza cumulativa relativa P_j somma delle frequenze relativa di tutte le modalità:

$$P_j = \sum_{k:s_k \leq s_j} p_k$$

Si definisce **distribuzione di frequenza** una funzione $F : S \to \mathbb{N}$ che associa ad ogni modalità la corrispondente frequenza per cui esiste la distribuzione di frequenza assoluta, relativa, frequenza cumulativa assoluta e relativa.

Quando il carattere da studiare è continuo (o discreto con un gran numero di valori) è conveniente ricondursi a raggruppamenti come quelli appena trattati, per cui si suddivide S, l'insieme delle modalità, in alcune classi (sottoinsiemi di S) che formano una partizione e la scelta delle classi con cui si suddivide l'insieme S è del tutto arbitraria anche se è necessario che esse formino una partizione di S.
Le partizioni devono essere significative e sufficientemente numerose ed inoltre ad ogni classe si associano le grandezze:

- confine superiore e inferiore (valori estremi della classe)
- ampiezza (differenza tra confine superiore ed inferiore)
- valore centrale (media tra i due confini)

Nel caso in cui il carattere esaminato sia continuo occorre specificare come le classi sono chiuse, a destra o a sinistra, ossia specificare se gli elementi dell'indagine il cui dato coincide con il confine della classe sono da raggruppare all'interno della classe stessa oppure no.

3.1 Indici di tendenza Generale

Fino ad ora abbiamo visto come rappresentare i dati, sia discreti che continui, ora iniziamo ad analizzare gli indici che ci forniscono un valore che rappresenta un certo aspetto della serie di dati, incominciando dagli **indici di tendenza generale**:

media è la media aritmetica tra tutti i valori dei dati osservati

$$\overline{x} = \frac{1}{n}\sum x_i = \frac{x_1 + \cdots + x_n}{n}$$

Considerando le distribuzioni di frequenza definite, posssiamo fornire definizioni equivalenti di media:

$$\overline{x} = \frac{1}{n}\sum s_j f_j = \sum s_j p_j$$

La dimostrazione dell'uguaglianza di queste definizioni alternative è banale e si riconduce alla definizione di frequenza relativa ed assoluta.

mediana è l'elemento in mezzo ai valori dei dati, ordinati in maniera crescente in cui se il numero degli elementi n è dispari è l'elemento $\frac{n+1}{2}$ altrimenti è la media degli elementi di posto $\frac{n}{2}$ e $\frac{n}{2}+1$.

moda $\widetilde{x}$ valore o classe corrispondente alla massima frequenza assoluta e viene usata solitamente in caso sia impossibile definire la media e la mediana (quando si hanno caratteri qualitativi).
La moda non è unica infatti parliamo di:

- **distribuzione uni-modale** nel caso in cui vi sia un unica moda
- **distribuzione multi-modale** nel caso in cui vi siano più mode

Gli indici di tendenza centrale non sono utili per fornire informazioni circa l'omogeneità dei dati in quanto forniscono informazioni sui valori centrali e medi del campione statistico per cui per risolvere questo problema introduciamo i seguenti indici:

varianza è la media dello scarto quadratico di ogni elemento dalla sua media

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

La varianza ovviamente è tanto più grande quanto i singoli elementi si discostano dalla media, ossia significa che i dati in tal caso sono molto disomogenei. Come abbiamo già visto per la media sono presenti le seguenti definizioni alternative di varianza:

$$s^2 = \frac{1}{n}\sum_{j=1}^{N} f_j(s_j - \overline{x})^2$$

$$s^2 = \sum_{j=1}^{N} p_j(s_j - \overline{x})^2$$

Le prime due definizioni alternative derivano dalla definizione di frequenza assoluta e frequenza mentre l'ultima proviene da passaggi algebrici, dimostrati di seguito formalmente:

Dimostrazione.

$$\begin{aligned} s^2 &= \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 \\ &= \frac{1}{n}\sum(x_i^2 - 2x_i\overline{x} + \overline{x}^2) \\ &= \frac{1}{n}(\sum x_i^2 - 2\overline{x}\sum x_i + \sum \overline{x}^2) \\ &= \frac{1}{n}(\sum x_i^2 - 2n\overline{x}^2 + n\overline{x}^2) \\ &= \frac{1}{n}\sum x_i^2 - \overline{x}^2 \end{aligned}$$

Si dimostra $\sum x_i = n\overline{x}$ in quanto $\overline{x} = \frac{1}{n}\sum x_i$ e il resto sono soltanto passaggi algebrici elementari □

scarto quadratico medio misura quanto sono distanti gli elementi di un campione ed è calcolata come:

$$s = \sqrt{s^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

Nel calcolo della varianza si utilizza il quadrato per la differenza tra l'elemento e la sua media in quanto per come è definita la media si ha $\sum(x_i - \overline{x}) = 0$ e per evitare ciò si eleva la differenza tra un elemento e la sua media al quadrato.

La varianza è definito come il momento secondo rispetto alla media, espresso tramite la formula:

$$M_{k,y} = \frac{1}{n}\sum(x_i - y)^2$$

3.2 Il caso bidimensionale

Fino ad ora noi abbiamo considerato il caso unidimensionale in cui consideriamo solo un carattere del campione ma molte analisi richiedono di analizzare due o più caratteri del campione contemporaneamente, per riconoscere leggi ed analogie tra i diversi caratteri.
Considereremo solo due caratteri contemporanei, sia perché un analisi con più caratteri si comporta uguale e sia per non aggravare troppo la rappresentazione dei dati negli esempi e assumiamo che entrambi i caratteri sono di tipo quantitativo e discreto, in quanto se fossero quantitativi continui subirebbero prima un raggruppamento a classi.

L'insieme dei dati viene rappresentato come l'insieme delle coppie

$$E = \{(x_1, y_1), (x_2, y_2) \dots, (x_n, y_n)\}$$

mentre l'insieme delle coppie di valori assumibili si rappresentati con l'insieme

$$S = \{(s_j, u_k),\ j = 1 \dots N \quad k = 1 \dots M\}$$

Come abbiamo fatto anche per il caso unidimensionale definiamo le seguenti quantità:

frequenza assoluta (s_j, u_k) è la quantita f_{jk} corrispondente al numero di elementi con valore (s_j, u_k)

frequenza relativa

$$p_{jk} = \frac{f_{jk}}{n}$$

rapporto tra la frequenza e il numero di elementi

frequenza cumulata assoluta

$$F_{jk} = \sum_{r:s_r \le s_j; l:u_l \le u_k} f_{rl}$$

frequenza cumulata relativa

$$P_{jk} = \sum_{r:s_r \leq s_j; u_l \leq u_k} p_{rl}$$

frequenza

Si definisce *distribuzione di frequenza doppia* una qualsiasi funzione f, F, p, P che associa ad ogni coppia (s_j, u_k) la corrispondente frequenza ma esistono anche altri tipi di distribuzioni, infatti noi vediamo le **distribuzioni marginali**: distribuzioni dei singoli caratteri presi indipendemente degli altri.

Le distribuzioni marginali hanno la definizione delle seguenti funzioni:

frequenza assoluta marginale quantità di elementi $f_x j$ data dagli elementi di E, il cui primo carattere ha valore s_j

frequenza relativa marginale rapporto tra la frequenza assoluta marginale e il numero di osservazioni n.

frequenza cumulata assoluta marginale F_{xj} somma delle frequenze assolute marginali di tutti gli s_k con $s_k \leq s_j$

frequenza cumulata relativa marginale P_{xj} somma delle frequenze relative marginali di tutti gli s_k con $s_k \leq s_j$

Oltre agli indici definiti fino ad ora, esiste un indice che fornisce un grado di interdipendenza tra i due caratteri, indice importante in quanto molti problemi concreti necessitano di analizzare gradi di correlazione tra due o più serie di dati, iniziando prima di tutto da un esempio.

Considero due serie $\{x_i\}, \{y_i\}, \ i = 1, ..., n$ ponendo a confronto le variazione delle coppie di dati rispetto ai corrispondenti valori medi, considerando le coppie di scarti:

$$x_i - \overline{x}$$

$$y_i - \overline{y}$$

si ha una relazione di dipendenza tra i due caratteri se i due scarti corrispondono sistematicamente o quasi valori positivi o negativi.

Si definisce quindi la **covarianza** c_{xy}, dei dati o campionaria, delle due serie di dati :

$$c_{xy} = \frac{1}{n} \sum (x_i - \overline{x})(y_i - \overline{y})$$

La covarianza assume un valore positivo (negativo) che diviene grande in valore assoluto nel caso in cui i termini prodotto abbiano segni concordi e in questo caso si parla di serie statistiche fortemente correlate o per meglio dire

di dati delle serie fortemente correlati.
Nel caso opposto vale a dire nel caso in cui i dati delle serie siano incorrelati avremo che i prodotti avranno segni diversi (saranno discordi in segno) e la covarianza, per come definita, risulterà piccola in valore assoluto, prossima al valore 0.

Si ha anche la seguente formula per la covarianza:

$$c_{xy} = \frac{1}{n}\sum x_i y_i - \overline{xy}$$

Dimostrazione.

$$\begin{aligned}
c_{xy} &= \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) \\
&= \frac{1}{n}\sum_{i=1}^{n}(x_i y_i - x_i\overline{y} - y_i\overline{x} + \overline{xy}) \\
&= \frac{1}{n}(\sum x_i y_i - \overline{y}\sum x_i - \overline{x}\sum y_i + \sum \overline{xy}) \\
&= \frac{1}{n}(\sum x_i y_i - n\overline{yx} - n\overline{yx} + n\overline{yx}) \\
&= \frac{1}{n}(\sum x_i y_i - n\overline{xy}) \\
&= \frac{1}{n}\sum x_i y_i - \overline{xy}
\end{aligned}$$

□

Nel caso in cui i dati si riferiscano a caratteri quantitativi discreti, di cui è nota la distribuzione di frequenza doppia, è possibile utilizzare le seguenti formule per il calcolo della covarianza:

$$c_{xy} = \sum_{1}^{N}\sum_{1}^{M}(s_j - \overline{x})(u_k - \overline{y})p_{jk}$$

$$c_{xy} = \sum_{1}^{N}\sum_{1}^{M} s_j u_k p_{jk} - \overline{xy}$$

Date due serie di dati si ha che sono:

- **statisticamente incorrelate** se la loro covarianza è nulla

- **statisticamente indipendenti** se vale:

$$\forall j = 1, \ldots, N \;\; k = 1, \ldots, M \;\; p_{jk} = p_j p_k$$

con:

$$p_{jk} = \frac{f_{jk}}{n}$$
$$p_j = \frac{f_j}{n}$$
$$p_k = \frac{f_k}{n}$$

inoltre due serie di dati statisticamente indipendenti sono incorrelate mentre non è necessariamente vero il contrario, infatti:

$$\sum\sum(x_i - \overline{x})(y_i - \overline{y}) = \sum(x_i - \overline{x})\sum(y_i - \overline{y}) = 0$$

Nel caso bidimensionale, con variabili x e y, la covarianza si può rappresentare attraverso una matrice 2×2:

$$C = \begin{vmatrix} c_{xx} & c_{xy} \\ c_{xy} & c_{yy} \end{vmatrix} = \begin{vmatrix} var(x) & cov(x,y) \\ cov(x,y) & var(y) \end{vmatrix}$$

Per una misura indipendente dalla variabilità delle grandezze si usa la matrice di correlazione:

$$Corr = \begin{vmatrix} \frac{c_{xx}}{\sigma_x^2} & \frac{c_{xy}}{\sigma_x^2\sigma_y^2} \\ \frac{c_{xy}}{\sigma_x^2\sigma_y^2} & \frac{c_{yy}}{\sigma_y^2} \end{vmatrix} = \begin{vmatrix} 1 & corr(x,y) \\ corr(x,y) & 1 \end{vmatrix}$$

che ovviamente può crescere in m dimensioni.

3.3 Regressione Lineare

In molti casi ci si pone la questione se tra tali caratteri x ed y esista un legame di tipo funzionale o una relazione di tipo funzionale che ne descriva in modo soddisfacente corretto il legame realmente esistente.
Si parla di un'**analisi di regressione**, in cui si pensa ad uno dei due caratteri come variabile indipendente e si cerca una funzione che stabilisce la relazione tra i due caratteri.
Se fisso x, come **variabile indipendente**, cerco $y = f(x)$ in modo che essa descriva al meglio il legame tra la variabile indipendente x e il carattere y che a questo punto viene interpretato come **variabile dipendente**.

Si determina quindi la funzione f che minimizza le distanze tra i valori osservati del carattere y e quelli che si otterrebbero per il carattere y se la relazione che lega il carattere y ad x fosse proprio quella descritta da f, quindi cerco la funzione f che minimizza la quantità:

$$g(f) = \sum [f(x_i) - y_i]^2$$

dove il quadrato si utilizza affinché le distanze vengano tutte considerate con segno positivo.
Se f è vincolata ad essere una funzione lineare allora si parla di **regressione lineare**, con la retta rappresentata da:

$$y = mx + q$$

con q intercetta e m coefficiente angolare, tale per cui risulti minima la quantità:

$$g(m, q) = \sum [mx_i + q - y_i]^2$$

con $mx_i + q = f(x_i)$ che sono l'approssimazione alle y_i mediante f. Si ha che:

$$m = \frac{c_{xy}}{s_x^2}$$

$$q = \overline{y} - \frac{c_{xy}}{s_x^2}\overline{x}$$

Questo metodo consente di determinare la retta che meglio descrive la relazione tra i due caratteri senza peraltro fornire alcuna indicazione circa il grado di approssimazione che è in grado di offrire.
Per tale motivo è stata introdotta una nuova grandezza detta **coefficiente di correlazione lineare:**

$$r_{xy} = \frac{c_{xy}}{s_x s_y}$$

L'importanza di tale coefficiente deriva dal fatto che esso assume valori sempre appartenenti all'intervallo $[-1, 1]$ ed inoltre è nullo se le serie sono statisticamente incorrelate; il valore assoluto risulta tendente a 1 se le coppie sono tutte sulla retta $y = mx + q$ quindi rappresenta il grado di allineamento delle coppie di dati

Abbiamo accennato in precedenza al fatto che non si è sempre vincolati alla scelta di una retta tra le funzioni che possono descrivere la relazione tra le due serie di dati ma quanto esposto in precedenza può essere applicato anche nel caso in cui si considerino relazioni funzionali di diversa natura, la cui scelta può essere suggerita da una qualche impressione derivante da ispezioni visive dei dati o da altre forme di conoscenza circa il fenomeno analizzato, avendo quindi il modello non lineare di regressione.

Molte relazioni funzionali non lineari possono essere ricondotte a tali (lineari) con opportune trasformazioni delle variabili, infatti prendendo per esempio la relazione:

$$y = a \cdot e^{bx}$$

che si può riscrivere come:

$$\widetilde{y} = \beta \cdot \widetilde{x} + \alpha$$

con:

$$\widetilde{y} = \log(y)$$

$$\widetilde{x} = x$$

$$\alpha = \log(a)$$

$$\beta = b$$

si ottiene quindi una sorta di curva e non più una retta.

La determinazione dei coefficienti a e b che meglio permettono di approssimare una serie di punti $\{x_i, y_i\}$ può essere effettuata riconducendosi ad una regressione lineare ovvero determinando i coefficienti α, β che meglio approssimano, linearmente, la serie dei punti $\{\widetilde{x}_i, \widetilde{y}_i\}$, con:

$$\widetilde{y}_i = \log(y_i)$$

$$\widetilde{x}_i = x_i$$

Una volta determina ti tali coefficienti il calcolo di a e b risulta immediato.

Ecco alcune funzioni riconducibili a lineari:

$$y = a \log(x) + b$$

$$y = ax^b$$

$$y = \frac{1}{a + b \cdot e^{-x}}$$

Capitolo 4

Calcolo delle probabilità

Disciplina di carattere matematico che permette di affrontare l'analisi delle situazioni che hanno un esito imprevedibile a priori e pertanto con conseguenze incerte ed è lo strumento di base della statistica, che invece trae conclusioni su una popolazione, utilizzando i dati osservati su una collezione di individui appartenenti alla popolazione: estraendo un conveniente campione e analizzandolo, si spera di poter trarre conclusioni circa l'intero infatti da un campione traggo un'inferenza riguardo la popolazione.

Si hanno 4 impostazioni per la probabilità:

1. classica
2. frequentista
3. soggettivista
4. assiomatica

Iniziamo a considerare l'impostazione assiomatica, supponendo di voler studiare una situazione con un insieme Ω di possibili esiti ben distinti tra loro, detto *spazio campione*, in cui tutti i suoi sottoinsiemi A sono eventi(elementari in caso contengono solo un elemento altrimenti sono eventi composti) e ad ogni evento si ha una quantità numerica $P(A)$ detta probabilità.

Un passo fondamentale nel superare le polemiche sull'interpretazione del concetto di probabilità fu compiuto da Kolmogorov (1933) che abbandonò il tentativo di fondare la teoria della probabilità su una interpretazione sperimentale del concetto e costruire una teoria secondo una **impostazione assiomatica**, trascendendo il significato effettivo di probabilità, rinviandone l'interpretazione al momento delle applicazioni.

Due eventi sono **incompatibili** se non hanno elementi in comune, formando un intersezione vuota e l'insieme delle parti di Ω, $\wp(\Omega)$, definisce ovviamente l'insieme di tutti i sottoinsiemi; la probabilità deve essere definita per tutti gli elementi di $\wp(\Omega)$ con particolari proprietà (algebre di Boole o σ-algebre).

Si dice **misura di probabilità** ogni applicazione $P : \wp(\Omega) \to \mathbb{R}_0^+$ che associa un valore reale ad ogni sottoinsieme di Ω e per cui valgono le seguenti proprietà:

- $\forall A \subseteq \Omega$ esiste ed è un unico numero $P(A) \geq 0$
- $P(\Omega) = 1$
- data la famiglia $\{A_i \in I \subseteq N\}$ di eventi incompatibili vale:

$$P\left(\bigcup_{i\in I} A_i\right) = \sum_{i\in I} P(A_i)$$

Ogni misura di probabilità è una funzione che assegna valori numerici a sottoinsiemi di Ω e non ai suoi elementi (eventi elementari), come contrariamente si è portati a credere intuitivamente, invece avviene che la funzione di misura di probabilità associa un valore agli eventi elementi come una conseguenza dell'assegnazione di valori ad eventi composti.
La definizione di probabilità non fornisce indicazioni su quali valori numerici devono essere assegnati dato che dipende, come sempre dal particolare problema da analizzare.

Dalla definizione assiomatica derivano facilmente le seguenti proprietà aggiuntive: Sia P una misura di probabilità definita sull'insieme delle parti $\wp(\Omega)$ di uno spazio campione Ω ($\overline{A}$ evento complementare di A cioè somma di probabilità pari a 1) allora:

- $\forall A, B \subseteq \Omega$eventi anche incompatibili risulta che $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

  Dimostrazione. Si dimostra che:

  $$P(A \cup B) = P(A \cap \overline{B}) + P(A \cap B) + P(\overline{A} \cap B)$$

  Si ricava e si sa che sono soddisfatte le seguenti equazioni

  $$P(A) = P(A \cap \overline{B}) + P(A \cap B)$$

  $$P(B) = P(A \cap B) + P(\overline{A} \cap B)$$

quindi si ricava dall'equazione iniziale, sostituendo $P(A)$ e $P(B)$ la seguente equazione:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

□

- $\forall A \subseteq \Omega \rightarrow P(\overline{A}) = 1 - P(A)$

  Dimostrazione.

$$\forall A \subseteq \Omega \rightarrow P(\overline{A}) = 1 - P(A)$$
$$\downarrow$$
$$1 = P(A \cup \overline{A} = P(A) + P(\overline{A}) - P(A \cap \overline{A}) = P(A) + P(\overline{A}) - 0$$

□

- $\forall A \subseteq \Omega$ risulta che $P(A) \leq 1$
- $\forall A, B \subseteq \Omega, A \subseteq B$ risulta che $P(A) \leq P(B)$

si ha che il terzo assioma della probabilità generalizza la definizione della $P(A \cup B)$, nel caso di tutti gli eventi incompatibili mentre la terza e la quarta proprietà si dimostrano facilmente, considerando gli elementi basilari di teoria degli insiemi e dei 3 assiomi della probabilità.

Si ha uno **spazi campione con elementi equiprobabili** se dato $\Omega = \{1, 2, \dots, N\}$ tale che:

$$P(\{1\}) = \cdots = P(\{N\}) = p$$

Da questa formula si ricava facilmente le seguenti due equazioni, fondamentali per il calcolo della probabilità:

$$P(\{1\}) + \cdots + P(\{N\}) = Np = p(\Omega) = 1$$
$$P(\{i\}) = p = \frac{1}{n}$$

Da quest'ultima equazione, considerando anche il terzo assioma della probabilità si ricava che la probabilità di evento E è data dal rapporto tra i numeri degli eventi elementari dell'elemento e la numerosità degli elementi dello spazio campione.

In caso si facciano due esperimenti: se l'esperimento 1 può avere n possibili esiti equiprobabili, e l'esperimento 2 può avere m possibili esiti equiprobabili, allora i due esperimenti hanno $n \times m$ possibili esiti. Se si espande a r esperimenti si avranno $n_1 \times n_2 \times ... \times n_r$ possibili esiti.

4.1 Coefficiente Binomiale

Una **permutazione** è un ordinamento senza ripetizione di simboli.
Si voglia determinare il numero di differenti gruppi di r oggetti che è possibile costruire (formare) usando n oggetti diversi (non importa lordine degli oggetti). Si ottiene che:

$$\frac{n(n-1)...(n-r+1)}{r!} = \frac{n!}{(n-r)!\,r!} = \binom{n}{r}$$

detto **coefficiente binomiale**, ovvero il numero delle combinazioni di n oggetti presi a gruppi di r.
Si ha inoltre che:

$$P(E) = \frac{\text{num eventi elementari}\, E}{N}$$

se ho più casi faccio:

$$P = \frac{modo_1 \cdot modo_2}{tutti\ i\ possibili\ gruppi}$$

4.2 Probabilità condizionata

In molti casi, quando si desideri studiare un fenomeno con comportamenti aleatori, oppure si vuole effettuare un esperimento con esiti imprevidibili a priori, si usano alcune informazioni complementari per restringere il campo dei possibili risultati, raggiungibili dal fenomeno da analizzare e per poter trattare matematicamente queste situazioni viene introdotto il concetto di **probabilità condizionata**.

Definizione 1. *Siano dati uno spazio campione Ω ed una misura di probabilità P definita su $\wp(\Omega)$ e secondo limpostazione assiomatica di probabilità, considerati due eventi, A e B con $P(B) > 0$, viene definita probabilità dell'evento A condizionata dall'evento B la quantità:*

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

Questa definizione non è limitata solo alla probabilità assiomatica, infatti secondo l'approccio frequentista si definisce la probabilità dell'evento A, limitandosi a considerare solo i casi appartenenti a B come totale dei casi possibili.

Si possono avere diversi casi:

- la probabilità dellevento B sotto condizionamento diviene maggiore rispetto a quella che avrebbe assunto senza condizionamento
- il valore di probabilità diminuisca a fronte del condizionamento
- può accadere che il condizionamento rispetto ad un evento non inficia in alcun modo la probabilità di un altro evento ed in sto caso i due eventi sono **stocasticamente indipendenti**.

Due eventi $A, B \in \wp(\Omega)$ sono stocasticamente indipendenti in caso in cui $P(A) = P(A|B) = P(B)$ da cui risultano, attraverso semplici trasformazioni ed applicazioni delle formule, i seguenti risultati:

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

L'ultima formula presentata può essere generalizzata ad n eventi, venendo chiamata **formula del prodotto**, nel seguente modo:

Definizione 2. *Sia $n \in \mathbb{N}_+$ e data la famiglia $\{A_i i = 1, \cdots, n\}$ di sottoinsiemi di Ω allora risulta*

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(A_n|A_1 \cap A_2 \cap \cdots \cap A_{n-1}$$

Considerata una partizione dello spazio campione Ω, si hanno le seguenti formule:

- $P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i)$ (formula delle probabilità totali)
- $P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{i=j}^{n} P(B|A_j)P(A_j)}$ (formula di Bayes)

4.3 Variabili Aleatorie

In base a quanto visto fino ad ora per affrontare problemi di calcolo delle probabilità occorre di volta in volta definire in maniera appropriata lo spazio campione e la misura di probabilità.
Questo fatto comporta delle difficoltà se ci si pone come obiettivo la formulazione di una teoria generale della probabilità, in quanto spazio campione e misura di probabilità sono diversi ogni volta, inoltre in quasi tutti i problemi concreti si ha a che fare con situazioni il cui esito, imprevedibile a priori, è di tipo numerico, è questo il caso, ad esempio del lancio di un dado, o più realisticamente della spesa complessiva per la costruzione di un immobile o del ricavato della sua vendita.

Queste considerazioni hanno portato i matematici ad introdurre delle opportune trasformazioni, chiamate **variabili aleatorie**, che consentono di ricondursi sempre ad $\mathbb{R}$ come spazio campione ed a considerare quali suoi sottoinsiemi tutti gli intervalli del tipo (a, b) o $[a, b]$ con $-\infty \leq a < b \leq +\infty$ in cui sono comprese tutte le possibili unioni ed intersezioni(finite o infinite) e i loro complementi.

Dato uno spazio campione Ω si definisce **variabile aleatoria o casuale** un'applicazione $X : \Omega \to \mathbb{R}$ che associa un numero reale ad ogni elemento di Ω.
In base a questa definizione è possibile assegnare delle probabilità ad eventi del tipo $X \in B \subseteq \mathbb{R}$ in quanto $P(X \in B) = P(\{\omega \in \Omega : X(\omega) \in B\})$ ma ciò appesantisce troppo la notazione per cui utiliziamo per comodità $P(X \in B)$ con però $X \in B$ un evento in Ω, invece di rappresentare la funzione della variabile aleatoria.

Dal punto di vista intuitivo la definizione di variabile aleatoria è poco chiara ma nel nostro caso assumiamo, senza andare ad affrontare in maniera dettagliata, che essa rappresenta le quantità d'interesse determinate dall'esecuzione dell'esperimento su cui si vuole conoscere la probabilità di avvenimento dell'evento.

Come notazione adotteremo quella solitamente utilizzata in campo statistico, indicando con lettere maiuscole le variabili aleatorie e con lettere minuscole le rispettive possibili realizzazioni.
Essendo imprevedibile a priori il valore assunto da una variabile aleatoria, tutto ciò che si può fare relativamente ad essa è esprimere delle valutazioni di tipo probabilistico sui valori che essa assumerà.

Per tale ragione ad ogni variabile aleatoria X è associata una funzione, la **funzione di ripartizione** $F_X : \mathbb{R} \to [0, 1] \subset \mathbb{R}$ definita come:

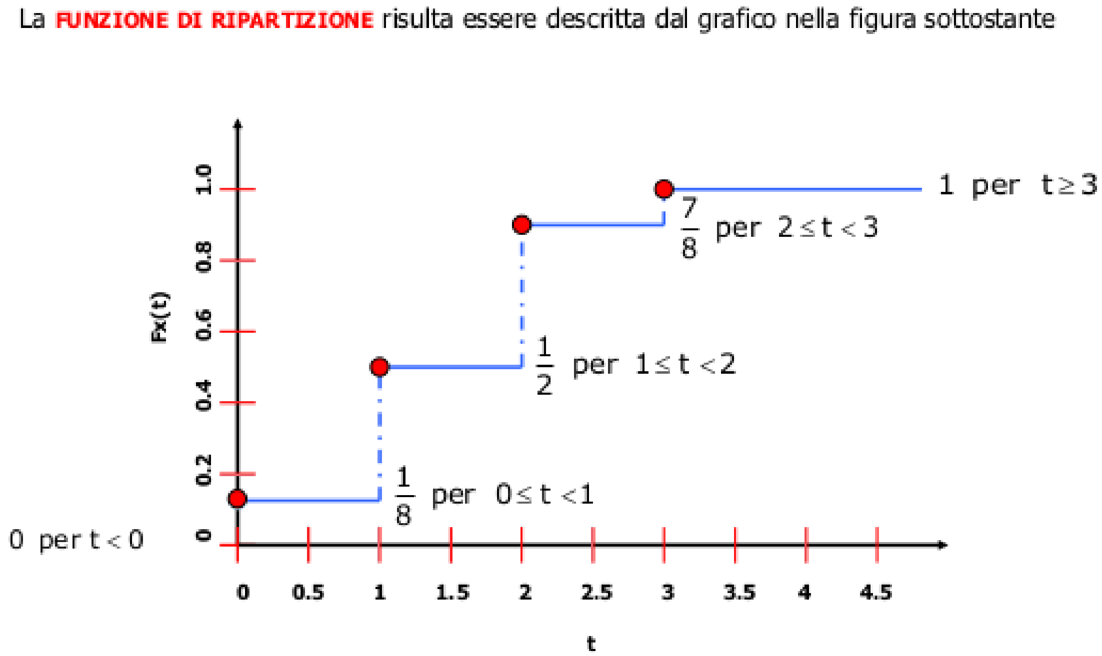
$$F_X(t) = P(X \leq t) \quad \forall t \in \mathbb{R}$$

che ci indica la probabilità che la variabile casuale X assuma un valore minore o uguale a t.

Come si può notare nella figura img:ripartizione 4.1, la funzione di ripartizione è una funzione a gradini, in cui ad ogni valore intero si ha la presenza di un punto di discontinuità, con un salto in avanti.

Avendo definito la funzione di ripartizione, adesso tutte le questioni riguardanti la variabile casuale X possono trovare una soluzione attraverso di essa, infatti supponiamo di calcolare $P(\{a < X \leq b\})$ con l'evento $X \leq b$ esprimibile come l'unione dei due eventi indipendenti $X \leq a$ e $a < X \leq b$ da

Figura 4.1: Funzione di ripartizione img:ripartizione



cui si ricava, usando il terzo assioma della probabilità la seguente formula:

$$P(\{X \leq b\}) = P(\{X \leq a\}) + P(\{a < X \leq b\}) \text{ da cui si ricava}$$
$$P(\{a < X \leq b\}) = F(b) - F(a)$$

Questa formula ci permette di determinare la probabilità che una variabile casuale possa assumere valori in intervalli reali e ciò ha un notevole utilizzo in statistica e nella probabilità.

In genere la funzione di ripartizione non è nota, altrimenti tutti gli eventi della nostra vita sarebbero facilmente analizzabili senza nessuna incertezza, ossia ad esempio si sarebbe in grado di prevedere come vincere al superenalotto e tutti i giochi d'azzardo, per cui l'obiettivo della statistica è di determinarla o determinare le grandezza ad essa associate mentre la probabilità e le sue applicazioni assumono che essa sia sempre nota.

Si dimostra che sono delle funzioni di ripartizione tutte e sole le funzioni del tipo $F : \mathbb{R} \to [0, 1]$ che godono simultaneamente delle seguenti proprietà:

- F è monotona crescente
- $\lim_{t\to+\infty} F(t) = 1$
- $\lim_{t\to-\infty} F(t) = 0$

- $\lim_{t \to t_0^+} F(t) = F(t_0), \ \forall t_0 \in \mathbb{R}$

Le variabili aleatorie si distinguono in due categorie, in base a che valori può assumere l'insieme di valori S di supporto:

- **variabili aleatorie discrete**: l'insieme S è finito oppure costituito da un insieme infinito di valori discreti, ossia definiti in $\mathbb{Z}$.
- **variabili aleatorie continue**: l'insieme S assume valori infiniti continui

Iniziamo ad analizzare prima le variabili discrete, più semplici da analizzare per poi considerare il caso continuo, in cui si estendono i valori assumibili dalle variabili.

4.3.1 Variabile Aleatoria Discreta

Una variabile aleatoria è detta **variabile aleatoria discreta** nel caso in cui l'insieme dei valori S che essa può assumere è finita oppure da un infinito di valori discreti, in cui si associa anche, oltre alla funzione di ripartizione, una funzione di probabilità assunta da valori specifici.
Sia S il supporto della variabile aleatoria X e si definisce**distribuzione discreta di probabilità** la funzione: $p_X : \mathbb{R} \to [0, 1]$ così definita:

$$\begin{cases} p(X = t) \forall t \in S \\ 0 \text{altrimenti} \end{cases}$$

Una funzione rappresenta una distribuzione di probabilità in caso sia soddisfatte entrambe le seguenti proprietà:

- $p_X(t) \geq 0 \quad \forall t \in R$
- $\sum_{s \in S} p_X(s) = 1$

L'ultima proprietà deriva dal secondo e dal terzo assioma della probabilità.

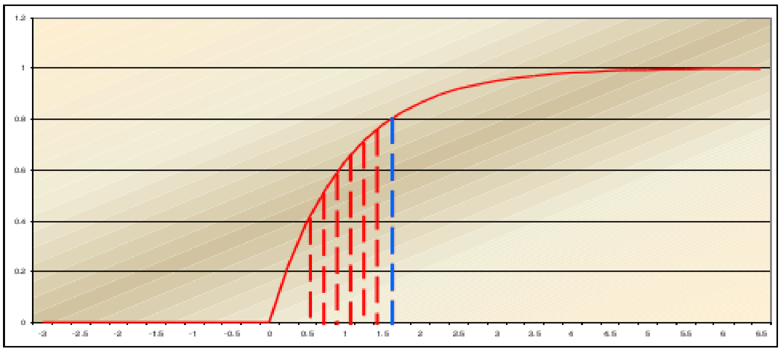
Tra le funzioni di ripartizione e le distribuzioni discrete esiste una corrispondenza biunivoca, in quanto si hanno le seguenti equivalenze:

- $F_X(t) = \sum_{s \in S : s \subseteq t} p_X(s) \ \ \forall t \in \mathbb{R}$
- $p_X(s) = F_X(s) - \lim_{t \to s^-} F_X(t) \ \ \forall s \in S$

Dalla prima di tali relazioni se ne deduce che le funzioni di ripartizione delle variabili aleatorie discrete presentano dei *salti* in corrispondenza dei valori s mentre sono costanti per gli altri valori, come si può anche notare nella figura `img:ripartizione` 4.1

Figura 4.2: Funzione di ripartizione continua

img:ripartizione(



4.3.2 Variabile Aleatoria Continua

Una variabile aleatoria è detta continua nel caso in cui la corrispondente funzione di ripartizione F_X sia continua, ed in particolare viene detta **assolutamente continua** se esiste una funzione $f_X : \mathbb{R} \to \mathbb{R}_+$ tale che $F_X(t) = \int_{-\infty}^{t} f_X(u)du \ \forall t \in \mathbb{R}$. Una tale funzione, in caso esista ossia è definito l'integrale, viene detta **densità di probabilità** di X.

È detto poi **supporto** della variabile X l'insieme $S = \{t \in \mathbb{R} : f_X(t) \neq 0\}$ e si osservi che se la densità di probabilità esiste allora la sua funzione di ripartizione è una sua primitiva.

Per semplicità supporremo nel seguito che le variabili aleatorie assolutamente continue abbiano funzione di ripartizione derivabile e che la funzione di densità di probabilità sia la derivata della funzione di ripartizione.
Come per le distribuzioni discrete di probabilità anche le funzioni di densità di probabilità per essere tali devono soddisfare le seguenti due proprietà:

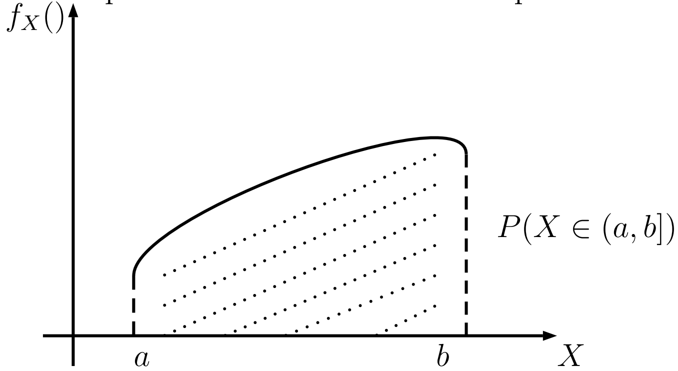
1. $f_X(t) \geq 0 \ \ \forall t \in \mathbb{R}$
2. $\int_{-\infty}^{+\infty} f_X(t)dt = 1$

La probabilità che una variabile aleatoria continua (o assolutamente continua) assuma un ben determinato valore è sempre nulla, in quanto se X è una variabile aleatoria continua allora $\forall t_0 \in R$ risulta

$$\begin{aligned} P(X = t_0) &= P(X \leq t_0) - \lim_{t \to t_0^-} P(X \leq t) \\ &= F_X(t_0) - \lim_{t \to t_0^-} F_X(t) \\ &= F_X(t_0) - F_X(t_0) = 0 \end{aligned}$$

Pertanto quando si pensa a variabili aleatorie continue, non ha mai senso domandarsi,quale sia la probabilità che esse assumano valori esatti, ma ha

Figura 4.3: Esempio di calcolo della densità di probabilità continua



graficoRipartizione

senso invece domandarsi quale è la probabilità che essi assumano specifici valori appartenenti in specifici intervalli dell'asse reale.

Per calcolare la probabilità che una variabile casuale continua X assuma un valore in un intervallo $(a, b] \subseteq \mathbb{R}$ è possibile far ricorso alla seguente formula:

$$P(X \in (a, b]) = \int_a^b f_X(u)du \ \ \forall t \in \mathbb{R}$$

ossia si avrebbe che $P(X \in (a, b]) = F_X(b) - F_X(a) = \int_{-\infty}^{b} f_X(u)du - \int_{-\infty}^{a} f_X(u)du$.

In pratica la probabilità che sia soddisfatto l'evento $X \in (a, b]$ corrisponde all'area sottesa dalla densità f_X nell'intervallo $(a, b]$, come si può notare nella figura graficoRipartizione 4.3

quindi:

$$[P(X \in (a, b]) = \int_a^b f_X(u)du \ \ \forall a, b \in \mathbb{R}, \ \ a < b$$

inoltre essendo $P(X = a) = 0$ si ha:

$$[P(X \in (a, b]) = p(X \in [a, b]) \ \ \forall a, b \in \mathbb{R}, \ \ a < b$$

Si è visto che per definizione:

$$F_X(t) = \int_{-\infty}^{t} f_X(u)du \ \ \forall t \in \mathbb{R}$$

Ovvero che la funzione di ripartizione è la funzione integrale della funzione di densità di probabilità. Quindi, data la funzione di ripartizione, si ottiene la funzione di densità di probabilità tramite derivazione:

$$\frac{d}{dt} F_X(t) = f_X(t)$$

4.3.3 Variabili Aleatorie Multidimensionali

In molti casi è lecito considerare situazioni (esperimenti) il cui esito è rappresentato, anziché da un valore numerico, da una coppia o da una n-pla di valori, in tal caso si parla di *variabili aleatorie multidimensionali.*
Anche qui le variabili sono definite da uno spazio campione Ω a $\mathbb{R}^n$ con n dimensione della variabile ed è comodo pensare a queste variabili aleatorie come a risultati esprimibili da n-ple di valori numerici, quindi si considerano funzioni che esprimano le valutazioni probabilistiche sui valori assumibili dalle variabili.

Consideriamo quindi le **variabili aleatorie bidimensionali assolutamente continue**, anche se ovviamente si può estendere a n variabili anche non continue.

Sia quindi una variabile aleatoria così definita:

$$(X, Y) : \Omega \to \mathbb{R}^2$$

con Ω, uno spazio campione al quale è associata una probabilità P definita sui sottoinsiemi di Ω.
Si definisce la **funzione di ripartizione congiunta** la funzione bidimensionale:

$$F_{X,Y}(t, s) : \mathbb{R}^2 \to [0, 1] \subseteq \mathbb{R}$$

definita come:

$$F_{X,Y} = P(\{X \leq t\} \cap \{Y \leq s\} \;\; \forall (t, s) \in \mathbb{R}$$

inoltre se la variabile (X, Y) è assolutamente continua esiste la **funzione dei densità congiunta**:

$$f_{X,Y} : \mathbb{R}^2 \to \mathbb{R}_+$$

tale che:

$$F_{X,Y}(t, s) = \inf\limits_{-\infty}^{t} \inf\limits_{-\infty}^{s} f_{X,Y}(u, v) du\, dv \;\; \forall (t, s) \in \mathbb{R}^2$$

conoscendo una delle due funzioni sopra è possibile determinare la probabilità che la coppia (X, Y) assuma valori in un qualsiasi sottoinsieme rettangolare $(a_1, b_1] \times (a_2, b_2] \in \mathbb{R}^2$, in cui risulta:

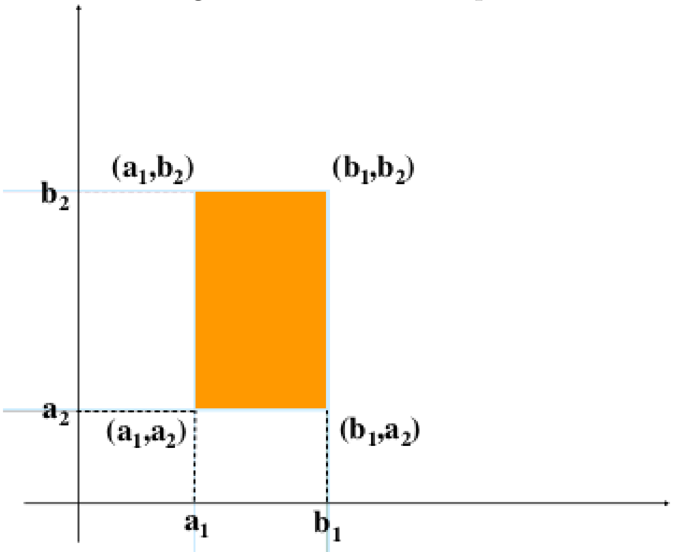
$$P((X, Y) \in (a_1, b_1] \times (a_2, b_2]) = F_{X,Y}(b_1, b_2) - F_{X,Y}(a_1, b_2) - F_{X,Y}(b_1, a_2) + F_{X,Y}(a_1, a_2)$$

$$= \int_{a_1}^{b_1} \int_{a_2}^{b_2} f_{X,Y}(u, v) du\, dv \;\; \forall (a_1, b_1] \times (a_2, b_2] \in \mathbb{R}^2$$

In molti casi, benché ci si trovi di fronte a situazioni i cui esiti sono di tipo multidimensionale, capita di essere interessati ai valori che possono

Figura 4.4: Densita di probabilità multidimensionale

img:densitaMulti



essere assunti solamente da una delle variabili per cui sono state introdotte le **funzioni marginali**.

Data una variabile aleatoria bidimensionale (X, Y) assolutamente continua, avente funzione di ripartizione congiunta $F_{X.Y}$ e funzione di densità disgiunta $f_{X,Y}$ è detta **funzione di ripartizione marginale di X** in caso $F_X(t) = P(\{X \leq t\} \cap \{Y \leq +\infty\}) = F_{X,Y}(t, +\infty)$ mentre si definisce la funzione p detta **funzione di densità marginale di X** come

$$f_X(t) = \int_{-\infty}^{+\infty} f_{X,Y}(t, s)\, ds$$

Data una variabile bidimensionale $(X; Y)$ diciamo che le due variabili considerate singolarmente sono **stocasticamente indipendenti** se e solo se $\forall(t, s) \in \mathbb{R}^2$ vale:

$$F_{X,Y}(t, s) = F_X(t) \cdot F_Y(s)$$

che discende dalla definizione di stocasticamente indipendente fornita nella teoria assiomatica della probabilità.

4.4 Indici delle variabili aleatorie

Iniziamo ad affrontare gli indici associati alle variabili aleatorie, partendo prima dagli indici centrali, grandezze numeriche associate alle variabili aleatorie in grado di sintetizzare, con un solo valore, le principali caratteristiche delle

loro distribuzioni.
Gli indici risultano strettamente legati agli indici introdotti nella prima parte in relazione alla statistica descrittiva, il più importante degli indici di tendenza centrale è detto **valore atteso** corrispondente alla media matematica dei dati statistici.

Data una variabile aleatoria unidimensionale X avente supporto $S \subseteq \mathbb{R}$ è detto valore atteso di X la quantità:

$$E[X] \begin{cases} \sum_{s \in S} \; s \cdot p_X(s) \text{ se X è discreta} \\ \int_{-\infty}^{+\infty} \; u \cdot f_X(u)\, du \text{ se X è assolutamente continua} \end{cases}$$

In effetti il valore atteso, così come la media di una serie di dati, va pensato come una *media pesata* dei valori assumibili dalla variabile, e fornisce un'indicazione di massima del posizionamento della variabile lungo l'asse dei numeri reali.
Il valore atteso gode delle seguenti tre proprietà:

1. $\forall a \in \mathbb{R}$, se $X = a$ con probabilità uguale ad 1 allora $E[X] = a$
2. $E[a \cdot X + b] = a \cdot E[X] + b$ per ogni variabile X e per ogni $a, b \in \mathbb{R}$
3. data una funzione $y = g(X)$ della variabile aleatoria X si ha che il valore atteso è:
$$E(g(X)] = \int_{-\infty}^{+\infty} g(u) f_X(u)\, du$$

Il valore atteso non è detto che esista, infatti in caso l'integrale o la sommatoria convergono il valore atteso risulta non definito e come avevamo già visto con la media, il valore atteso è in realtà un caso particolare di momento centrale di ordine $r = 1$, la cui formula è definita come:

$$E[X^r] = \begin{cases} \sum_{s \in S} \; s^r \cdot p_X(s) \text{se X è discreta} \\ \int_{-\infty}^{+\infty} \; u^r \cdot f_X(u)\, du \text{ se X è assolutamente continua} \end{cases}$$

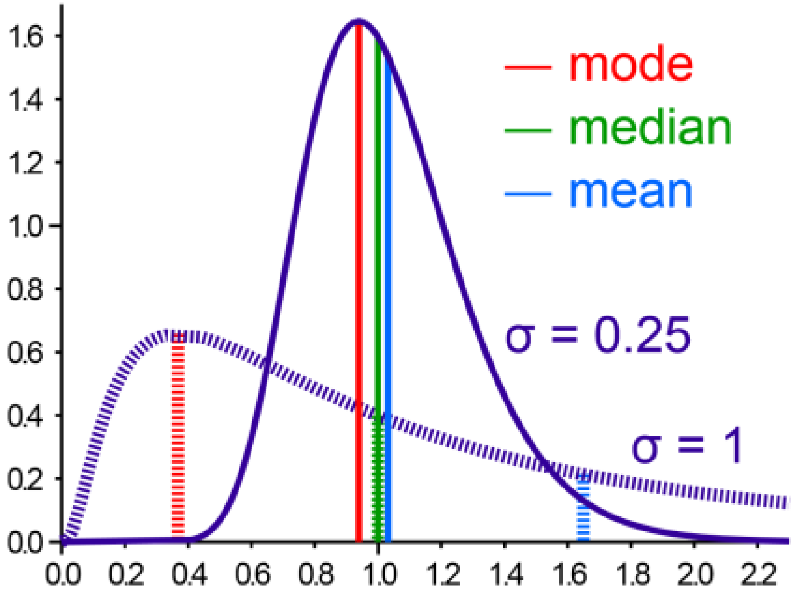
Un altro indice di tendenza centrale importante è la **moda** di una variabile aleatoria X, indicata con $\widetilde{X} \in \mathbb{R}$, corrispondente al valore per cui è massima la distribuzione discreta di probabilità, se X è discreta, oppure rappresenta la funzione di densità; tale valore non è detto che sia unico, ma può avere la presenza di più valori di moda, e ciò porta a parlare di distribuzione multimodale.

Un terzo indice di tendenza centrale è la **mediana** di una variabile aleatoria X, indicata con $\hat{X} \in \mathbb{R}$, che soddisfa la diseguaglianza:

$$lim_{t \to \hat{X}^-} F_X(t) \leq \frac{1}{2} \leq F_X(\hat{X})$$

Figura 4.5: grafico di moda, mediana e media

fig:centralValue



Nel caso in cui la funzione di ripartizione della variabile sia continua ed invertibile allora:

$$\hat{X} = F_X^{-1}(\frac{1}{2})$$

mentre nel caso di variabili discrete la mediana è il valore dell'ascissa in cui la funzione di ripartizione passa da un valore minore di $\frac{1}{2}$ ad uno superiore. La mediana può non essere unica, e ciò avviene in caso esistano più valori t per cui risulta $F_X(t) = \frac{1}{2}$

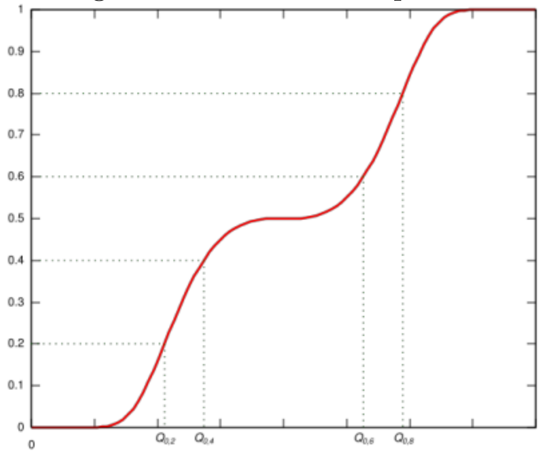
Unitamente alla mediana è possibile considerare altri indici definiti in maniera simile e che dividono la retta dei reali in due intervalli di probabilità assegnata e che sono detti **quantili**.
Dato un valore $P \in [0, 1] \subseteq \mathbb{R}$ è detto **quantile p-esimo della variabile aleatoria X** il valore:

$$x_p \in \mathbb{R} : \lim_{t \to x_p^-} F_X(t) \leq p \leq F_X(x_p)$$

Nel caso in cui la funzione di ripartizione sia continua ed invertibile allora $x_p = F_X^{-1}(p)$ e questa definizione ci porta a pensare a x_p come ad un valore in cui risulta $P(X \leq x_p) = p$ e $P(X > x_p) = 1 - p$.

Come visto in precedenza, nella statistica descrittiva, oltre agli indici di tendenza centrale è conveniente anche considerare indici che forniscano un'idea del grado di dispersione dei valori assumibili da una variabile aleatoria.

Figura 4.6: Grafico del quantile



Questi vengono detti **indici di variabilità**: il più famoso tra tutti è la **varianza** di una variabile aleatoria X definita come:

$$V[X] = \begin{cases} \sum_{s \in S} (s - E[X])^2 \cdot p_X(s) \text{ se X è discreta} \\ \int_{-\infty}^{+\infty} (u - E[X])^2 \cdot f_X(u)\, du \text{ se X è assolutamente continua} \end{cases}$$

Così come il valore atteso anche la varianza talvolta può non esistere, quando la sommatoria o l'integrale divergono e la varianza viene solitamente indicata con σ_X^2.
La radice della varianza è anch'esso un altro indice importante, detto **deviazione standard**, il cui vantaggio rispetto alla varianza è quello di avere la stessa unità di misura del valore atteso.

La varianza ha le seguenti proprietà:

1. $\forall a \in \mathbb{R}$, se $X = a$ con probabilità uguale ad 1 allora $V[X] = 0$
2. $V[a \cdot X + b] = a^2 \cdot V[X] + b$ per ogni variabile X e per ogni $a, b \in \mathbb{R}$
3. $V[X] = E[X^2] - (E[X])^2$ $\forall X$ variabile aleatoria

Anche per le variabili multidimensionali ed in particolare per quelle bidimensionali esistono indici di tendenza centrale e variabilità.
Sia (X, Y) una variabile aleatoria bidimensionale discreta o continua, sono detti **valori attesi marginali** e **varianza marginali** le quantità:

$$E[X]\ E[Y]\ V[Y]\ V[Y]$$

ottenute considerando le distribuzioni marginali di X ed Y ed integrando (o sommando) in accordo ai sistemi visti sopra.

Si hanno le seguenti relazioni:

- $E[a \cdot X + b \cdot Y] = a \cdot E[X] + b \cdot E[Y]$ $\forall$ coppia X, Y e $\forall a, b \in \mathbb{R}$
- $E[X \cdot Y] = E[X] \cdot E[Y]$ $\forall$ coppia X, Y stocasticamente indipendenti
- $V[X + Y] = V[X] + V[Y]$ $\forall$ coppia X, Y stocasticamente indipendenti

Oltre ai valori attesi ed alle varianze marginali, un altro indice estremamente importante, è la **covarianza** definita come:

$$Coc[X, Y] = E[(X - E[X]) \cdot (Y - e[X])] = \iint_{\mathbb{R}^2} (t - E[X]) \cdot (s - E[Y]) \cdot f_{X,Y}(t, s)\, dt\, ds$$

$$= \iint_{\mathbb{R}^2} t \cdot s \cdot f_{x,y}(t, s)\, dt\, ds - \int_{\mathbb{R}} t \cdot f_X(t) dt \cdot \int_{\mathbb{R}} s \cdot f_Y(s) dt$$

La covarianza è un indice della correlazione che sussiste tra due variabili ovvero del loro grado di dipendenza reciproca, in cui tanto essa è grande quanto è forte il legame di dipendenza tra le variabili.

In caso la covarianza è nulla le due variabili aleatorie vengono dette **incorrelate**, relazione meno forte della indipendenza stocastica, come avevamo già notato nel capitolo sulla statistica descrittiva.

Un ultimo indice è quello detto **coefficiente di correlazione di Pearson**, strettamente legato alla covarianza ed utilizzato per esprimere più chiaramente il grado di dipendenza tra due variabili:

$$p_{XY} = \frac{Cov[X, Y]}{\sqrt{V[X] \cdot V[Y]}} = \frac{Cov[X, Y]}{\sigma_x \cdot \sigma_y}$$

questo indice ha le seguenti proprietà:

1. $p_{XY} = 0$ se X e Y sono incorrelate
2. $|p_{XY}| = 1$ se vale la relazione $Y = a \cdot X + b$ $\forall a, b \in \mathbb{R}$

Ovvero se vale $+1$ allora $a > 0$ e se vale -1 si ha che $a < 0$.

4.5 Esercizi

Esercizio 1. *2 carte con reintroduzione si estraggono da un mazzo da 52.*

1. *P che prima sia di picche seconda di quadri*
2. *P di due figure dello stesso seme*
3. *se la prima è > 4 calcolo P che la seconda sia 10*

risposte:

1. *P1 prima carta picche e Q2 seconda carta quadri. Sono eventi indipendenti per la reintroduzione*

$$P(P1 \cap Q2) = P(P1) \cdot P(Q2) = \frac{13}{52} \cdot \frac{13}{52} = \left(\frac{13}{52}\right)^2$$

2. *A carta di seme X e B carta di seme X, si ricorda reintroduzione*

$$P(A \cap B) = P(A) \cdot P(B|A) = \frac{12}{52} \cdot \frac{3}{52} = \frac{9}{876}$$

3. *C carta > 4 e D carta = 10, la carta è sempre la stessa*

$$P(D|C) = \frac{P(D \cap C)}{P(C)} = \frac{P(D)}{P(C)} = \frac{1}{9}$$

Esercizio 2. *Gli abitanti di abc raggiungono xyz col treno con con la macchina. Il 90% arriva in ritardo, il 40% di questi usa il treno. Il 20% di quelli che non arriva in ritardo usano la macchina.*

1. *probabilità che un abitante usi il treno*

2. *se usa il treno percentuali di ritardo o meno*

risposte

1. *R = 0.9 persona in ritardo,* $\overline{R} = 0.1$ *non in ritardo, T usa treno e M usa macchina*

$$P(T) = P(R) \cdot P(T|R) + P(\overline{R}) \cdot P(T|\overline{R}) = 0.9 * 0.4 + 0.1 * 0.8 = 0.44$$

2.

$$P(R|T) = \frac{P(R \cap T)}{P(T)} = \frac{P(T|R) \cot P(R)}{P(T)} = \frac{0.9 * 0.4}{0.44} = 0.8182$$

quindi per la macchina è $1 - 0.8182 = 0.1818$

Esercizio 3. *Un urna contiene una pallina rossa e una bianca, ne estraggo una e la rimetto dentro con un'altra dello stesso colore. All'i-sima estrazione se ne estrae una rossa (Ri) e all'i-esima estrazione una bianca(Bi)*

1. *P(R2) alla seconda estrazione una rossa e P(R3) alla terza estrazione una rossa*

2. *se la seconda è rossa la prima era più probabile bianca o rossa?*

3. *P di avere prima rossa e seconda bianca*

risposta:

1.
$$P(R2) = P(R1 \cap R2) + P(B1 \cap R2) =$$
$$P(R1 \cap R2) = P(R1) \cdot P(R2|R1) = \frac{1}{2} \cdot \frac{2}{3} = \frac{1}{3}$$
$$P(B1 \cap R2) = P(B1) \cdot P(R2|B1) = \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6}$$
$$P(R2) = \frac{1}{2}$$
$$P(R3) = P(R1 \cap R2 \cap R3) + P(R1 \cap B2 \cap R3) + P(B1 \cap R2 \cap R3) + P(B1 \cap B2 \cap R3)$$
$$P(R1 \cap R2 \cap R3) = P(R1) \cdot P(R2|R1) \cdot P(R3|R1 \cap R2) = \frac{1}{2} \cdot \frac{2}{3} \cdot \frac{3}{4} = \frac{1}{4}$$
$$P(R1 \cap B2 \cap R3) = P(R1) \cdot P(B2|R1) \cdot P(R3|R1 \cap B2) = \frac{1}{2} \cdot \frac{1}{3} \cdot \frac{1}{2} = \frac{1}{12}$$
$$P(B1 \cap R2 \cap R3) = P(B1) \cdot P(R2|R1) \cdot P(R3|R1 \cap R2) = \frac{1}{2} \cdot \frac{1}{3} \cdot \frac{1}{2} = \frac{1}{12}$$
$$P(B1 \cap B2 \cap R3) = P(B1) \cdot P(B2|R1) \cdot P(R3|R1 \cap B2) = \frac{1}{2} \cdot \frac{2}{3} \cdot \frac{1}{4} = \frac{1}{12}$$
$$P(R3) = \frac{1}{4} + \frac{1}{12} + \frac{1}{12} + \frac{1}{12} = \frac{1}{2}$$

2.
$$P(R1|R2) = \frac{P(R1 \cap R2}{P(R2)} = \frac{P(R1) \cdot P(R2|R1}{P(R2)} = \frac{\frac{1}{2} \cdot \frac{2}{3}}{\frac{1}{2}} = \frac{2}{3}$$
$$P(B1|R2) = \frac{P(B1 \cap R2}{P(R2)} = \frac{P(B1) \cdot P(R2|B1}{P(R2)} = \frac{\frac{1}{2} \cdot \frac{1}{3}}{\frac{1}{2}} = \frac{1}{3}$$

3.
$$P(R1 \cap B2) = P(R1) \cdot P(B2|R1) = \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6}$$

Capitolo 5

Distribuzioni Notevoli

Essendoci una correlazione, come notato nel precedente capitolo riguardante le variabili aleatorie, tra la funzione di ripartizione e la sua distribuzione/-densità di probabilità, si parla di **distribuzione** di una variabile indentendo indifferentemente la sua ripartizione o la sua densità/distribuzione.

Adesso affrontiamo una serie di distribuzioni famose, che hanno un notevole successo ed utilizzo in campo statistico e probabilistico, incominciando dalle distribuzioni discrete e poi si analizzano quelle assolutamente continue.

Con $X \sim F$ si indica che la variabile X è distribuita secondo la distribuzione F.

5.1 Distribuzioni discrete

Le distribuzioni discrete maggiormente utilizzate, definite su un parametro $p \in [0, 1]$, sono le seguenti:

- bernoulliana
- binomiale
- poisson
- geometrica

Una variabile aleatoria X è detta distribuita secondo una Bernoulliana di parametro $p \in [0, 1]$, indicata con $X \sim B(p)$ se essa può assumere solo i valori 1 e 0 rispettivamente con probabilità p e $(1 - p)$.
Questa distribuzione presenta le seguenti funzioni di ripartizione e la sua corrispondente distribuzione di probabilità:

- $p_X(t) = \begin{cases} 1 - p \text{ se} t = 0 \\ p \text{ se} t = 1 \\ 0 \text{ altrimenti} \end{cases}$

- $F_X(t) = \begin{cases} 0 \text{ se} t < 0 \\ 1 - p \text{ se} 0 \leq t < 1 \\ 1 \text{ se} t \geq 1 \end{cases}$

Limportanza di questa semplice distribuzione è ovvia, in quanto sono variabili di Bernoulli tutte quelle che individuano il verificarsi di uno specifico evento e che valgono 1 se questo si verifica e 0 altrimenti.

Attraverso l'applicazione della definizione di valore atteso e di varianza si ottiene:

$$E[X] = 0 \cdot (1 - p) + 1 \cdot p = p$$
$$V[X] = [0^2 \cdot (1 - p) + 1^2 \cdot p] = (1 - p) \cdot p$$

Siano $X_1, \cdots, X_n\, n$ variabili Bernoulliane di identico parametro p e stocasticamente indipendenti tra loro, e sia anche $X = \sum X_i$, variabile definita distribuita secondo una binomiale di parametri n e p, espressa con $X \sim Bin(n, p)$ se tale variabile può assumere qualsiasi valore intero k compreso tra 0 e n, in accordo con la seguente probabilità:

$$P(X = k) = \binom{n}{k} \cdot p^k \cdot (1 - p)^{n-k}$$

La parte $p^k \cdot (1 - p)^{n-k}$ fornisce la probabilità che k delle n variabili X_i assumano il valore 1 e che le restanti $n - k$ variabili assumano valore 0 mentre la prima parte $\binom{n}{k}$, come ovvio dal corso di matematica discreta, fornisce il numero di combinazioni possibili delle variabili k.

In questa distribuzione vengono definite le seguenti funzioni di ripartizione e di distribuzione:

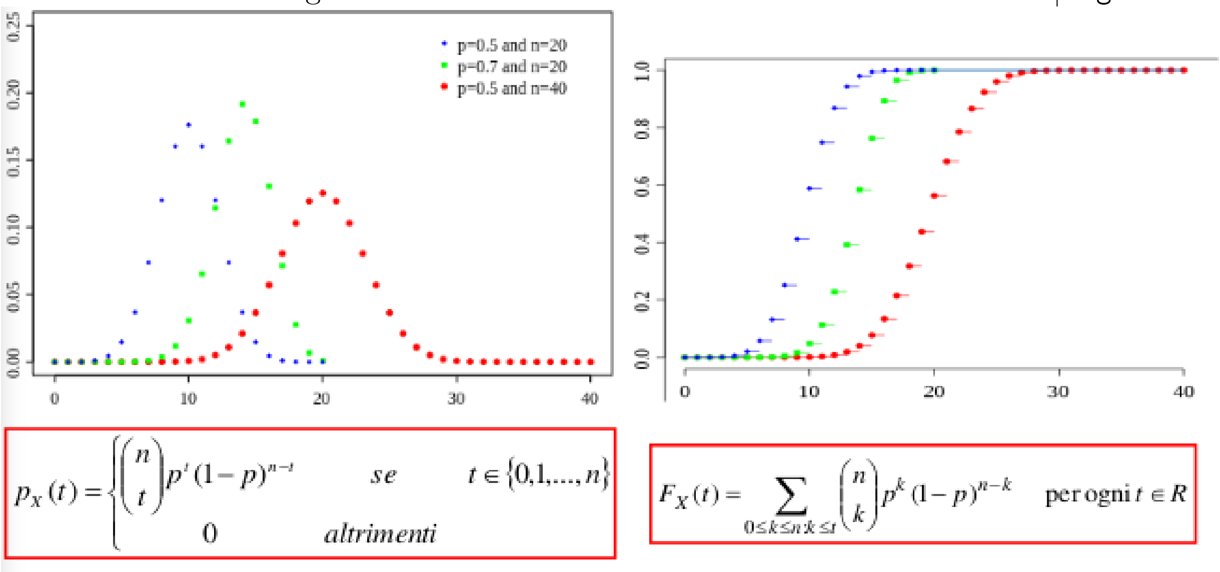
- $p_X(t) = \begin{cases} \binom{n}{t} p^t (1 - p)^{n-t} \text{ se } t \in \{0, \cdots, n\} \\ 0 \text{ altrimenti} \end{cases}$

- $F_X(t) = \sum_{0 \leq k \leq n; k \leq t} \binom{n}{k} p^k (1 - p)^{n-k}, \ \forall t \in \mathbb{R}$

Le variabili X_i sono indipendenti quindi posso calcolare il valore atteso e la varianza di X:

$$E[X] = E[X_1 + \cdots + X_n] = E[X_1] + \cdots + E[X_n] = n \cdot p$$

Figura 5.1: Distribuzione binomiale `fig:binomiale`



$$V[X] = V[X_1 + \cdots + X_n] = V[X_1] + \cdots + V[X_n] = n \cdot (1-p) \cdot p$$

La principale applicazione della distribuzione binomiale consiste nella definizione di variabili che "contano" le realizzazioni di eventi quando questi siano da considerarsi indipendenti e con identica probabilità di verificarsi.

La distribuzione di Poisson può essere vista come un caso particolare della distribuzione Binomiale che si ottiene quando il numero di variabili X_i che compaiono in $X = \sum X_i$ tende ad infinito mentre il valore del parametro p tende a zero in modo tale che il prodotto $\lambda = n \cdot p$ resti costante.

In caso ciò viene rispettato definiamo X distribuita secondo una Poisson con parametro $\lambda \in \mathbb{R}_+$, indicata con $X \sim Poi(\lambda)$, applicabile sui valori in $\mathbb{R}$. La probabilità associata a questa distribuzione è la seguente:

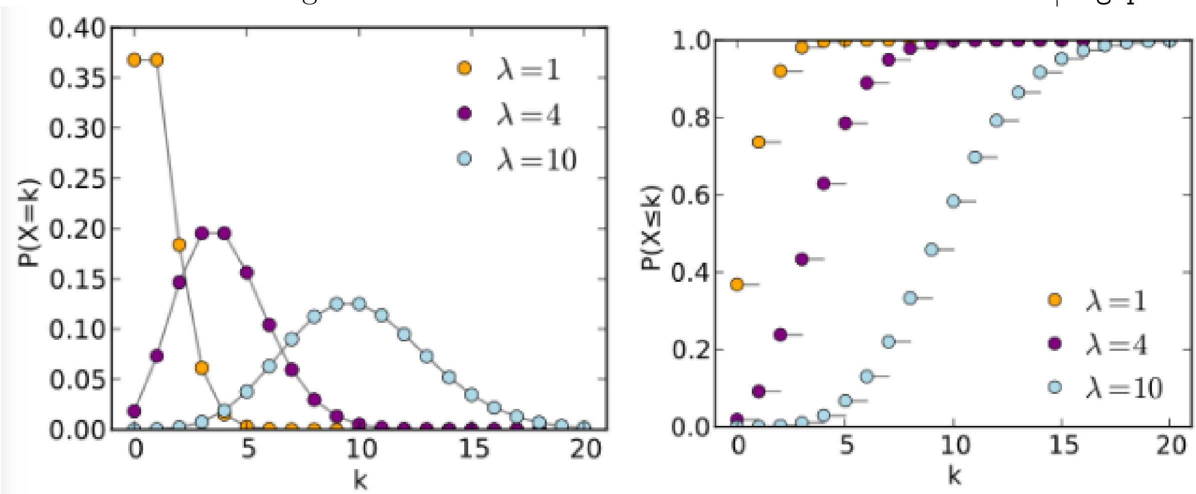
$$P(X = k) = \lim_{n \to +\infty} \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} = \frac{\lambda^k}{k!} e^{-lambda} \; \forall k \in \mathbb{N}$$

la *distribuzione di probabilità* e la *funzione di ripartizione* risultano essere quindi:

$$p_X(t) = \begin{cases} \frac{\lambda^k}{k!} e^{-\lambda} \text{ se } t \in \{0, \cdots, n\} \\ 0 \text{ altrimenti} \end{cases}$$

$$F_X(t) = \sum_{k \in \mathbb{N}: k \le t} \frac{\lambda^k}{k!} e^{-\lambda}, \; \forall t \in \mathbb{R}$$

Figura 5.2: Distribuzione di Poisson `fig:poisson`



$$p_X(t) = \begin{cases} \frac{\lambda^t}{t!} \cdot e^{-\lambda} & se \quad t \in \{0, 1, \dots, \} \\ 0 & altrimenti \end{cases}$$

$$F_X(t) = \sum_{k \in N : k \leq t} \frac{\lambda^k}{k!} \cdot e^{-\lambda} \quad \text{per ogni } t \in R$$

valore atteso e varianza le ricavo da quelli delle bernoulliane con $\lambda = n \cdot p$:

$$E[X] = n \cdot p = \lambda$$

$$V[X] = n \cdot (1-p) \cdot p = n \cdot p - n \cdot p^2 = \lambda - \frac{\lambda^2}{n}$$

che con $n \to +\infty$ diventa $V[X] = \lambda$.

La distribuzione di Poisson viene utilizzata quando si considerino grandi popolazioni di individui in cui ogni individuo ha una probabilità p molto piccola di essere soggetto ad uno specifico evento in esame; per tale ragione la distribuzione di Poisson viene anche detta degli eventi rari.

Una variabile aleatoria X è detta distribuita secondo una Geometrica di parametro $p \in [0,1]$, indicata con $X \sim Geo(p)$, se può assumere qualsiasi valore intero non negativo k con probabilità $P(X = k) = p \cdot (1-p)^k$.

La distribuzione di probabilità e la funzione di ripartizione risultano essere quindi:

$$p_X(t) = \begin{cases} p \cdot (1-p)^t \text{ se } t \in \mathbb{N} \\ 0 \text{ altrimenti} \end{cases}$$

$$F_X(t) = \sum_{k \in \mathbb{N} : k \leq t} p \cdot (1-p)^k, \ \forall t \in \mathbb{R}$$

con valore atteso e varianza:

$$E[X] = \frac{1-p}{p}$$

$$V[X] = \frac{1-p}{p^2}$$

Questa distribuzione ha la proprietà di **assenza di memoria**, ossia risulta $P(X = k + m|X \geq m) = P(X = k)$.

Per comprenderne il significato, supponiamo che X sia il tempo di vita di una macchina soggetta a guasti , che possono avvenire solo in corrispondenza di intervalli di tempo unitari, e supponiamo di aver rilevato che per m unità di tempo essa non si sia guastata.
La proprietà di assenza di memoria A asserisce che la probabilità che la macchina si guasti all'istante $(k+m)$-esimo, condizionata dall'evento $X \geq m$, è uguale alla probabilità iniziale che essa si guasti all'istante k-esimo.
Quindi questa proprietà asserisce che il tempo trascorso da quando abbiamo iniziato ad esaminare il funzionamento della macchina non influisce sulla distribuzione del tempo restante al verificarsi del guasto.

5.2 Distribuzioni continue

Parliamo ora della **distribuzione uniforme**. Rappresenta la più semplice distribuzione assolutamente continua e viene adottata nel caso in cui la variabile considerata possa assumere qualsiasi valore compreso in un dato intervallo con probabilità costante.
Si dice che la variabile X è distribuita secondo una uniforme di supporto $[a, b]$, indicata con $X \sim U[a, b]$ se essa è assolutamente continua con densità e funzione di ripartizione:

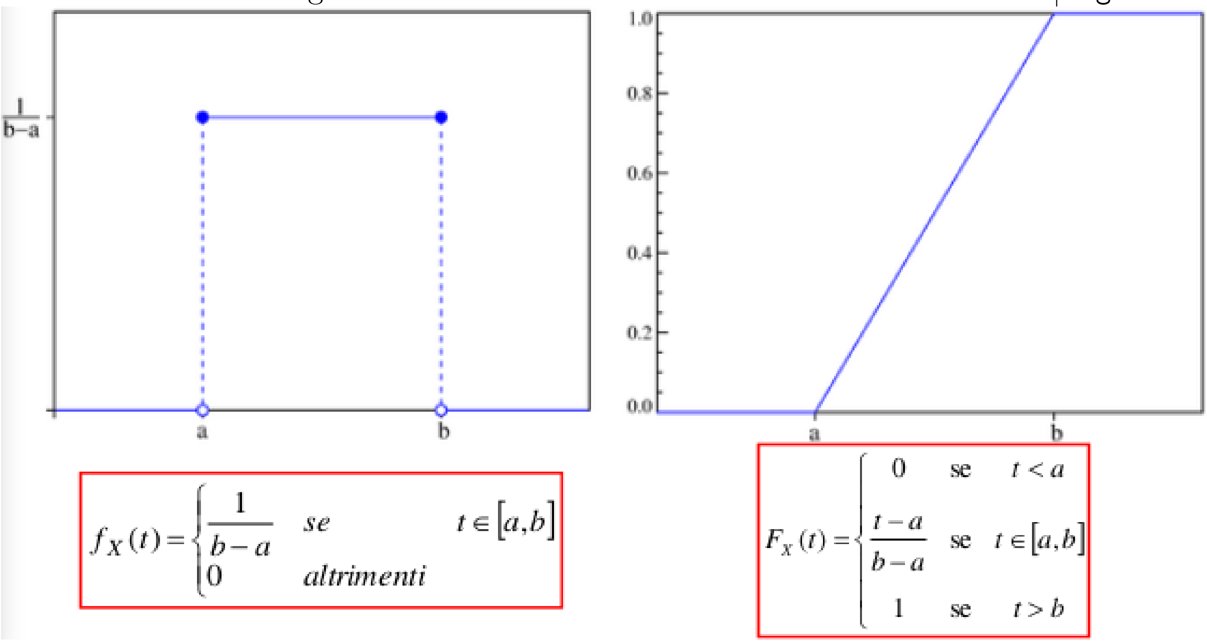
$$f_X(t) = \begin{cases} \frac{1}{b-a} \text{ se } t \in [a, b] \\ 0 \text{ altrimenti} \end{cases}$$

$$F_X(t) = \begin{cases} 0 \text{ se } t < a \\ \frac{t-a}{b-a} \text{ se } t \in [a, b] \\ 1 \text{ se } t > b \end{cases}$$

e con semplici integrazioni è possibile ricavare il valore atteso e la varianza:

$$E[X] = \frac{a+b}{2}$$

Figura 5.3: Distribuzione Uniforme img:uniforme



$$V[X] = \frac{(b-a)^2}{12}$$

Come accennato in precedenza l'interesse in questa distribuzione è giustificato dal fatto che essa descrive bene situazioni nelle quali le variabili possono assumere valori in intervalli finiti di $\mathbb{R}$ con probabilità uniforme, ovvero tale da essere identica per intervalli di medesima ampiezza, purchè contenuti nel supporto della variabile stessa.

Ovviamente non è certo che i valori assumibili abbiano tutti la stessa probabilità di presentarsi, per questo sono stati introdotti alcune generalizzazioni della distribuzione uniforme;una di queste è la **distribuzione triangolare**, che assegna alla densità di probabilità valori maggiori al centro del supporto e minori in prossimità degli estremi.
Formalmente diciamo che la variabile X è distribuita secondo una Triangolare di supporto $[a, b]$, indicata con $X \sim T[a, b]$ se essa è assolutamente continua con le seguenti funzioni di densità e ripartizione:

$$f_X(t) = \begin{cases} \frac{4(t-a)}{(b-a)^2} \text{ se } t \in \left[a, \frac{a+b}{2}\right) \\ \frac{4(b-t)}{(b-a)^2} \text{ se } t \in \left[\frac{a+b}{2}, b\right] \\ 0 \text{ altrimenti} \end{cases}$$

$$F_X(t) = \begin{cases} 0 \text{ se } t < a \\ \frac{2(t-a)^2}{(b-a)^2} \text{ se } t \in \left[a, \frac{a+b}{2}\right) \\ 1 - 2\frac{(b-t)^2}{(b-a)^2} \text{ se } t \in \left[\frac{a+b}{2}, b\right] \\ 1 \text{ se } t > b \end{cases}$$

Attraverso l'integrazione delle funzioni presentate si ottiene il valore atteso e la varianza:

$$E[X] = fraca + b2$$

$$V[X] = \frac{(b-a)^2}{24}$$

Passiamo alla **distribuzione esponenziale**, importante nello studio di variabili che descrivono i tempi necessari per il verificarsi di un evento. Formalmente, una variabile aleatoria X è distribuita secondo una Esponenziale di parametro $\lambda \in \mathbb{R}$, indicata con $X \sim Exp(\lambda)$ se essa è assolutamente continua con le seguenti funzioni:

$$f_X(t) = \begin{cases} \lambda e^{-\lambda t} \text{ se } t \geq 0 \\ 0 \text{ altrimenti} \end{cases}$$

$$F_X(t) = \begin{cases} 1 - e^{-\lambda t} \text{ se } t \geq 0 \\ 0 \text{ altrimenti} \end{cases}$$

Attraverso integrazione si ottengono **valor medio** e **varianza**:

$$E[X] = \frac{1}{\lambda}$$

$$V[X] = \frac{1}{\lambda^2}$$

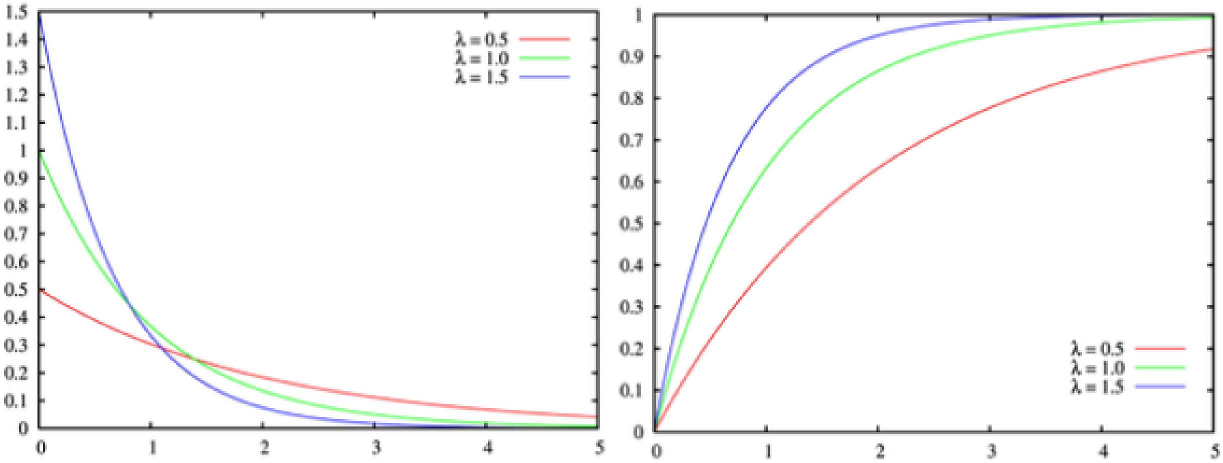
L'importanza della distribuzione esponenziale in numerosi campi applicativi è dovuta al fatto che essa è lunica distribuzione assolutamente continua che gode della **proprietà di assenza di memoria**, vista anche nella distribuzione geometrica, di tipo discreto.

5.2.1 Distribuzione Normale

Una variabile aleatoria X è detta **distribuita secondo una Normale con parametri** $\mu \in \mathbb{R}$ e $\sigma \in \mathbb{R}_+$, indicata attraverso $X \sim N(\mu, \sigma)$, se essa è assolutamente continua con densità:

$$f_X(t) = \frac{1}{\sqrt{2 \cdot \pi \cdot \sigma^2}} \cdot e^{-\frac{(t-\mu)^2}{2 \cdot \sigma^2}} \quad \forall t \in \mathbb{R}$$

Figura 5.4: Distribuzione esponenziale



$$f_X(t) = \begin{cases} \lambda \cdot e^{-\lambda t} & se \quad t \geq 0 \\ 0 & se \quad t < 0 \end{cases}$$

$$F_X(t) = \begin{cases} 1 - e^{-\lambda t} & se \quad t \geq 0 \\ 0 & se \quad t < 0 \end{cases}$$

Questa distribuzione è fondamentale per il **teorema centrale del limite**, studiata nel prossimo capitolo e poi perchè molti fenomeni studiati dalla statistica e dalla probabilità si comportano come una normale, chiamata anche **gaussiana**.

Integrando si ottengono il valore atteso e la varianza:

$$E[X] = \mu$$

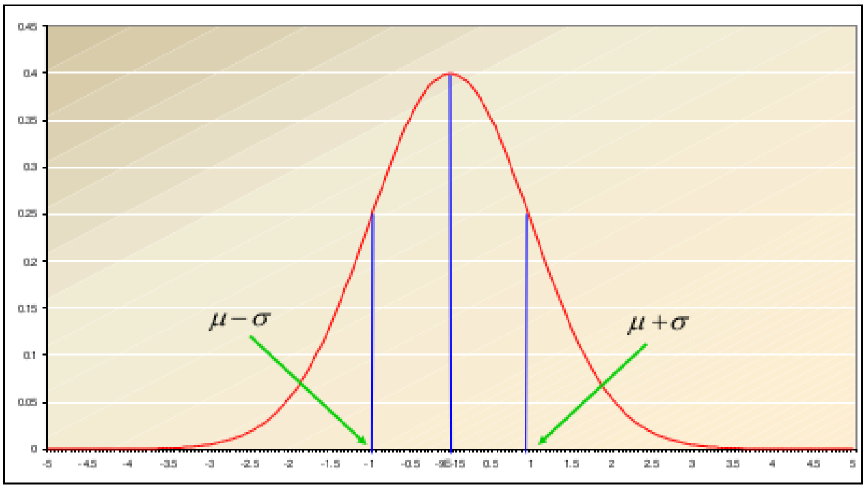
$$V[X] = \sigma^2$$

Inoltre, la variabile aleatoria ha supporto su tutto l'asse reale: Calcolare la probabilità P, assunta da una variabile X in un intervallo $[a, b]$ è notevolmente difficile in quanto si deve risolvere la seguente equazione:

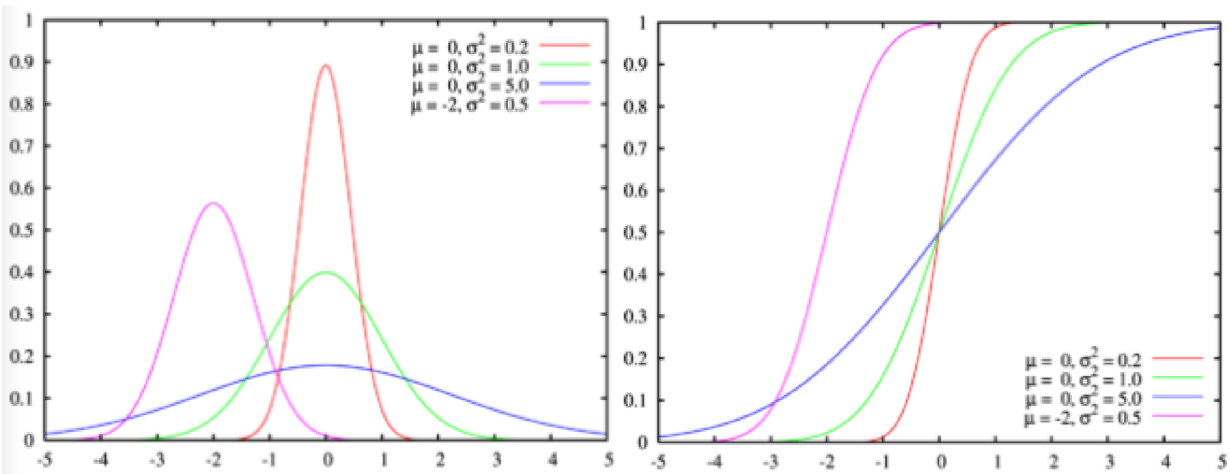
$$P(X \in [a, b]) = \int_a^b \frac{1}{\sqrt{2 \cdot \pi \cdot \sigma^2}} \cdot e^{-\frac{(t-\mu)^2}{2 \cdot \sigma^2}} \, dt$$

Per questa ragione si ricorre ad opportune tavole che si riferiscono alla distribuzione normale standard ovvero con parametri $\mu = 0$ e $\sigma = 1$ e che forniscono i valori di $\int_0^z f_X(t(\, dt$ per un elevato numero di valori $z \in \mathbb{R}$.
Quando si sia interessati a determinare delle probabilità associate ad una generica normale ci si riconduce al caso sopra osservando che la variabile $Z = \frac{X-\mu}{\sigma}$ è distribuita se è distribuita secondo una **Normale Standard**

Graficamente la densità di una Normale risulta come quella presentata sotto



Si osservi che la moda coincide con la media e che in corrispondenza dei valori $\mu-\sigma$ $\mu+\sigma$ vi sono dei punti di flesso.



ovvero si ha che:

$$\text{se } X \sim N(\mu, \sigma) \text{ allora } Z = \frac{X\mu}{\sigma} \sim N(0, 1)$$

quindi ogni variabile $X \sim N(\mu, \sigma)$ può essere ricondotta ad una può essere ricondotta ad una Normale standarizzata ovvero ancora per ogni ovvero ancora $\forall [a, b] \subseteq \mathbb{R}$ si avrà:

$$P(X \in [a, b]) = P(a \leq X \leq b) = P\left(\frac{a-\mu}{\sigma} \leq \frac{X-\mu}{\sigma} \leq \frac{b-\mu}{\sigma}\right) = P\left(Z \in \left[\frac{a-\mu}{\sigma}, \frac{a-\mu}{\sigma}\right]\right)$$

La distribuzione gaussiana possiede la seguente proprietà:

Definizione 3. *se $X_1 \sim N(\mu_1, \sigma_1)$ e $X_2 \sim N(\mu_2, \sigma_2)$ e se X_1 e X_2 sono indipendenti allora la variabile $Y = X_1 + X_2$ tale che:*

$$Y \sim N(\mu_1 + \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2})$$

In altri termini la variabile somma di due variabili aleatorie stocasticamente indipendenti con distribuzioni normali è ancora una variabile aleatoria distribuita secondo una normale i cui parametri sono ricavabili facilmente da quelli delle distribuzioni degli addendi.

La tabella della z

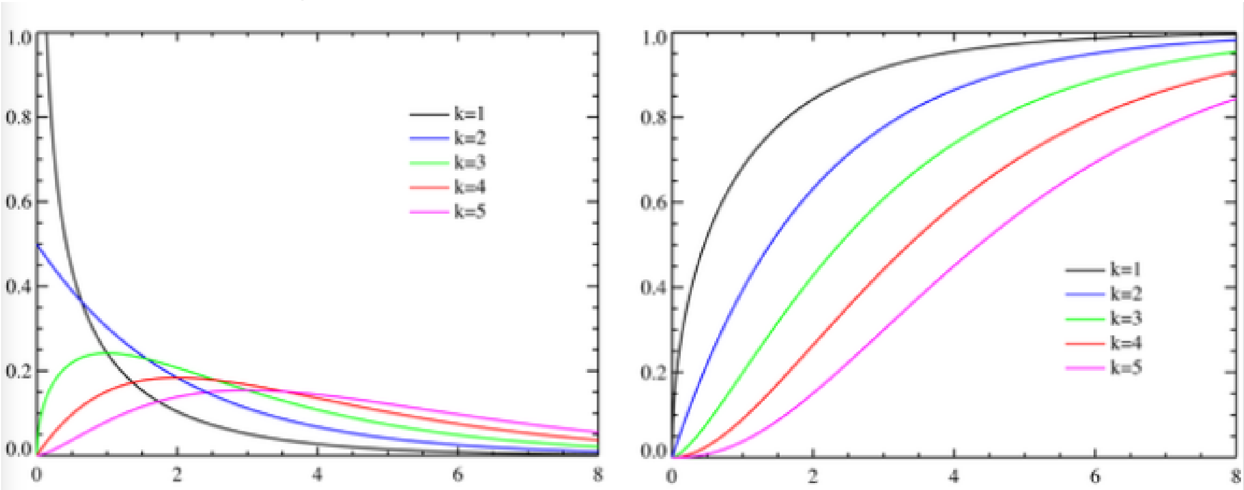
una volta ottenuto $P\left(Z \in \left[\frac{a-\mu}{\sigma}, \frac{a-\mu}{\sigma}\right]\right)$ poniamo per semplicità $\gamma = \frac{a-\mu}{\sigma}$ e $\delta = \frac{a-\mu}{\sigma}$. Per simmetria notiamo che è indifferente valutare γ e δ sia che siano positivi che negativi e per comodità la tabella presenta unicamente valori positivi. Valutemo quindi $|\gamma|$ e $|\delta|$. Sappiamo che $P(Z \in [\gamma, \delta]) = P(Z \in [0, |\gamma|]) + P(Z \in [0, |\delta|])$. Per calcolare, per esempio, $P(Z \in [0, |\gamma|])$ prendo la cifra intera e la prima cifra decimale e trovo la riga corrispondente nella prima colonna (se $\gamma = 1.35$ cercherò 1.3) e poi cerco scelgo la colonna corrispondente al valore della seconda cifra decimale presente nella prima riga (nel caso di prima cerco nella prima riga il valore 0.05 e ne scelgo la colonna). L'incrocio fra la riga scelta prima e la colonna scelta dopo mi daranno il valore ricercato. Ecco un esempio con 0.08 e 1.35:

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990

Regole di calcolo per normali standardizzati da tabelle per integrali:

- integrali della forma $\int_{-\infty}^{b} f(u)\,du$:
  - $b > 0$, finito:
$$\int_{-\infty}^{b} f(u)\,du = \frac{1}{2} + \int_{0}^{b} f(u)\,du$$
  - $b < 0$, finito:
$$\int_{-\infty}^{b} f(u)\,du = \frac{1}{2} - \int_{0}^{-b} f(u)\,du$$
- integrali della forma $\int_{a}^{+\infty} f(u)\,du$:
$$\int_{a}^{+\infty} f(u)\,du = 1 - \int_{-\infty}^{a} f(u)\,du$$
- integrali della forma $\int_{a}^{b} f(u)\,du$:
$$\int_{a}^{b} f(u)\,du = \int_{-\infty}^{b} f(u)\,du - \int_{-\infty}^{a} f(u)\,du$$

Figura 5.5: Distrubuzione chi-quadro



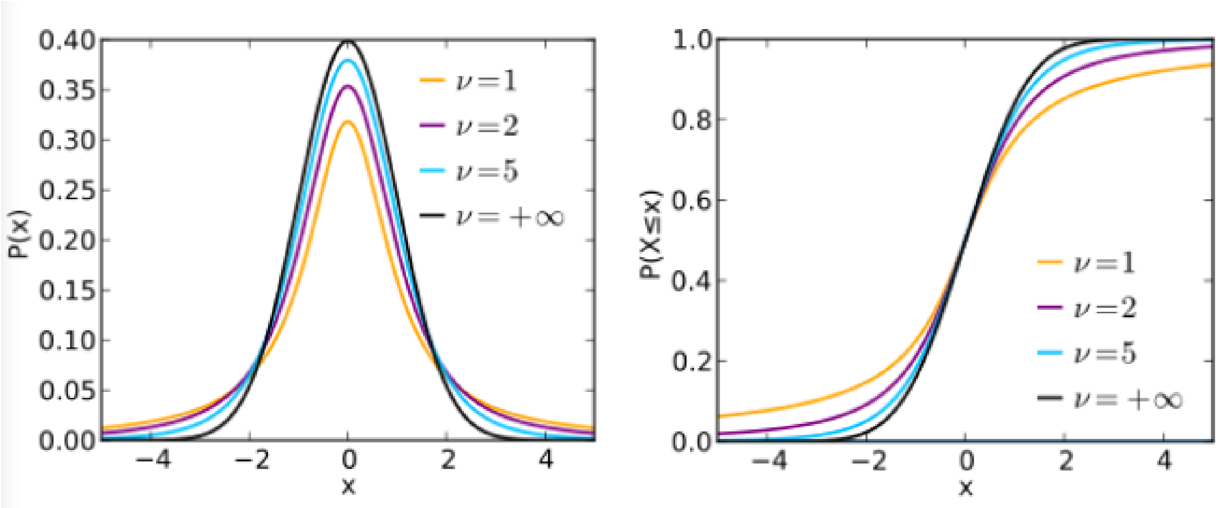
Dopo aver considerato le distribuzioni principali continue, consideriamo 3 distribuzioni utili per la statistica inferenziale:

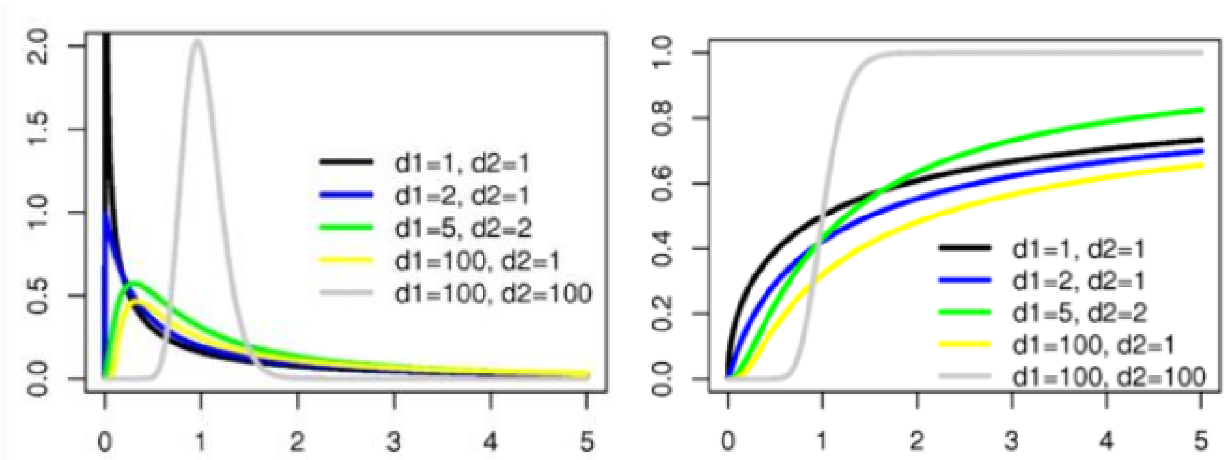
- **chi-quadro**: siano $X_1, \cdots, X_n n$ variabili con distribuzione normale standard ed indipendenti tra loro e sia X una variabile aleatoria definita come $X = \sum X_i^2$, definita distribuita secondo una chi-quadro con n gradi di liberta, indicata con $X \sim \chi_n^2$, che ovviamente essendo definita come somma di quadrati può assumere solo valori non negativi.

- **t di student**: siano $Z \sim N(0,1)$ e $Y \sim \chi_n^2$ due variabili indipendenti e sia X una variabile aleatoria definita come $X = \frac{Z}{\sqrt{\frac{Y}{n}}}$, definita distribuita secondo una t di student con n gradi di libertà, indicata con $X \sim t_n$.

- **distrubuzione f**: siano $U \sim \chi_m^2$ e $V \sim \chi_n^2$ due variabili indipendenti, si ha che X una variabile aleatoria definita come $X = \frac{\frac{U}{m}}{\frac{V}{n}}$, definita distribuita secondo una F con m e n gradi di libertà, indicata con $X \sim F(m,n)$.

Figura 5.6: Distribuzione t di student





Capitolo 6

Teoremi di Convergenza

Consideriamo una successione $\{X_n, n \in \mathbb{N}\}$ di variabili aleatorie e sia F_n la funzione di ripartizione della generica variabile X_n della successione.
Diremo che la successione converge in distribuzione alla variabile X avente funzione di ripartizione F se vale:

$$\lim_{n \to \infty} F_n(t) = F(t)$$

per ogni $t \in \mathbb{R}$, punto di continuità per la funzione di ripartizione F.

Si useranno le seguenti notazioni:

$$X_n \xrightarrow{d} X$$

$$F_n \xrightarrow{d} F$$

Nella definizione di convergenza in distribuzione vengono esclusi, nel passaggio al limite, i punti in cui la funzione di ripartizione limite F è discontinua, per avere un concetto di convergenza simile all'intuizione.

Consideriamo ad esempio una successione di numeri reali $\{a_n, n \in \mathbb{N}\}$ tale che:

$$a_n \xrightarrow{d} a \in \mathbb{R} \text{ per } n \to \infty$$

e pensiamo alle variabili aleatorie X_n aventi funzione di ripartizione:

$$F_n(t) = \begin{cases} 0 \text{ se } t < a_n \\ 1 \text{ se } t \geq a_n \end{cases}$$

ovvero la generica variabile X_n assume valore a_n con probabilità 1.
Da un punto di vista intuitivo siamo portati a pensare che valga $X_n \xrightarrow{d} X$, con X avente la funzione di ripartizione:

$$F(t) = \begin{cases} 0 \text{ se } t < a \\ 1 \text{ se } t \geq a \end{cases}$$

ovvero $X = a$ con probabilità uguale a uno ma la successione $\{a_n, n \in N\}$ tale che:

$$\begin{cases} a < a_n \text{ con n pari} \\ a_n < a \text{ con n dispari} \end{cases}$$

risulta quindi:

$$\begin{cases} F_n(a) = 0 \text{ per n pari} \\ F_n(a) = 1 \text{ per n dispari} \end{cases}$$

quindi il limite $\lim_{n\to\infty} F_n(a)$ non esiste quindi è scorretto dire che la successione converge in distribuzione in quanto non è definibile in a la funzione ripartizione limite F per cui per evitare questi problemi si esclude, nella definizione di convergenza, i valori in cui la F limite non è continua.

Segnaliamo che la convergenza in distribuzione non è lunico tipo di convergenza tra variabili aleatorie definito in letteratura, come ad esempio sono importanti convergenza quasi certa e in probabilità.
In questo corso di statistica e probabilità abbiamo deciso di non considerarla in quanto esula del nostro corso.

6.1 Legge dei Grandi Numeri

Considero la successione $\{X_i \in \mathbb{N}_+\}$ di variabili aleatorie indipendenti ed identicamente distribuite e considero poi la variabile aleatoria $\overline{X}_n = \frac{X_1+\cdots+X_n}{n}$, detta **media aritmetica n-esima della successione**.

Avendo $E[X_i] = \mu$ e $V[X_i] = \sigma^2$, si ha allora $\overline{X}_n \xrightarrow{d} M$, con M variabile aleatoria che assume il valore μ con probabilità 1.
La proprietà appena introdotta costituisce una forma debole del risultato, noto con il nome di **Legge dei Grandi Numeri**.

Nel dettaglio questa legge asserisce che:

Teorema 1. *Se considero una successioni di variabili aleatorie indipendenti ed identicamente distribuite $\{X_i, i \in \mathbb{N}_+\}$ per cui esistono finiti valore atteso e varianza allora si può affermare che la successione:*

$$\{\overline{X}_n, n \in \mathbb{N}_+$$

delle corrispondenti medie aritmetiche tende, al crescere di n, ad una variabile che assume certamente il valore:

$$E[X_i] = \mu$$

Quindi se considero una successione $\{x_i, i \in \mathbb{N}_+\}$ di realizzazioni delle variabili $\{X_i, i \in \mathbb{N}_+\}$ e se considero la successione $\{\overline{x}, n \in \mathbb{N}_+\}$ delle corrispondenti realizzazioni delle medie aritmetiche, abbiamo che questa seconda successione tende, per n tendente ad infinito, al valore $E[X_i] = \mu$.

Nella realtà le ipotesi della legge dei grandi numeri possono essere indebolite rispetto a quelle definite da noi, infatti esistono versioni alternative in cui non è richiesta l'ipotesi che le variabili X_i siano identificamente distribuite.

La legge dei grandi numeri assicura la convergenza $\overline{X}_n \xrightarrow{d} M$ ma non ci fornisce alcuna informazione riguardo la rapidità con cui ciò avviene, infatti non sappiamo per quale n è lecito supporre che la realizzazione $\overline{x}_n$ assuma valore μ o prossimo ad esso.
È intuitivo pensare che la convergenza avvenga con maggiore rapidità in caso di una varianza molto piccola, e questo è il *teorema del limite centrale*, in cui viene specificata quale sia la distribuzione della variabile aleatoria $\overline{X}_n$, con n sufficientemente grande, e quali siano il valore atteso e la varianza della stessa.

Sia $\{X_i, \in \mathbb{N}_+\}$ una successione di variabili aleatorie che soddisfa le ipotesi della Legge dei Grandi Numeri, ovvero siano le X_i indipendenti ed identicamente Distribuite ed aventi valore atteso:

$$E[X_i] = \mu$$

e varianza

$$V[X_i] = \sigma^2$$

entrambi esistenti e finiti.
Si consideri ora la variabile aleatoria S_n, $S_n = \sum X_i$, in cui vale:

$$S_n \xrightarrow{d} X \sim N(n \cdot \mu, \sqrt{n} \cdot \sigma) = N(n \cdot \mu, n \cdot \sigma^2)$$

. Ovvero S_n converge in distribuzione ad una variabile distribuita come una normale di media $n \cdot \mu$ e deviazione standard $\sqrt{n} \cdot \sigma$.

Osservato che vale $\overline{X}_n = \frac{S_n}{n}$ dalla formula:

$$S_n \xrightarrow{d} X \sim N(n \cdot \mu, \sqrt{n} \cdot \sigma) = N(n \cdot \mu, n \cdot \sigma^2)$$

si ha che $E[a \cdot X + b] = a \cdot E[X] + b, \ \forall X, \ \forall a, b \in \mathbb{R}$ e quindi:

$$E[\overline{X}_n] = E\left[\frac{S_n}{n}\right] = \frac{1}{n}E[S_n] = \frac{1}{n}E[X_1 + \cdots + X_n] = \frac{1}{n}(n\mu) = \mu$$

inoltre si ha che $V[a \cdot X + b] = a^2 \cdot V[x], \ \forall X \ \forall a, b \in \mathbb{R}$ e quindi

$$V[\overline{X}_n] = V\left[\frac{S_n}{n}\right] = \frac{1}{n^2}V[S_n] = \frac{1}{n^2}V[X_1 + \cdots + X_n] = \frac{1}{n^2}(n\sigma^2) = \frac{\sigma^2}{n}$$

quindi si conclude che:

$$S_n \xrightarrow{d} X \sim N(\mu, \frac{\sigma}{\sqrt{n}}) = N(\mu, \frac{\sigma^2}{n})$$

quindi per n sufficientemente grande possiamo approssimare le variabili S_n e $\overline{X}_n$ con delle variabili aventi distribuzione normale i cui parametri dipendono da quelli delle variabili X_i. Si ha inoltre la seguente approssimazione:
n è ritenuto sufficientemente grande sse $n \geq 30$

$$p_X(t) = \begin{cases} P(X = s) \forall s \in S \\ 0 \text{altrimenti} \end{cases}$$