

GeMan

Gestione della Manutenzione

Statistica e Affidabilità dei Sistemi



SAPIENZA
UNIVERSITÀ DI ROMA

Lorenzo Fedele
Dipartimento di Ingegneria Meccanica e Aerospaziale

Dalla Statistica all'Affidabilità

La Fidatezza

FIDATEZZA (*Dependability*)

- Affidabilità (*Reliability*)
 - Disponibilità (*Availability*)
 - Manutenibilità (*Maintainability*)
 - Sicurezza (*Safety*)
- **RAMS**

Dalla Statistica all'Affidabilità

La Fidatezza

RAMS

Acronimo per le grandezze su cui si fonda la teoria dell'Affidabilità che attraverso lo studio delle leggi di occorrenza dei guasti cerca di approfondire la conoscenza dei sistemi relativa ai problemi di:

1. Sopravvivenza
2. Durata media di vita
3. Percentuale di tempo di buon funzionamento.

Dalla Statistica all'Affidabilità

La Fidatezza

L'**Affidabilità** si propone di indagare le caratteristiche di fidatezza di un sistema di componenti attraverso:

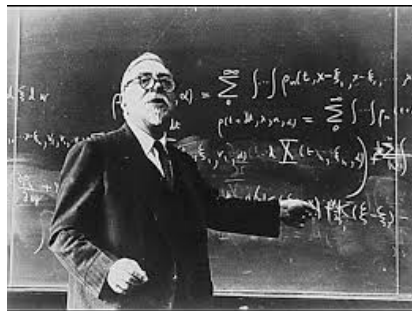
- teorie e metodi matematici e statistici
- metodi organizzativi
- pratiche operative
- tecniche e metodi di analisi
- metodi di analisi numerica e computazionale (algoritmi)
- modellazione e strumenti simulazione

Dalla Statistica all'Affidabilità

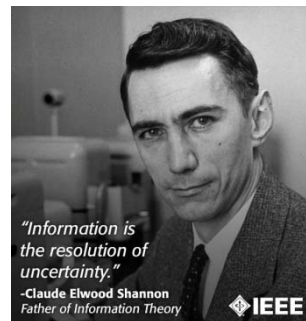
Ingegneria dei Sistemi

I sistemi

Un **sistema** è un insieme di **elementi** o **componenti** caratterizzati da particolari condizioni funzionali ed affidabilistiche, ciascuno dei quali contribuisce a realizzare il funzionamento del sistema complessivo con un certo **livello affidabilistico**.



Warren Weaver
(Reedsburg, 17 luglio 1894 – New Milford, 24 novembre 1978)



Claude Shannon
(Petoskey, 30 aprile 1916 – Medford, 24 febbraio 2001)



*Herbert A. Simon
(1916-2001)*

Herbert Simon
(Milwaukee, 15 giugno 1916 – Pittsburgh, 9 febbraio 2001)

Dalla Statistica all'Affidabilità

Ingegneria dei Sistemi

SCIENCE AND COMPLEXITY

By WARREN WEAVER
Rockefeller Foundation, New York City

SCIENCE has led to a multitude of results that affect men's lives. Some of these results are embodied in mere conveniences of a relatively trivial sort. Many of them, based on science and developed through technology, are essential to the machinery of modern life. Many other results, especially those associated with the biological and medical sciences, are of unquestioned benefit and comfort. Certain aspects of science have profoundly influenced men's ideas and even their ideals. Still other aspects of science are thoroughly awesome.

How can we get a view of the function that science should have in the developing future of man? How can we appreciate what science really is and, equally important, what science is not? It is, of course, possible to discuss the nature of science in general philosophical terms. For some purposes such a discussion is important and necessary, but for the present a more direct approach is desirable. Let us, as a very realistic politician used to say, let us look at the record. Neglecting the older history of science, we shall go back only three and a half centuries and take a broad view that tries to see the main features, and omits minor details. Let us begin with the physical sciences, rather than the biological, for the place of the life sciences in the descriptive scheme will gradually become evident.

A Mathematical Theory of Communication

By C. E. SHANNON

INTRODUCTION

THE recent development of various methods of modulation such as PCM and PPM which exchange bandwidth for signal-to-noise ratio has intensified the interest in a general theory of communication. A basis for such a theory is contained in the important papers of Nyquist¹ and Hartley² on this subject. In the present paper we will extend the theory to include a number of new factors, in particular the effect of noise in the channel, and the savings possible due to the statistical structure of the original message and due to the nature of the final destination of the information.

The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. Frequently the messages have *meaning*; that is they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem. The significant aspect is that the actual message is *one selected from a set of possible messages*. The system must be designed to operate for each possible selection, not just the one which will actually be chosen since this is unknown at the time of design.

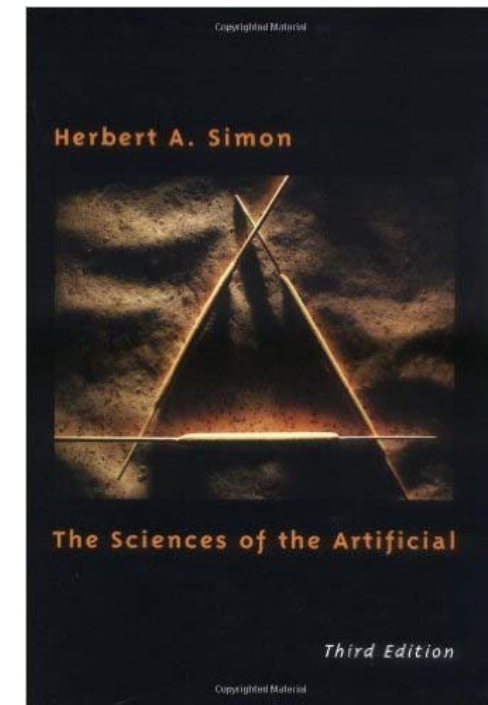
If the number of messages in the set is finite then this number or any monotonic function of this number can be regarded as a measure of the information produced when one message is chosen from the set, all choices being equally likely. As was pointed out by Hartley the most natural choice is the logarithmic function. Although this definition must be generalized considerably when we consider the influence of the statistics of the message and when we have a continuous range of messages, we will in all cases use an essentially logarithmic measure.

The logarithmic measure is more convenient for various reasons:

1. It is practically more useful. Parameters of engineering importance

¹ Nyquist, H., "Certain Factors Affecting Telegraph Speed," *Bell System Technical Journal*, April 1924, p. 324; "Certain Topics in Telegraph Transmission Theory," *A. I. E. E. Trans.*, v. 47, April 1928, p. 617.

² Hartley, R. V. L., "Transmission of Information," *Bell System Technical Journal*, July 1928, p. 535.



Dalla Statistica all’Affidabilità

Ingegneria dei Sistemi

✓ I Sistemi e la Complessità

*L’evoluzione dell’**ingegneria di manutenzione** è strettamente legata allo sviluppo dell’**ingegneria dei sistemi**.*

Il termine **stato di un sistema** in un certo istante denota l’insieme delle proprietà rilevanti di tutti gli elementi del sistema in tale istante.

La complessità di un sistema cresce, in generale, con la **numerosità degli elementi** componenti, con la **molteplicità delle relazioni** sussistenti e con la **molteplicità degli stati** che il sistema può assumere.

Dalla Statistica all'Affidabilità

Ingegneria dei Sistemi

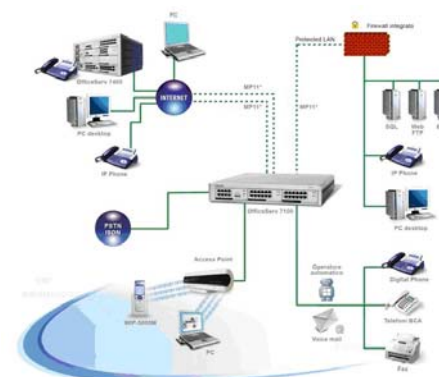
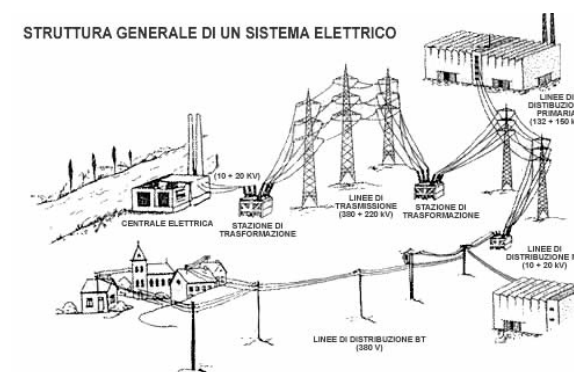
✓ I Sistemi e la Complessità

- La complessità ha i caratteri del **molteplice**, del **ricorsivo**, dell'**incerto**, dell'**instabile**: essa attiene anche al **dominio dell'informazione**.
- La complessità è la proprietà e grandezza caratteristica di ogni sistema, cresce all'aumentare del **numero degli stati possibili**, di modo che la **molteplicità** (o l'**informazione richiesta per individuare ogni stato**) può essere assunta, in prima istanza, come **misura della complessità del sistema**.

Dalla Statistica all'Affidabilità

Ingegneria dei Sistemi

✓ I Sistemi e la Complessità



Dalla Statistica all’Affidabilità

Ingegneria dei Sistemi

✓ I Sistemi e la Complessità



1940



1970



2000



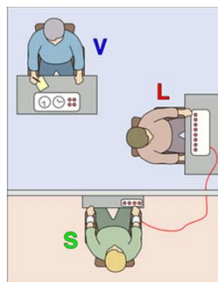
Dalla Statistica all’Affidabilità

Piccoli mondi

La **teoria dei piccoli mondi** è una teoria che sostiene che tutte le reti complesse presenti in natura sono tali che 2 qualunque nodi possono essere collegati da un percorso costituito da un numero relativamente piccolo di collegamenti.

Matematicamente, la teoria è studiata come branca della teoria dei grafi.

La teoria dei piccoli mondi si deve ad una serie di esperimenti condotti da **Stanley Milgram** (psicologo, New York, 1933-1984) che esaminavano la lunghezza media del percorso per reti sociali tra residenti negli Stati Uniti (“**sei gradi di separazione**”).



Nel 1961 Stanley Milgram conduce una serie di controversi esperimenti all'Università di Yale. Gli esperimenti coinvolgevano gente comune, a cui veniva fatto credere di inviare scosse elettriche ad altre persone, al fine di testare il facile condizionamento dell'essere umano nei confronti dell'autorità.



Dalla Statistica all’Affidabilità

Piccoli mondi

La teoria venne sviluppata nel 1998 nell'articolo apparso sulla rivista *Nature*, "*Collective dynamics of «smallworld» networks*" dei matematici Duncan Watts e Steven Strogatz:

Reti di lucciole, router, attori, partner sessuali, etc. hanno almeno due caratteristiche simili: l'alto livello di aggregazione e il basso grado di separazione: nonostante ogni elemento tenda ad avere relazioni prevalentemente con pochi altri (alta aggregazione), ciò non impedisce di ottenere comunque una sua "vicinanza", tramite pochi intermediari, con qualsiasi altro elemento della rete (basso grado di separazione)

Il matematico Paul Erdos si occupò di studiare i grafi casuali, aggiungendo archi a caso tra nodi nell'insieme dato.

Erdos dimostrò che basta una piccola percentuale di archi rispetto al totale per avere un grafo connesso.

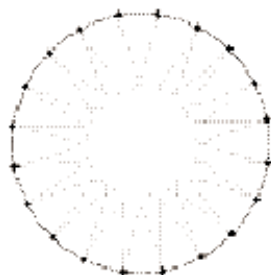
Il grado di separazione di tali grafi è straordinariamente piccolo.

Dalla Statistica all'Affidabilità

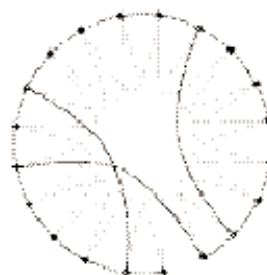
Piccoli mondi

Making Connections

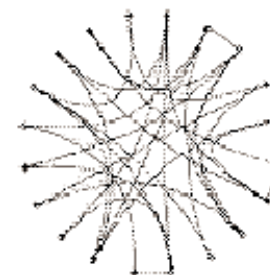
Regular network:
connections to 4 nearest neighbors



Small-world network:
a few long-range connections



Random network:
points connected haphazardly



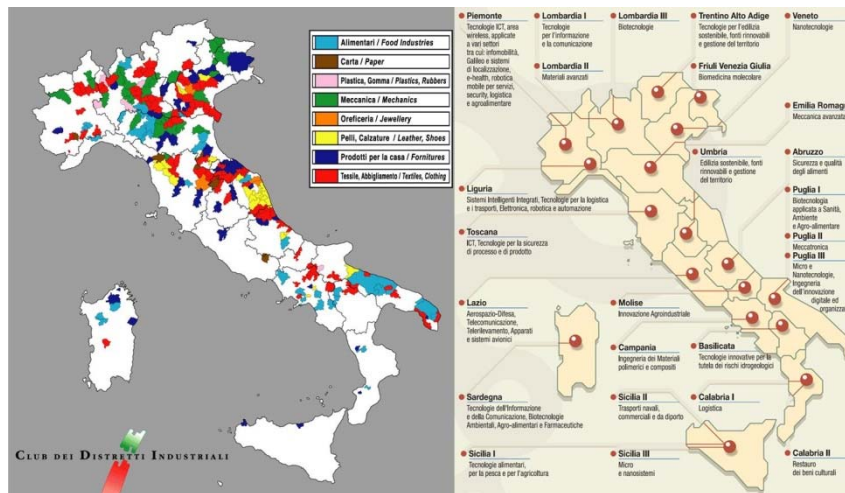
Source: Watts & Strogatz, "Collective Dynamics of Small-World Networks," *Nature*, June 1998.

Dalla Statistica all’Affidabilità

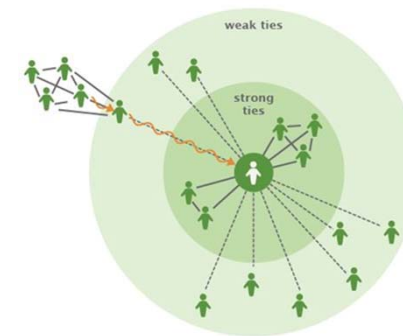
Piccoli mondi: l’importanza dei legami deboli

Mark Granovetter (Jersey City, 20 ottobre 1943, sociologo statunitense) ha dimostrato che:

i **soggetti inseriti in legami deboli**, fatti cioè di conoscenze amicali non troppo strette, hanno più possibilità di accesso ad informazioni e quindi di potenziali posizioni lavorative di proprio interesse, rispetto a coloro che investono socialmente soltanto nei legami forti, cioè i familiari, i parenti e gli amici intimi.



E’ da questa teoria che derivano in Italia i **distretti industriali**, gli studi sul capitale sociale, sull’economia informale etc.



Elementi di Statistica

La Statistica

Analisi affidabilistica di componenti e sistemi

STRUMENTO

Statistica



De ludo Aleæ Liber. 271

tum ut gemina punda his, cum seorsus vinei non contingat, neque talia huiusmodi: videamus tum etiam declinationem, inde persequi mutationem, tum verò, & omnia quedam magis calamitatis, aut felicitatis, quomodoque & ceterorum, tum verò id statuendum conseruim esse aliquid, licet causam nesciamus, quod putamus colligere. Vel, quod aliquid debet, aut possit: tum maxime cum ex his, aliquid minus sequitur, ut ille, qui ludo discedens, amissa omni pecunia, pigrescit imaginem beate Virginis vultuante. Vnde deprehensus à hostilibus, damnatus est poena suspendij. Quando autem causa illa felicitatis, seu sfortium conseruus sit, seu factum, & odo aliqui Vniuersi possit mouere chartas, que ex arbitrio hominum bene, aut male censentur, cum per se nihil significent, dubitatione dignum adeo est, ut facilius sit huius facti causam inuenire sine hoc fine, quam cum illo, nam absque eo ad casum bene reduci potest, ut in constitutione nobium, & spargendo fabas, atque huiusmodi.

dorum, qui fortuna constant, & non robore, vel agilitate corporis quatuor sunt prima genera.

CAPVT XXIII.

De ludis Chartarum, in quibus industria locat est.

¶ Vm ergo in huiusmodi exercitiis indoluit in re incognita, necesse est, ut memoria eorum, qua depositimus, aut teximus, aut reliquimus, aliquid possit, in quibuldam autem plurimum, ut in tripolis, Venetis, ludo. Hoc estimatur ita, quatuor, quinq; & sex punda. In quatuor generibus sunt sexdecim, relinquuntur triginta sex. Dantur quinq; post quatuor, in duobus ludibus sunt decem odo charte, totidem in cumulo relinquuntur primo placent, recinet, cum secundo duplicent, mutat, & nouem primas recipit superiores in cumulo si placent, & ipse suas retinet, si non cumulat cum secundis cumulis, & vides igitur

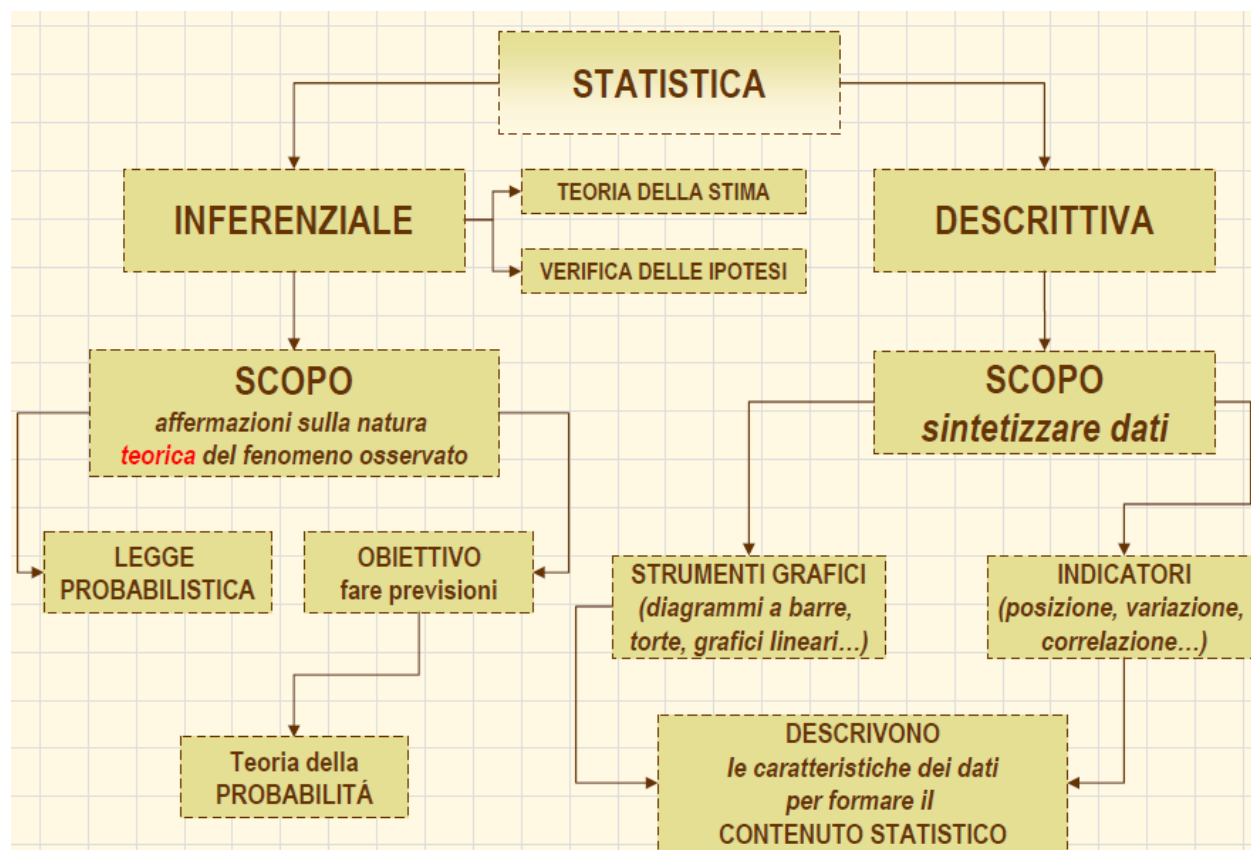
Girolamo Cardano
(Pavia, 24 settembre 1501 – Roma, 21 settembre 1576 circa)



Fra Luca Bartolomeo de Pacioli o anche *Paciolo* (Borgo Sansepolcro, 1445 circa – Roma, 19 giugno 1517)

Elementi di Statistica

La Statistica



Elementi di Statistica

La probabilità

➤ Una delle grandezze fondamentali su cui si basa la Statistica è la **probabilità**, introdotta per la prima volta nel diciassettesimo secolo con i giochi d'azzardo.

➤ Questi giochi, dunque, prevedono lo svolgimento di azioni dall'esito incerto, ma con un ***esito di lungo termine prevedibile***.



➤ Nel caso di lancio di una moneta sufficientemente simmetrica e bilanciata, ci si può aspettare una sostanziale equivalenza fra le volte che esce una faccia della moneta e l'altra.

Elementi di Statistica

La probabilità

- Tale tipo di incertezza e di regolarità su di un arco temporale sufficientemente lungo, tale da consentire il ripetersi di numerosi eventi relativi al fenomeno che si sta analizzando, si verifica spesso nelle **scienze sperimentali**, rendendo interessante l'analisi statistica che diviene uno strumento prezioso e, in molti casi, essenziale.
- La probabilità è stata oggetto di **successive definizioni nel tempo**, in seguito allo svilupparsi di nuove teorie e scuole di pensiero sulla disciplina statistica.

Elementi di Statistica

La Statistica

Secolo	XVI	XVII	XVIII
	<i>Liber de ludo aleæ</i> Girolamo Cardano (1526, pub. nel 1663)	<i>Sulla scoperta dei dadi</i> Galileo Galilei (1656) <i>Si spiega come mai, lanciando tre dadi, il 10 sia più probabile del 9 nonostante che entrambi i risultati si ottengano da un uguale numero di combinazioni.</i>	Corrispondenza tra Pascal e Fermat concetto di probabilità nell'accezione frequentista
			<i>Ars conjectand</i> Jakob Bernoulli (1713) <i>Si dimostrat il teorema di Bernulli, noto anche come legge dei grandi numeri o legge empirica del caso.</i>
Secolo	XX	Bruno de Finetti e poi Leonard Jimmie Savage "Concezione soggettiva della probabilità come "grado di fiducia che una persona ha nel verificarsi dell'evento".	



Elementi di Statistica

La probabilità classica

definizione classica di probabilità:

se un fenomeno casuale può dar luogo, nel suo verificarsi ripetitivo, ad n eventi che si escludono a vicenda e ugualmente possibili, e se n_A di questi danno luogo al verificarsi dell'evento A , allora la probabilità che si verifichi A è data dal rapporto n_A/n .

Elementi di Statistica

La probabilità classica

Esempio classico è il **lancio di dado** (o l'estrazione di una carta, etc.)

- Se consideriamo un dado con sei facce, gli eventi associati all'uscita di una faccia qualunque, sono **ugualmente probabili ed esclusivi** (se esce una faccia, infatti, non può uscirne un'altra).
- Dunque la probabilità che esca una faccia dal lancio del dado, in base alla definizione fornita, è **$1/6$** .

Elementi di Statistica

La probabilità frequentista

Un limite della definizione classica si manifesta quando si incontrano domande del tipo:

qual è la probabilità che un uomo cominci a fumare prima di 50 anni?

Tale quesito è plausibile in statistica, ma non trova una risposta alla luce della probabilità classica.

Per questo motivo si è sviluppata la **teoria frequentista**, che definisce una probabilità di più vasta applicabilità.

Elementi di Statistica

La probabilità frequentista

- Si supponga di considerare l'**esperimento del lancio di una moneta** perfettamente bilanciata; come si è visto nella trattazione classica, la probabilità che esca una faccia o l'altra è circa la stessa, **a priori**.

La definizione classica, infatti, fornisce una valutazione a priori.

- Se a questa considerazione teorica si fa seguire l'esperimento, si scopre che il numero di volte che esce una faccia della moneta, o l'altra, è circa $1/2$, ma non proprio $1/2$.

La definizione frequentista, infatti, fornisce una valutazione della probabilità sulla base del verificarsi effettivo degli eventi, cioè della frequenza di accadimento di un determinato evento su un certo numero di tentativi.

Elementi di Statistica

La probabilità soggettiva

Le teorie probabilistiche appena viste (classica e frequentista) hanno un elemento in comune:

**entrambe fanno riferimento ad eventi ripetibili
in condizioni simili o abbastanza simili.**

Può accadere, però, di dovere **valutare la probabilità di un evento non necessariamente ripetibile, o comunque ripetibile in condizioni non simili** (ad esempio, può accadere di considerare la probabilità che scoppi una guerra mondiale prima di una certa data).

Questo tipo di probabilità, che non fa riferimento ad eventi ripetibili in condizioni simili, bensì a fatti non oggettivi o pressoché unici, viene detta appunto **probabilità soggettiva**.

Elementi di Statistica

Teoria generale della probabilità

La costruzione della teoria generale della probabilità richiede l'ausilio di uno strumento matematico importante, la **teoria degli insiemi**, a cui si fa riferimento nel seguito.

Si supponga di considerare la collezione di tutti gli oggetti che riguardano una certa trattazione (ad esempio, parlando di numeri, si considereranno tutti i numeri), detti nel seguito **elementi**.

Tale collezione costituisce un insieme, in particolare costituisce l'**insieme universale W** .

Elementi di Statistica

Teoria generale della probabilità

Si considera successivamente una collezione particolare di oggetti facenti parte di W , la quale costituisce a sua volta un insieme A che è un **sottoinsieme** di W :

$$A \subset W \quad (1)$$

La 1 indica che l'insieme A è contenuto in W . Diversamente si potrebbe considerare un altro sottoinsieme B , di A questa volta, e scrivere:

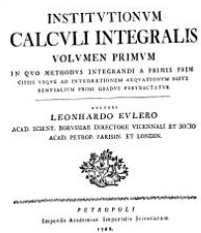
$$B \subseteq A \quad (2)$$

che dice che B è contenuto in A e, al limite, può coincidere con esso.

Elementi di Statistica

Teoria generale della probabilità

La trattazione degli insiemi può essere semplificata dal ricorso ai cosiddetti **diagrammi di (Eulero-) Venn**, che rappresentano graficamente le diverse situazioni. Ad esempio la 2 è rappresentata nella successiva Figura 1.



(Basilea, 15 aprile 1707 – San Pietroburgo, 18 settembre 1783)

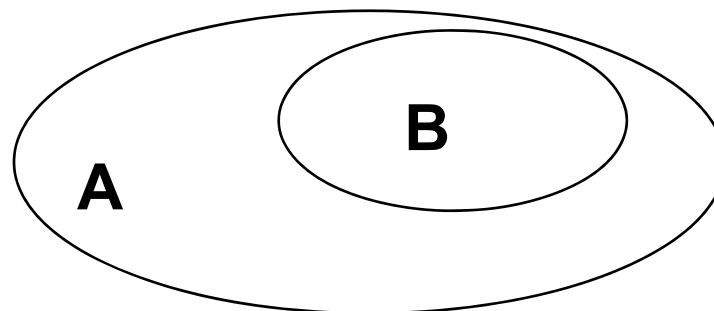
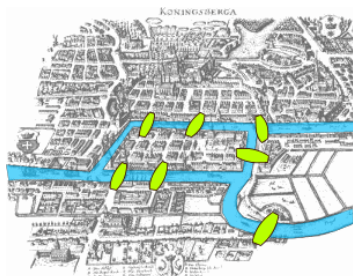


Fig. 1 - Rappresentazione di Venn di un insieme contenuto in un altro



(Kingston upon Hull, 4 agosto 1834 – Cambridge, 4 aprile 1923)

Elementi di Statistica

Teoria generale della probabilità

L'insieme privo di elementi è detto **insieme vuoto** e si indica con il simbolo ϕ .

Sugli insiemi, inoltre, è possibile definire alcune operazioni, l'unione, l'intersezione e la differenza.

L'**unione** di un insieme A ed uno B si indica con la seguente simbologia:

$$A \cup B \qquad (3)$$

Elementi di Statistica

Teoria generale della probabilità

L'unione di 2 insiemi dà luogo ad un altro insieme C, che è la collezione di tutti gli elementi che appartengono ad A o a B (Figura 2 o 3 nel caso di insiemi **disgiunti**).

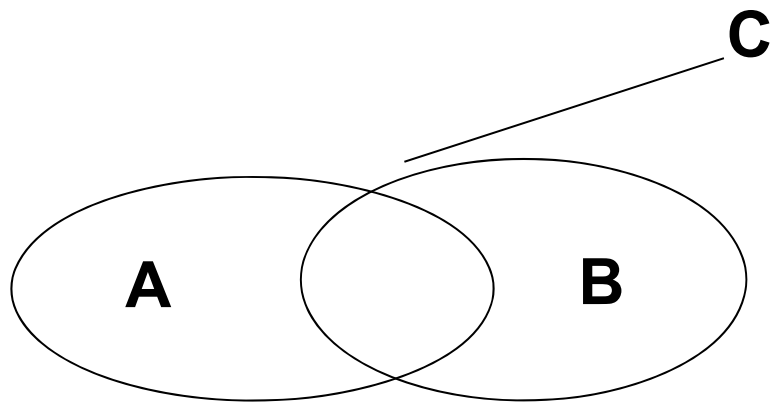


Fig. 2 - Operazione di unione fra due insiemi

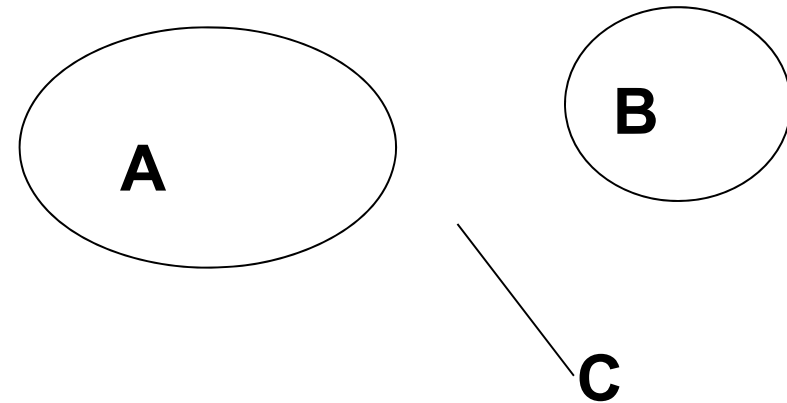


Fig. 3 - Operazione di unione fra due insiemi disgiunti

Elementi di Statistica

Teoria generale della probabilità

L'**intersezione** di due insiemi A e B rappresenta, invece, l'insieme C degli elementi che appartengono ad A e B allo stesso tempo (Figura 4):

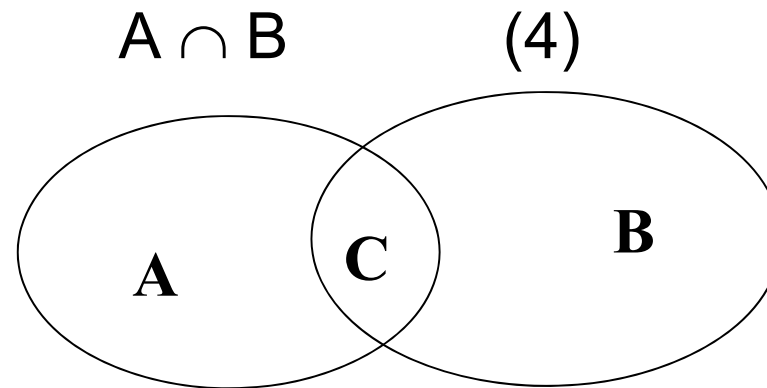


Fig. 4 - Operazione di intersezione fra due insiemi

E' evidente che l'intersezione di due insiemi disgiunti è pari all'insieme vuoto ϕ .

Elementi di Statistica

Teoria generale della probabilità

L'operazione di **differenza** (5) di due insiemi A e B rappresenta l'insieme i cui elementi sono la differenza fra gli elementi dei due insiemi di partenza, cioè l'insieme degli elementi di A che non sono presenti in B:

$$A - B \quad (5)$$

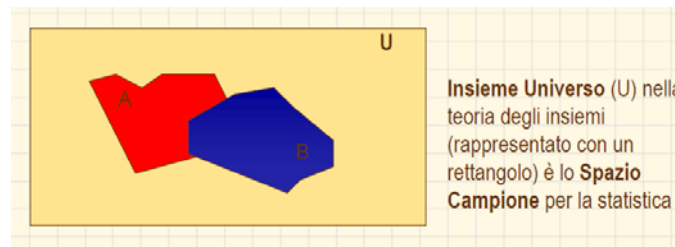
Un insieme $\overline{A}$, infine, è **complementare** di un altro insieme A se è costituito da tutti gli elementi che appartengono all'insieme universale W ma non ad A.

Elementi di Statistica

Teoria generale o assiomatica della probabilità

Gli accenni fatti alla teoria insiemistica e, soprattutto, l'introduzione delle rappresentazioni grafiche dovute a Venn, aiutano alla costruzione della teoria probabilistica che segue, di tipo **assiomatico** e **generale**.

Si definisce lo **spazio campionario** (o **campione**) come la collezione o la totalità di tutti i possibili risultati di un esperimento, ove la parola campionario sta ad indicare il fatto che il risultato di un esperimento è un **campione** di tutti i risultati possibili.



Elementi di Statistica

Teoria generale o assiomatica della probabilità

Un sottoinsieme dello spazio campionario appena visto costituisce un **evento**, e la famiglia di tutti gli eventi associati ad uno specifico esperimento (**fra tutti quelli possibili nell'Universo**) rappresenta lo **spazio degli eventi** o **insieme completo degli eventi**.

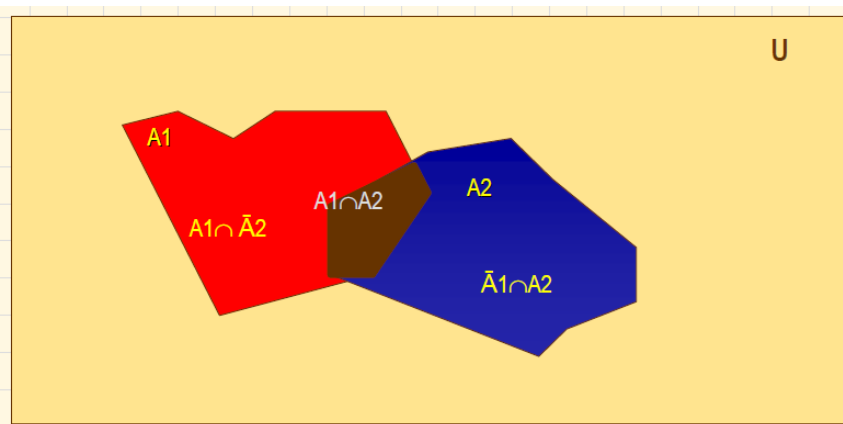
La probabilità, secondo il punto di vista assiomatico, è dunque una funzione fra insiemi che gode di alcune proprietà particolari.

Assegnato lo **spazio campionario** Ω e lo **spazio degli eventi** A associato ad un certo esperimento, la probabilità di un **generico evento** E , sottoinsieme di A , è un numero p tale che:

Elementi di Statistica

Teoria generale o assiomatica della probabilità

1. $p(E) \geq 0$;
2. $p(E) \leq 1$;
3. $p(E) = 1$ se l'evento è certo;
4. dati due eventi A e B è:



$$p(A \cup B) + p(A \cap B) = p(A) + p(B) \quad (6)$$

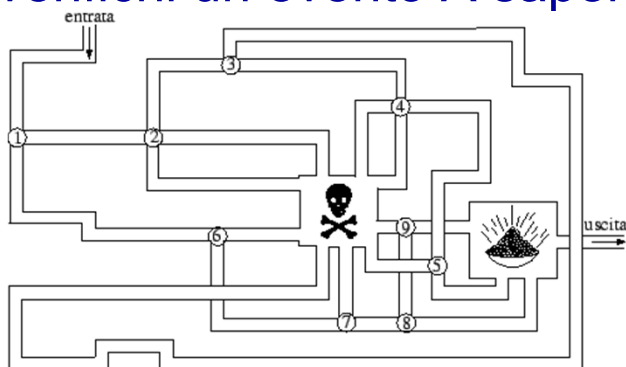
5. dati due eventi A e B:

$$p(A \cap B) = p(A/B)p(B) = p(B/A)p(A) \quad (7)$$

Elementi di Statistica

Teoria generale o assiomatica della probabilità

Gli assiomi 4 e 5, in quanto tali, non sono dimostrabili ma possono essere visualizzati alla luce della teoria insiemistica. L'assioma 5, inoltre, introduce la **probabilità condizionale** $p(A/B)$, cioè la probabilità che si verifichi un evento A sapendo che l'evento B si è verificato.

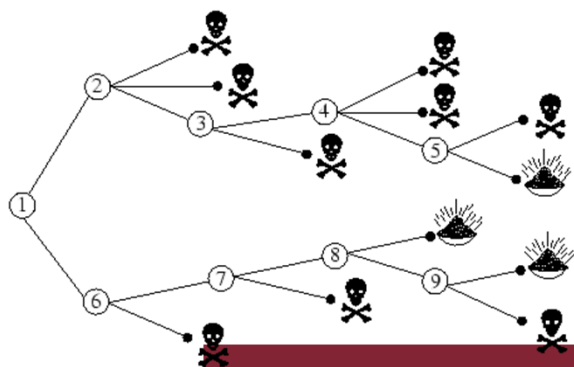


La probabilità di arrivare al tesoro (T) passando prima per il bivio 5 vale:

$$P(T | E_1 \cap E_2 \cap E_3 \cap E_4 \cap E_5)$$

$$P(1) \cdot P(E_2 | E_1) \cdot P(E_3 | E_1 \cap E_2) \dots$$

$$1 \cdot \frac{1}{2} \cdot \frac{1}{3} \cdot \frac{1}{2} \cdot \frac{1}{3} \cdot \frac{1}{2} = \frac{1}{72}.$$



Elementi di Statistica

Teoria generale o assiomatica della probabilità

Dall'assioma 4 discende il **teorema delle probabilità totali**:

se due eventi A e B sono incompatibili (cioè mutuamente esclusivi, $p(A \cap B) = 0$) allora la probabilità del verificarsi dell'uno dell'altro è la somma delle probabilità di ciascuno dei due eventi

$$p(A \cup B) = p(A) + p(B) \quad (8)$$

Elementi di Statistica

Teoria generale o assiomatica della probabilità

Dall'assioma 5, invece, discende il **teorema delle probabilità composte**:

se due eventi A e B sono indipendenti, allora la probabilità che si verifichino contemporaneamente è pari al prodotto delle probabilità associate ai singoli eventi

$$p(A \cap B) = p(A) p(B) \quad (9)$$

Elementi di Statistica

Teoria generale o assiomatica della probabilità

Assegnato uno spazio degli eventi, *se tutti gli eventi che si possono verificare sono mutuamente esclusivi*, in base alla 8 (*teorema della probabilità totali*), si ha:

$$\sum_i p(E_i) = p(E_1) + p(E_2) + \dots = p(E_1 \cup E_2 \cup \dots) = 1 \quad (10)$$

Elementi di Statistica

Teoria generale o assiomatica della probabilità

La **legge dei grandi numeri**, o **legge empirica del caso**, (**teorema di Bernoulli**) afferma che la frequenza di un evento associato ad una serie di prove effettuate su un sistema è pari alla sua probabilità se il numero delle prove è molto grande:

$$\lim_{N \rightarrow \infty} f = \lim_{N \rightarrow \infty} \frac{n}{N} = p \quad (11)$$

ove N è il numero di esperimenti effettuato e n il numero di volte che si è verificato l'evento prestabilito.

Elementi di Statistica

Teoria generale o assiomatica della probabilità

La trattazione delle probabilità condizionali induce ad introdurre il **teorema di Bayes** (da qualcuno chiamato **teorema della probabilità delle cause**).

Esso è impiegato per **calcolare la probabilità di una causa che ha prodotto l'evento verificato**:

assegnati due eventi R e S, vale l'assioma 7

da cui discende l'espressione:

$$\begin{aligned} p(R/S) &= \frac{p(R \cap S)}{p(S)} = \frac{p(S/R) \cdot p(R)}{p(S)} = \\ &= \frac{p(S/R) \cdot p(R)}{p(S/R) \cdot p(R) + p(S/\bar{R}) \cdot p(\bar{R})} \quad (12) \end{aligned}$$



Thomas Bayes
(Londra, 1702 – Royal Tunbridge Wells, 17 aprile 1761)

Elementi di Statistica

Teoria generale o assiomatica della probabilità

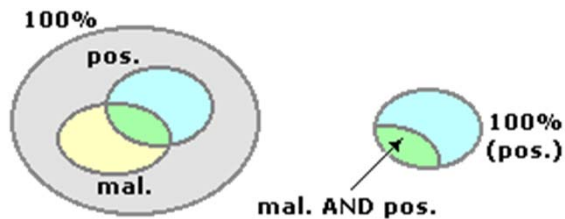
Supponendo che R significhi possedere certi assegnati requisiti qualitativi e S avere superato una certa prova,

$p(R/S)$ rappresenta la probabilità che un materiale
che ha superato una certa prova
possieda effettivamente i requisiti richiesti.

Tale teorema trova una importante applicazione nel caso della **accettazione di risorse produttive (materie prime) in una Azienda**, o - per esempio - si può calcolare la probabilità che una certa persona soffra della malattia per cui ha eseguito il **test diagnostico** (nel caso in cui questo sia risultato negativo) o viceversa non sia affetta da tale malattia (nel caso in cui il test sia risultato positivo).

Elementi di Statistica

Teoria generale o assiomatica della probabilità



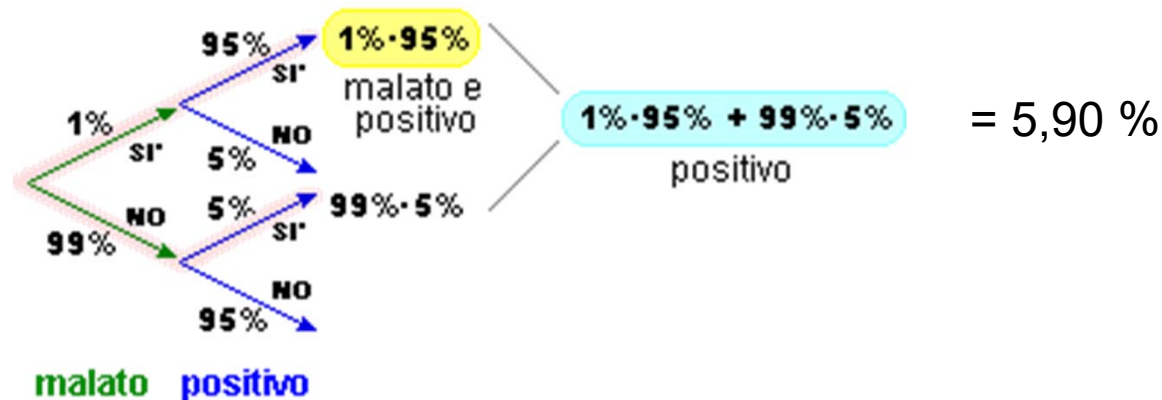
Un certo test sanitario per valutare la presenza (esito positivo) o assenza (esito negativo) della malattia X ha attendibilità del 95%:
- in caso di presenza c'è il 95% di probabilità che l'esito sia positivo
- in caso di assenza il 95% di probabilità che sia negativo.

Si sa da statistiche che l'1% della popolazione è affetta dalla malattia X.

probabilità condizionata di essere malato sotto la condizione di essere positivo

Se per una persona il test dà esito positivo, qual è la probabilità che essa sia realmente malata?

$$\begin{aligned} & p(\text{"essere malato"} \mid \text{"risultare positivo"}) = \\ & = p(\text{"essere malato"} \text{ et } \text{"risultare positivo"}) / p(\text{"risultare positivo"}) = \\ & = 95\% / 5,90\% = \mathbf{16\%} \end{aligned}$$



Elementi di Statistica

Teoria generale o assiomatica della probabilità

La probabilità di estrarre l'asso di picche da un mazzo di 52 carte è:

$$p(\text{Asso picche}) = 1/52$$

La probabilità di estrarre una carta di un seme nero (fiori o picche) è:

$$p(\text{carta nera}) = 1/2$$

La probabilità di aver estratto l'asso di picche, sapendo che la carta estratta è nera, è:

$$p(\text{Asso picche} \mid \text{carta nera}) = 1/26$$

La probabilità di aver estratto lo stesso asso, sapendo che la carta estratta è rossa, è:

$$p(\text{Asso picche} \mid \text{carta rossa}) = 0$$

La probabilità di aver estratto una carta nera, sapendo che è stato estratto l'asso di picche, è:

$$p(\text{carta nera} \mid \text{Asso picche}) = 1.$$

Elementi di Statistica

Teoria generale o assiomatica della probabilità

Si consideri una scuola che ha il **60% di studenti maschi** e il **40% di studentesse femmine**.

Le studentesse indossano **in egual numero gonne o pantaloni**; gli studenti indossano **tutti quanti i pantaloni**.

Un osservatore, da lontano, **nota un generico studente coi pantaloni**: qual è la probabilità che quello studente sia una femmina?

Il problema può essere risolto con il teorema di Bayes, ponendo **l'evento A che lo studente osservato sia femmina**, e **l'evento B che lo studente osservato indossi i pantaloni**. Per calcolare $P(A|B)$, dovremo sapere:

- $P(A)$, ovvero la probabilità che lo studente sia femmina senza nessun'altra informazione. Dato che l'osservatore vede uno studente a caso, ciò significa che tutti gli studenti hanno la stessa probabilità di essere osservati. Essendo le studentesse il 40% del totale, la probabilità risulterà **2/5**.

- $P(A')$, ovvero la probabilità che lo studente sia maschio senza nessun'altra informazione. Essendo A' l'evento complementare di A , risulta **3/5**.

- $P(B|A)$, ovvero la probabilità che uno studente femmina indossi i pantaloni (ossia la probabilità che, verificato l'evento che lo studente sia femmina, si verifichi l'evento che indossi i pantaloni). Poiché indossano gonne e pantaloni in egual numero, la probabilità sarà di **1/2**.

- $P(B|A')$, ovvero la probabilità che uno studente indossi i pantaloni, noto che lo studente è maschio. Tutti gli studenti maschi indossano i pantaloni, quindi vale **1**.

- $P(B)$, ovvero la probabilità che uno studente qualsiasi (maschio o femmina) indossi i pantaloni. Poiché il numero di coloro che indossa i pantaloni è di 80 (60 maschi + 20 femmine) su 100 studenti fra maschi e femmine, la probabilità $P(B)$ è di $80/100 = 4/5$.

Ciò detto, possiamo applicare il teorema:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{\frac{1}{2} \times \frac{2}{5}}{\frac{4}{5}} = \frac{1}{4}.$$

C'è pertanto una **probabilità pari a 1/4 che lo studente sia femmina, cioè il 25%**.

Elementi di Statistica

Distribuzioni di probabilità

In genere, si è interessati ad una **distribuzione di probabilità** piuttosto che ad un singolo valore di questa grandezza, cioè ad una distribuzione dei valori della probabilità associati ad un **insieme completo di eventi**.

Tutte le probabilità espresse nella distribuzione di valori devono essere comprese fra 0 ed 1 e devono essere tali che la loro somma sia pari ad 1 (**proprietà di chiusura**)

$$\sum_i p(E_i) = p(E_1) + p(E_2) + \dots = 1 \quad (13)$$

Elementi di Statistica

Distribuzioni di probabilità

In particolare, supponendo di considerare un insieme di numeri X posto in corrispondenza biunivoca con un insieme completo di eventi, è possibile valutare la probabilità associata al manifestarsi di uno di questi numeri.

Il generico x appartenente a tale insieme numerico viene detto **variabile aleatoria** e può assumere valori **discreti** oppure **continui**.

Elementi di Statistica

Distribuzioni di probabilità

Per comprendere il significato di una variabile aleatoria, si consideri il seguente caso particolare:

- un disco può ruotare intorno al suo centro ed ha un raggio tracciato in modo da essere facilmente visibile;
- assegnata un'ascissa angolare (cioè un riferimento fisso rispetto alla rotazione del disco), la variabile aleatoria (in questo caso continua) è l'angolo che durante la rotazione forma il raggio visibile rispetto al riferimento fisso; ci si chiede quale sia l'angolo alla fine della generica rotazione imposta;
- evidentemente, se la rotazione è imposta casualmente, anche l'angolo formato dal raggio predeterminato con il riferimento fisso avrà, alla fine della rotazione, un valore casuale; poiché tale angolo può assumere con continuità valori differenti, si ha a che fare con una variabile aleatoria continua.

Elementi di Statistica

Distribuzioni di probabilità

Assegnata una **variabile aleatoria**, si può considerare:

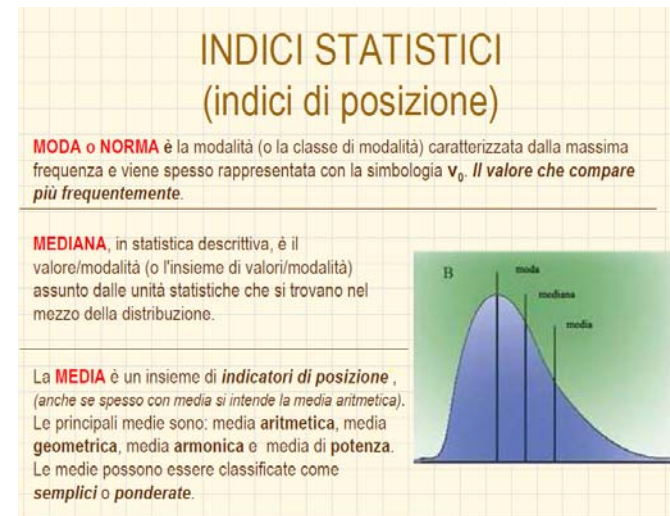
- una sua generica distribuzione

e

- le grandezze caratteristiche:

- il valore atteso, o speranza matematica o valore medio,

- la varianza o scarto quadratico medio, e la sua radice quadrata, la deviazione standard.



Elementi di Statistica

Distribuzioni di probabilità

Il valore atteso $\bar{X}$ rappresenta il valore baricentrico della distribuzione, dunque il valore di x più probabile.

Indicando con $D(x)$ la densità di probabilità ($D(x) = \frac{dp}{dx}$), cioè la probabilità infinitesima associata ad una variazione infinitesima di x , risulta:

$$\bar{X} = \frac{\int_{-\infty}^{+\infty} x \cdot D(x) dx}{\int_{-\infty}^{+\infty} D(x) dx} = \int_{-\infty}^{+\infty} x \cdot D(x) dx \quad (14)$$

Elementi di Statistica

Distribuzioni di probabilità

La **varianza** σ^2 è una grandezza indicativa dello scostamento dei valori della distribuzione dal valore più probabile, cioè il valore atteso. Essa, in sostanza, fornisce una valutazione della “apertura” del diagramma della funzione di distribuzione e la sua espressione è la seguente:

$$\sigma^2 = \int_{-\infty}^{+\infty} (x - \bar{x})^2 \cdot D(x) dx \quad (15)$$

Dalla quale discende la **deviazione standard** σ :

$$\sigma = \sqrt{\int_{-\infty}^{+\infty} (x - \bar{x})^2 \cdot D(x) dx} \quad (16)$$

Elementi di Statistica

Distribuzioni di probabilità

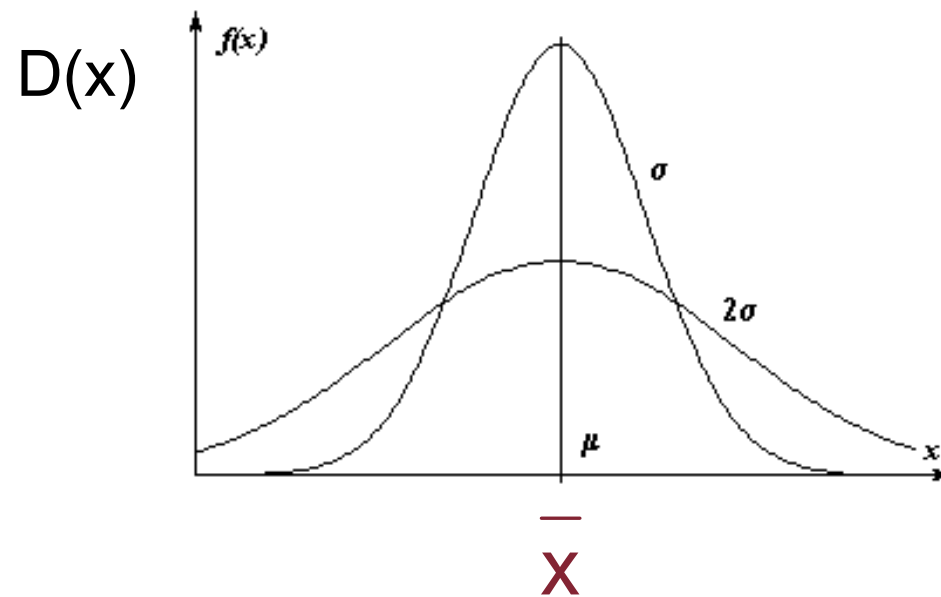


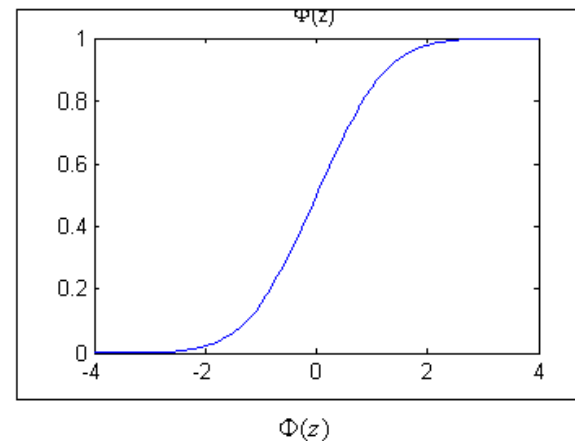
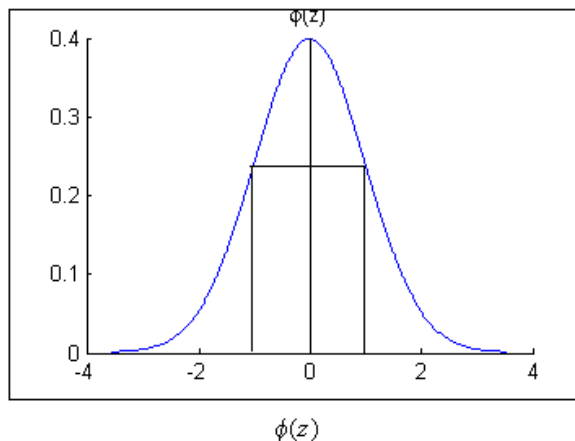
Fig. 5 - Esempio di distribuzione probabilistica

Elementi di Statistica

Distribuzioni di probabilità

Assegnata una certa distribuzione probabilistica è possibile risalire alla cosiddetta **funzione di ripartizione F(x)**, che rappresenta la probabilità che l'evento si verifichi dall'origine prescelta, in genere $-\infty$, ed x:

$$F(x) = \int_{-\infty}^x D(x)dx \quad (17)$$



Elementi di Statistica

Distribuzioni di probabilità

Principali distribuzioni statistiche discrete:

- Distribuzione binomiale o di Bernoulli
- Distribuzione di Poisson

Principali distribuzioni statistiche continue:

- Distribuzione normale o di Gauss
- Distribuzione lognormale
- Distribuzione del χ^2
- Distribuzione di t di Student
- Distribuzione F di Fisher-Snedecor
- Distribuzione esponenziale o di Laplace
- Distribuzione di Weibull



Elementi di Statistica

Distribuzione di Bernoulli

La distribuzione di Bernoulli o binomiale è una distribuzione discreta, su due soli valori: 0 e 1, detti anche *fallimento* e *successo*, espressa dalla seguente espressione:

$$p(k) = \binom{n}{k} p^k \cdot q^{n-k} \quad (18)$$



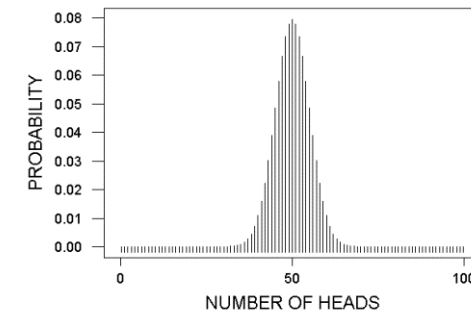
(Basilea, 27 dicembre 1654 – Basilea, 16 agosto 1705)

Descrive il numero di successi in un processo di Bernoulli, ovvero la variabile aleatoria $S_n = X_1 + X_2 + \dots + X_n$ che somma n variabili aleatorie indipendenti di uguale distribuzione di Bernoulli.

Esempi sono i risultati di una serie di lanci di una stessa moneta o di una serie di estrazioni da un'urna (con reintroduzione), ognuna delle quali può fornire due soli risultati: il *successo* con probabilità p e il *fallimento* con probabilità $q = 1 - p$.

Elementi di Statistica

Distribuzione di Bernoulli



Si considerino due eventi E_1 ed E_2 , e sia p la probabilità costante di E_1 e q la probabilità costante di E_2 , con $p + q = 1$, per la proprietà di chiusura.

Eseguendo n prove indipendenti, sia k il numero di volte che si verifica l'evento E_1 ; k è una variabile aleatoria discreta che può assumere i valori $0, 1, 2, \dots, n$.

La probabilità che in n prove si presenti una certa sequenza è $p^k \cdot q^{n-k}$.

Il numero di sequenze in cui E_1 compare k volte è dato dal coefficiente binomiale $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ (fornisce il numero delle combinazioni semplici di n elementi di classe k).

COMBINAZIONI DI TRE LANCI DI UNA MONETA	
TESTA TESTA TESTA	TTT
CROCE CROCE CROCE	CCC
TESTA CROCE CROCE	TCC
CROCE TESTA TESTA	CTT
TESTA CROCE TESTA	TCT
CROCE TESTA CROCE	CTC
TESTA TESTA CROCE	TTC
CROCE CROCE TESTA	CCT

Allora, essendo le n prove indipendenti, in base al teorema delle probabilità composte risulta valida la 18.

Elementi di Statistica

Distribuzione di Bernoulli

Il valore atteso è:

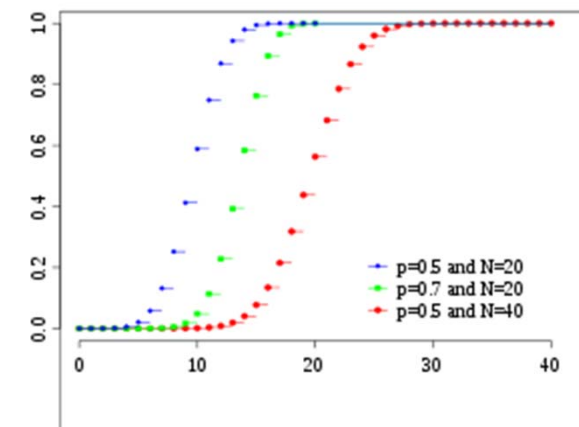
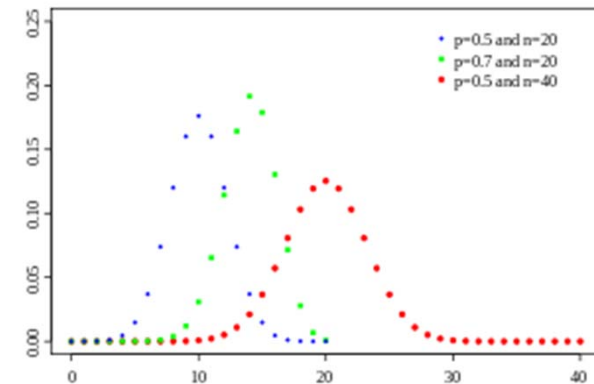
$$\bar{k} = n \cdot p \quad (19)$$

e la varianza

$$\sigma^2 = n \cdot p \cdot q \quad (20)$$

con:

- n, il numero di prove effettuate
- p, la probabilità di successo della singola prova di Bernoulli X_i



Elementi di Statistica

Distribuzione di Bernoulli

Esempio 1:

Si estrae per 5 volte una carta da un mazzo da 52, rimettendo ogni volta la carta estratta dal mazzo.

Determinare la funzione di probabilità della variabile aleatoria k = numero di volte in cui è uscita una carta di fiori.

Calcolare, inoltre la probabilità che:

- a) esca due volte una carta di fiori;
- b) non esca mai una carta di fiori.

$$p = \frac{1}{4} \quad q = \frac{3}{4} \quad \text{con } n=5 \qquad p_k = \binom{n}{k} p^k \cdot q^{n-k} \quad \text{con } 0 \leq k \leq n$$

$$p_0 = \binom{5}{0} \cdot \left(\frac{1}{4}\right)^0 \cdot \left(\frac{3}{4}\right)^5 = 1 \cdot 1 \cdot \frac{243}{1024} = \frac{243}{1024}$$

$$p_1 = \binom{5}{1} \cdot \left(\frac{1}{4}\right)^1 \cdot \left(\frac{3}{4}\right)^4 = 5 \cdot \frac{1}{4} \cdot \frac{81}{256} = \frac{405}{1024}$$

$$p_2 = \binom{5}{2} \cdot \left(\frac{1}{4}\right)^2 \cdot \left(\frac{3}{4}\right)^3 = \frac{5 \cdot 4}{2} \cdot \frac{1}{16} \cdot \frac{27}{64} = \frac{540}{2048} = \frac{270}{1024}$$

$$p_3 = \binom{5}{3} \cdot \left(\frac{1}{4}\right)^3 \cdot \left(\frac{3}{4}\right)^2 = \frac{5 \cdot 4 \cdot 3}{3 \cdot 2} \cdot \frac{1}{64} \cdot \frac{9}{16} = \frac{540}{6144} = \frac{90}{1024}$$

$$p_4 = \binom{5}{4} \cdot \left(\frac{1}{4}\right)^4 \cdot \left(\frac{3}{4}\right)^1 = \frac{5 \cdot 4 \cdot 3 \cdot 2}{4 \cdot 3 \cdot 2} \cdot \frac{1}{256} \cdot \frac{3}{4} = \frac{15}{1024}$$

$$p_5 = \binom{5}{5} \cdot \left(\frac{1}{4}\right)^5 \cdot \left(\frac{3}{4}\right)^0 = 1 \cdot \frac{1}{1024} \cdot 1 = \frac{1}{1024}$$

$$p_2 = \frac{270}{1024} = 0,263 \quad \text{probabilità che esca due volte una carta di fiori}$$

$$p_0 = \frac{243}{1024} = 0,237 \quad \text{probabilità che non esca mai una carta di fiori}$$

Elementi di Statistica

Distribuzione di Bernoulli

Esempio 2:

Una macchina produce pezzi con un tasso di difettosità del 10% ($p = 0,1$) in un lotto di $n = 20$ pezzi.

Studiare la variabile casuale $k =$ numero dei pezzi difettosi.

Calcolare il numero medio di pezzi difettosi, lo scarto quadratico medio.

Sviluppando col foglio di calcolo:

k	$\binom{n}{k}$	p^k	q^{n-k}	$p(x)$
0	1	1	0,121577	0,121577
1	20	0,1	0,135085	0,27017
2	190	0,01	0,150095	0,28518
3	1140	0,001	0,166772	0,19012
4	4845	0,0001	0,185302	0,089779
5	15504	0,00001	0,205891	0,031921
6	38760	0,000001	0,228768	0,008867
7	77520	1E-07	0,254187	0,00197
8	125970	1E-08	0,28243	0,000356
9	167960	1E-09	0,313811	5,27E-05
10	184756	1E-10	0,348678	6,44E-06

trattandosi di una distribuzione binomiale deve essere $\mu = n \cdot p = 20 \cdot 0,1 = 2$

mentre $\sigma^2 = n \cdot p \cdot q = 20 \cdot 0,1 \cdot 0,9 = 1,8$ da cui avremo $\sigma = \sqrt{1,8} = 1,34$

Elementi di Statistica

Distribuzione di Bernoulli

Esempio 3:

In un'azienda vi sono $n = 15$ macchine in funzione.

La probabilità che una macchina si guasti (k) durante la giornata è $p = 0,3$.

Se le macchine funzionano indipendentemente e non possono essere aggiustate in giornata, determinare:

- il numero più probabile di macchine in funzione a fine giornata ($n - k$);
- lo scarto quadratico medio delle macchine in funzione a fine giornata.

$\mu = np = 15 \cdot 0,3 = 4,5$ valor medio delle macchine guaste

quindi il valor medio delle macchine in funzione sarà 10,5.

Lo scarto quadratico medio delle macchine in funzione corrisponde allo scarto quadratico medio delle macchine guaste

$$\sigma^2 = npq = 15 \cdot 0,3 \cdot 0,7 = 3,15 \quad \text{quindi} \quad \sigma = \sqrt{3,15} = 1,77$$

k	$\binom{n}{k}$	p^k	q^{n-k}	$p(x)$
0	1	1	0,004748	0,004748
1	15	0,3	0,006782	0,03052
2	105	0,09	0,009689	0,09156
3	455	0,027	0,013841	0,17004
4	1365	0,0081	0,019773	0,218623
5	3003	0,00243	0,028248	0,20613
6	5005	0,000729	0,040354	0,147236
7	6435	0,000219	0,057648	0,08113
8	6435	6,56E-05	0,082354	0,03477
9	5005	1,97E-05	0,117649	0,01159
10	3003	5,9E-06	0,16807	0,00298

la probabilità massima si ha per $n=4$ macchine guaste, quindi, il numero più probabile di macchine in funzione a fine giornata è $15-4=11$ macchine.

Elementi di Statistica

Distribuzione di Poisson

La distribuzione di Poisson è anch'essa una distribuzione discreta.

Essa esprime la probabilità del numero di eventi che si verificano successivamente ed indipendentemente in un dato intervallo di tempo, sapendo che *mediamente* se ne verifica un certo numero *m*.

Ad esempio, si utilizza una distribuzione di Poisson per misurare il numero di chiamate ricevute in un call-center in un determinato arco temporale, come una mattinata lavorativa. Questa distribuzione, che ha a che vedere con eventi remoti, è anche nota come **legge degli eventi rari**.



Siméon-Denis Poisson
(Pithiviers, 21 giugno 1781 – Parigi, 25 aprile 1840)

Elementi di Statistica

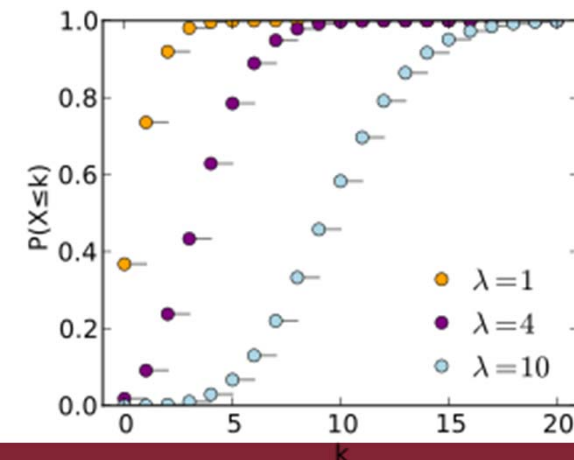
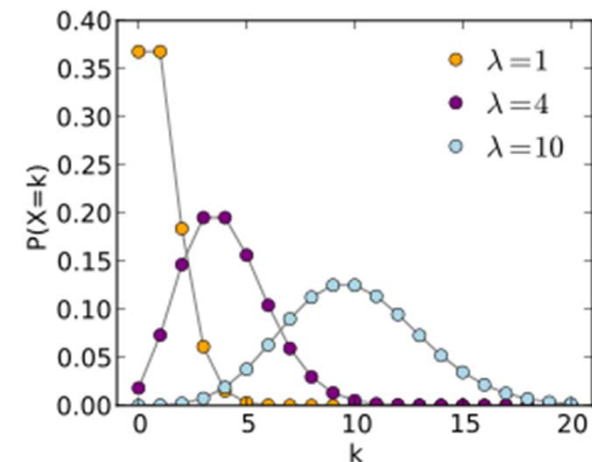
Distribuzione di Poisson

Assegnata una variabile aleatoria discreta k , noto o fissato m , risulta essere:

$$p(k) = \frac{m^k \cdot e^{-m}}{k!} \quad (21)$$

$$\bar{k} = \sigma^2 = m \quad (22)$$

Tale funzione di distribuzione si manifesta soprattutto in presenza di eventi rari, perciò è anche detta **distribuzione degli eventi rari**.



Elementi di Statistica

Distribuzione di Poisson

Esempio 1:

In una industria manifatturiera esiste una probabilità dello 0,4% che un pezzo fabbricato risulti difettoso (k).

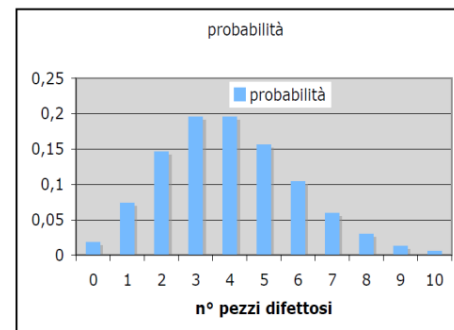
Si calcoli la probabilità che in una partita di $n = 1.000$ pezzi fabbricati:

- 1) nessun pezzo sia difettoso
- 2) ci sia un solo pezzo difettoso
- 3) ci sia più di un pezzo difettoso
- 4) ci siano esattamente 4 pezzi difettosi
- 5) ci siano meno di 6 pezzi difettosi

Si calcoli infine il numero medio e la deviazione standard del numero di pezzo difettosi per ogni campione di 1000 pezzi e si faccia un grafico della probabilità di trovare n pezzi difettosi sino a $n=10$.

Il valore medio dei pezzi difettosi è $m = 0,4/100 \cdot 1.000 = 4 = \bar{k} = \sigma^2$

$$p(0) = \frac{4^1 \cdot e^{-4}}{0!} = 0,0183...$$



Elementi di Statistica

Distribuzione di Gauss

La distribuzione di Gauss, o normale, è fra le più comuni.

E' una distribuzione di probabilità **continua**, spesso usata per descrivere variabili aleatorie a valori reali che tendono a concentrarsi attorno a **un singolo valore medio**.

E' considerata il caso base delle distribuzioni di probabilità continue a causa del suo ruolo nel teorema del limite centrale:

la somma di n variabili casuali con media e varianza finite tende a una distribuzione normale al tendere di n all'infinito



Johann Carl Friedrich Gauss
(Braunschweig, 30 aprile 1777 – Gottinga, 23 febbraio 1855)

Elementi di Statistica

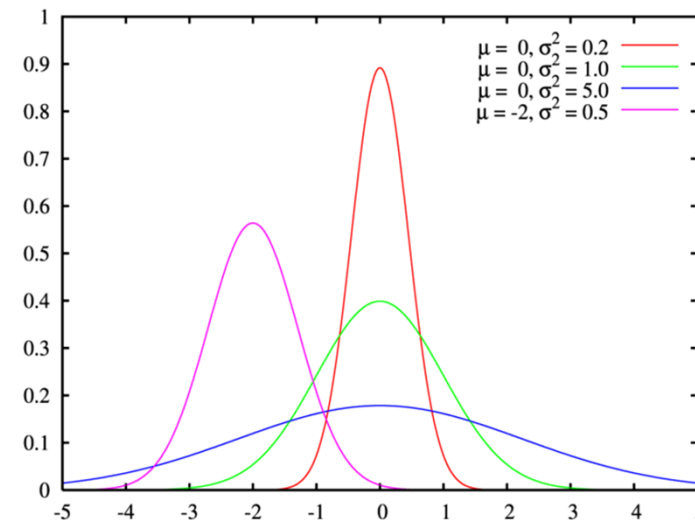
Distribuzione di Gauss

Si dice che una assegnata **variabile aleatoria continua**, essa segue la legge di Gauss se risulta:

$$D(x) = \frac{1}{\sigma \cdot \sqrt{2 \cdot \pi}} \cdot e^{-\frac{(x-m)^2}{2 \cdot \sigma^2}} \quad (23)$$

Il valore medio risulta essere:

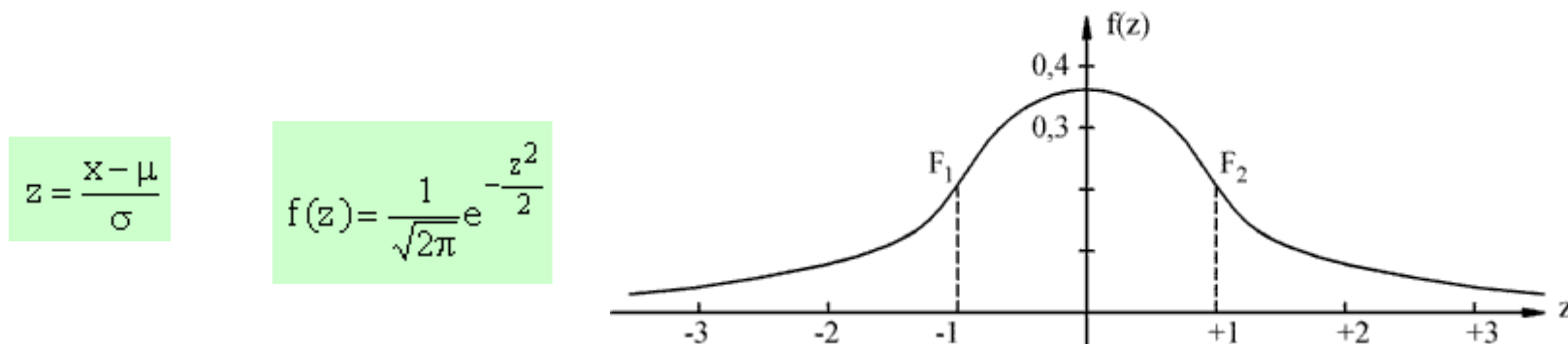
$$\bar{X} = m \quad (24)$$



Elementi di Statistica

Distribuzione di Gauss

La curva a cui ci si può sempre ricondurre (con un'opportuna trasformazione) è la curva della distribuzione normale standardizzata:



La curva che la rappresenta volge la sua concavità verso il basso nell'intervallo $(-1, 1)$ i punti di ascissa $z=\pm 1$ sono dei flessi:

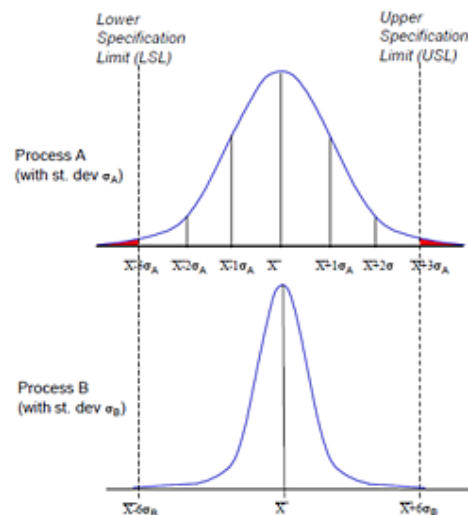
- Il 68,26% dei valori è compreso fra -1 e +1;
- Il 95,44% dei valori è compreso fra -2 e +2;
- Il 99,73% dei valori è compreso fra -3 e +3.

Elementi di Statistica

Distribuzione di Gauss

La curva che la rappresenta volge la sua concavità verso il basso nell'intervallo $(-1, 1)$ i punti di ascissa $z = \pm 1$ sono dei flessi:

- Il 68,26% dei valori è compreso fra -1 e $+1$;
- Il 95,44% dei valori è compreso fra -2 e $+2$;
- Il 99,73% dei valori è compreso fra -3 e $+3$.



$$C_p = \frac{USL - LSL}{6\hat{\sigma}}$$

$x\sigma$	C_p	$P\{\text{defect}\}$	ppm
1σ	0.33	0.317	317,000
2σ	0.67	0.0455	45,500
3σ	1.00	0.0027	2,700
4σ	1.33	0.0001	63
5σ	1.67	0.0000006	0,6
6σ	2.00	2×10^{-9}	0,00



Download from
Dreamstime.com

28848219
Nasir1154 | Dreamstime.com

Elementi di Statistica

Distribuzione di Gauss

La curva che la rappresenta volge la sua concavità verso il basso nell'intervallo $(-1, 1)$ i punti di ascissa $z = \pm 1$ sono dei flessi:

- Il 68,26% dei valori è compreso fra -1 e +1;
- Il 95,44% dei valori è compreso fra -2 e +2;
- Il 99,73% dei valori è compreso fra -3 e +3.

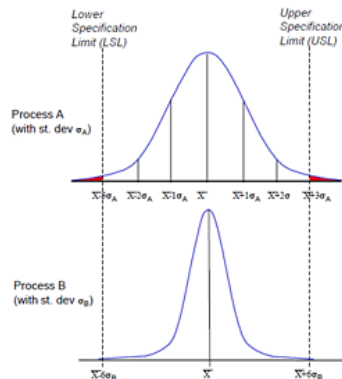
E' un programma di gestione della qualità basato sul controllo dello scarto quadratico medio che ha lo scopo di portare la qualità di un prodotto o di un servizio ad un determinato livello, introdotto per la prima volta dalla Motorola nella seconda metà degli anni '80 da Bob Galvin e Bill Smith. Si diffuse ad altre importanti compagnie, come General Electric, Toyota, Honeywell e Microsoft.

L'obiettivo del metodo è di raggiungere un tale controllo del processo da avere soltanto 3,4 parti difettose per milione:

l'obiettivo da raggiungere è quello di

avere 6 deviazioni standard tra il limite superiore di specifica ed il centro della produzione ed altrettanto tra questo ed il limite inferiore.

In altre parole, la produzione deve avere una deviazione standard non superiore ad un dodicesimo della larghezza delle specifiche.



$$C_p = \frac{USL - LSL}{6\hat{\sigma}}$$

$x\sigma$	C_p	$P(\text{defect})$	ppm
1σ	0.33	0.317	317,000
2σ	0.67	0.0455	45,500
3σ	1.00	0.0027	2,700
4σ	1.33	0.0001	63
5σ	1.67	0.0000006	0,6
6σ	2.00	2×10^{-9}	0,00



Elementi di Statistica

Distribuzione di Gauss

Esempio di approccio six sigma:

Un impiegato esce di casa tutti i giorni alle 8.00 e deve entrare al lavoro alle 8.30.

Per raggiungere l'ufficio in auto ha due possibilità: attraversare la città, o seguire un percorso di campagna, più lungo ma meno trafficato.

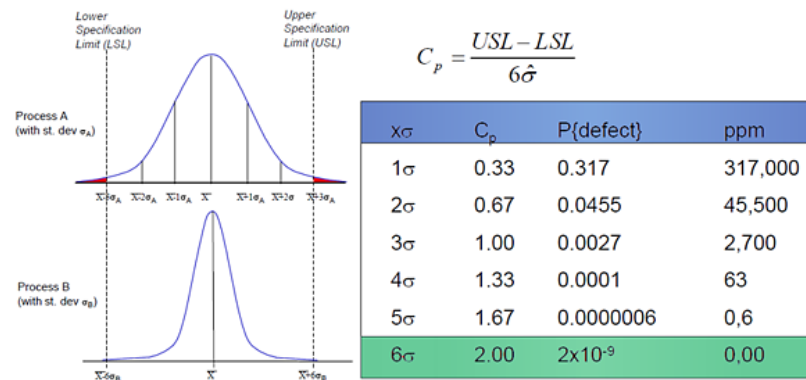
Per decidere quale sia il percorso più conveniente, misura il tempo di percorrenza più volte su entrambi i percorsi e trova che attraversando la città impiega mediamente 25 minuti, mentre per il percorso in campagna occorrono in media 28 minuti. Quale percorso gli conviene seguire?

Vecchia risposta: l'uomo dovrebbe scegliere il percorso cittadino, che in media è più veloce.

Risposta Sei Sigma: **la media non è un indicatore significativo per questo studio.**

Infatti l'impiegato è penalizzato quando arriva in ritardo, ma non ha alcun beneficio quando arriva in anticipo.

L'uomo definirebbe come difettosi i percorsi che richiedono più di 30 minuti di viaggio. Quindi si deve analizzare l'intera distribuzione dei dati nei due casi. Come si vede, il percorso cittadino presenta una forte variabilità dei dati, perché è molto influenzato (oltre che poco prevedibilmente) dal traffico; il percorso di campagna invece richiede un tempo praticamente costante. Visto l'alto numero di difetti nel caso del percorso cittadino, è evidente che quello di campagna è decisamente preferibile dal punto di vista dell'impiegato. La risposta fondata sul semplice paragone delle medie viene quindi ribaltata.



Elementi di Statistica

Distribuzione di Gauss

Esempio 1:

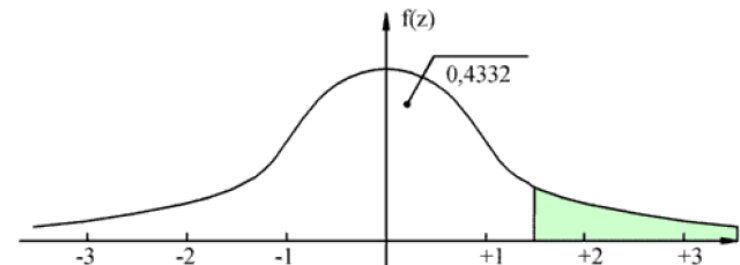
La vita media delle lavastoviglie prodotte da un'azienda è di 10 anni ($\mu=10$) con scarto quadratico medio $s=4$ anni.

Calcolare in un lotto di 500 lavastoviglie, quante:

- a) saranno in vita dopo 16 anni;
- b) avranno durata compresa fra 8 e 12 anni.

A) Al valore della variabile casuale $x=16$ anni corrisponde:

$$z = \frac{x - \mu}{\sigma} = \frac{16 - 10}{4} = \frac{6}{4} = 1,5 \quad \text{con} \quad p(z=1,5) = 0,4332$$



l'area di nostro interesse vale $p(x \geq 16) = p(z \geq 1,5) = 0,5 - 0,4332 = 0,0668$

su 500 lavastoviglie, dopo 16 anni, saranno ancora in funzione $500 \cdot 0,0668 \approx 33$ lavastoviglie.

B) Le macchine che avranno durata compresa fra 8 e 12 anni, possono essere ottenute considerando che:

$$z_1 = \frac{x_1 - \mu}{\sigma} = \frac{8 - 10}{4} = -\frac{2}{4} = -0,5$$

$$z_2 = \frac{x_2 - \mu}{\sigma} = \frac{12 - 10}{4} = \frac{2}{4} = 0,5$$

sapendo che $p(0,5) = 0,195$ avremo

$$p(8 \leq x \leq 12) = p(-0,5 \leq z \leq 0,5) = 2 \cdot 0,195 = 0,39$$

$0,39 \cdot 500 \approx 191$ lavastoviglie che avranno un tempo di funzionamento compreso fra 8 e 12 anni.

Elementi di Statistica

Distribuzione lognormale

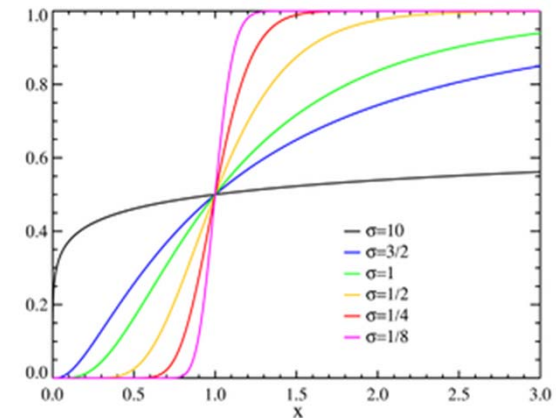
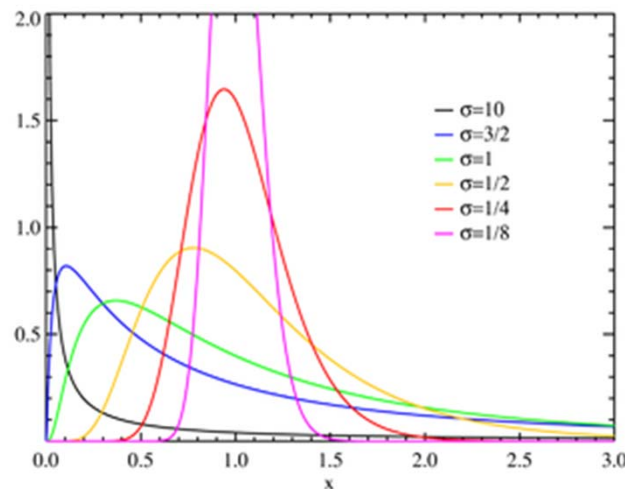
La **distribuzione lognormale** è una distribuzione di probabilità di una variabile Y il cui *logaritmo naturale* $X = \log(Y)$ segue una distribuzione normale.

E' spesso usata per descrivere la vita a fatica di componenti meccanici e in matematica finanziaria.

$$f(x) = \frac{1}{\sigma_x \sqrt{2\pi}} \cdot \exp \left[-\frac{1}{2} \left(\frac{x - \mu_x}{\sigma_x} \right)^2 \right] \quad \text{con } -\infty < x < +\infty$$

se si pone $x = \log(y)$, si ottiene la distribuzione log-normale di y

$$f(y) = \frac{1}{y \sigma_x \sqrt{2\pi}} \cdot \exp \left[-\frac{1}{2} \left(\frac{\log(y) - \mu_x}{\sigma_x} \right)^2 \right] \quad \text{con } 0 < y < +\infty$$



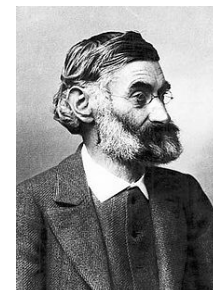
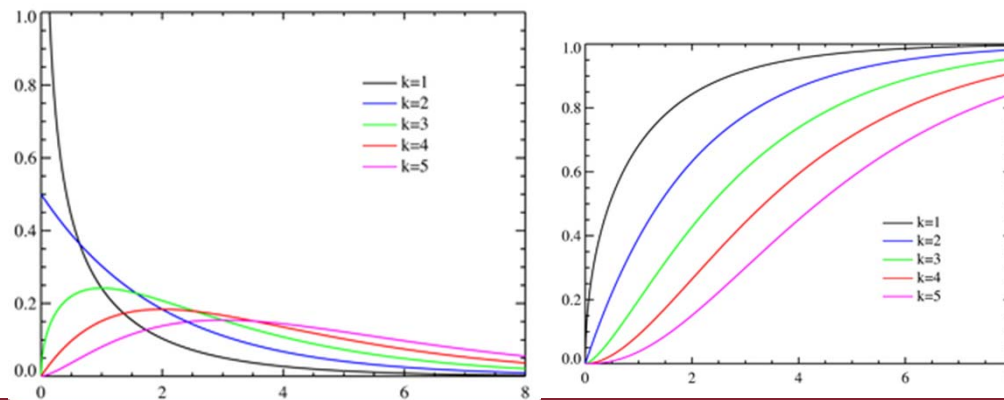
Elementi di Statistica

Distribuzione chi quadro

La **distribuzione** χ^2 è una distribuzione di probabilità che descrive la somma dei quadrati di alcune (k) variabili aleatorie **indipendenti** aventi distribuzione normale (k è detto grado di libertà).

$$x^2 = \sum_{i=1}^k x_i^2 = x_1^2 + \dots + x_k^2$$

In statistica viene particolarmente utilizzata per l'omonimo **test di verifica d'ipotesi** (test χ^2).



Ernst Karl Abbe

(Eisenach, 23 gennaio 1840 – Jena, 14 gennaio 1905)

Elementi di Statistica

Distribuzione chi quadro

Il **test di verifica d'ipotesi** si utilizza per verificare la **bontà di un'ipotesi**, ovvero un'affermazione che ha come oggetto fenomeni reali, che si presta ad essere confermata o smentita dai dati osservati sperimentalmente (**metodo sperimentale**):

- metodo deterministico, ad esempio attraverso l'uso di strumenti, etc.;
- metodo statistico
 - parametrico (media, varianza, etc.)
 - non parametrico.

Esempio

Un dado viene lanciato 2000 volte con i seguenti risultati (distribuzione uniforme con $p = 2000/6 = 333,33$):

- 1: 388 volte
- 2: 322 volte
- 3: 314 volte
- 4: 316 volte
- 5: 344 volte
- 6: 316 volte

Si può affermare che esso non è equilibrato?

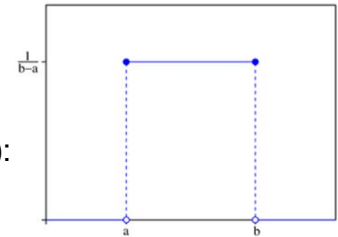
Effettivamente il risultato 1 è apparso un numero di volte sensibilmente superiore agli altri.

$$\chi^2 = \frac{(388 - 333,333)^2}{333,333} + \frac{(322 - 333,333)^2}{333,333} + \frac{(314 - 333,333)^2}{333,333} + \frac{(316 - 333,33)^2}{333,333} + \frac{(344 - 333,333)^2}{333,333} + \frac{(316 - 333,333)^2}{333,333} = 12,616$$

possiede 5 gradi di libertà = $(k - 1)$.

Se fissiamo l'errore tollerato al 5% dalle tavole della distribuzione chi quadrato con 5 gradi di libertà dobbiamo rifiutare l'ipotesi nulla con valori della statistica test superiori a 11,07.

La nostra statistica test è uguale a 12,616 pertanto dobbiamo **respingere l'ipotesi che il dado sia equilibrato**.

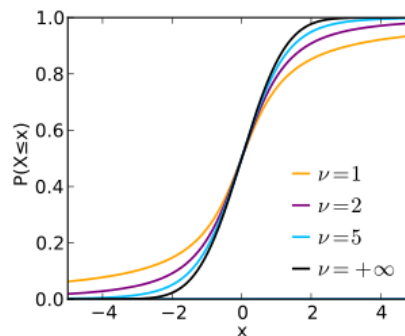
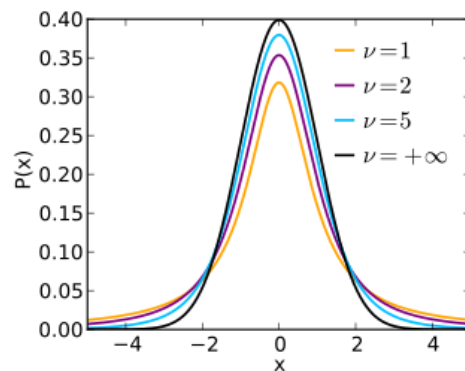


Elementi di Statistica

Distribuzione t di Student

La **distribuzione t di Student** è una distribuzione di probabilità continua che governa il rapporto tra due variabili aleatorie, la prima con **distribuzione normale** e la seconda il cui quadrato ha distribuzione **chi quadrato**.

Viene utilizzata negli omonimi test t di Student.



William Sealy Gosset
(Canterbury, 13 giugno 1876 – Beaconsfield, 16 ottobre 1937)

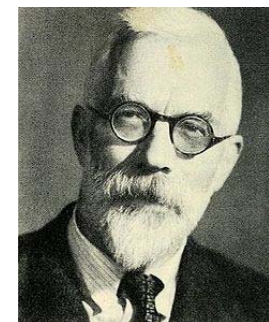
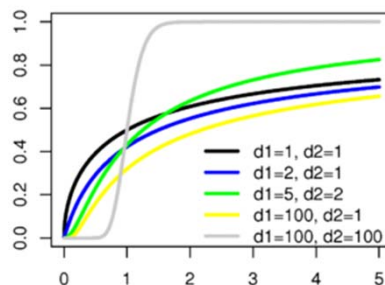
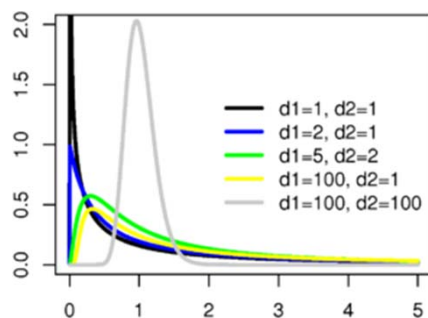


Elementi di Statistica

Distribuzione Fisher-Snedecor

La **distribuzione di Fisher-Snedecor** è una distribuzione di probabilità continua che regola il rapporto tra due **variabili aleatorie** che seguono **due distribuzioni χ^2** .

Viene utilizzata negli omonimi test F di Fischer o Z di Snedecor.

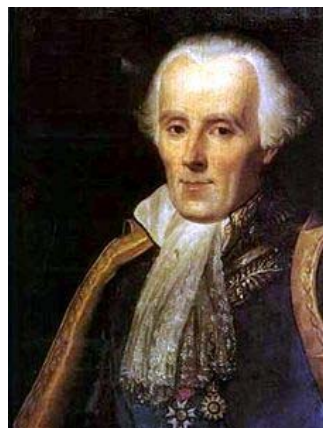


Ronald Aylmer Fisher
(Londra, 17 febbraio 1890 – Adelaide, 29 luglio 1962)

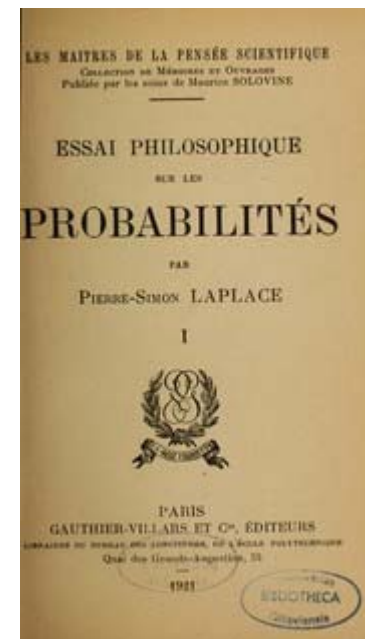
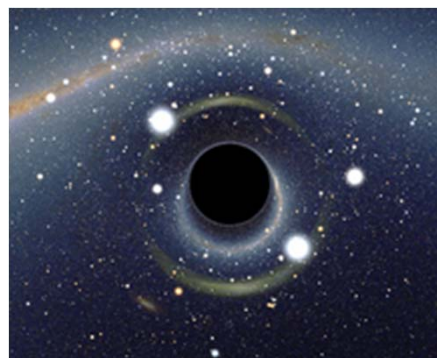
Elementi di Statistica

Distribuzione di Laplace (esponenziale)

La **distribuzione esponenziale** (o di Laplace) è una distribuzione di probabilità continua che descrive la "durata di vita" di un fenomeno/sistema che *non invecchia* (ovvero è priva di memoria).



Pierre-Simon Laplace, marchese di Laplace
(Beaumont-en-Auge, 23 marzo 1749 – Parigi, 5 marzo 1827)



Elementi di Statistica

Distribuzione di Laplace (esponenziale)

Un esempio è la *durata di vita* di una particella radioattiva prima di decadere, oppure la *durata di una richiesta di servizio* (arrivo di un cliente in una coda, l'arrivo di una telefonata).

Dunque essa è in relazione al tempo di attesa del primo successo, in fenomeni aleatori con distribuzione geometrica.

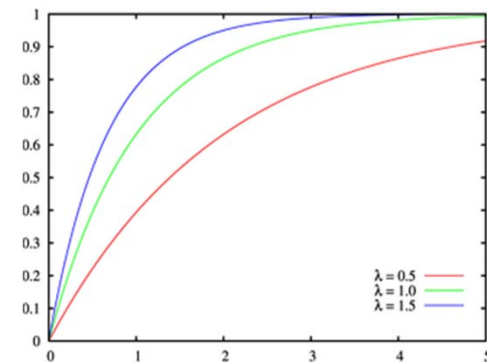
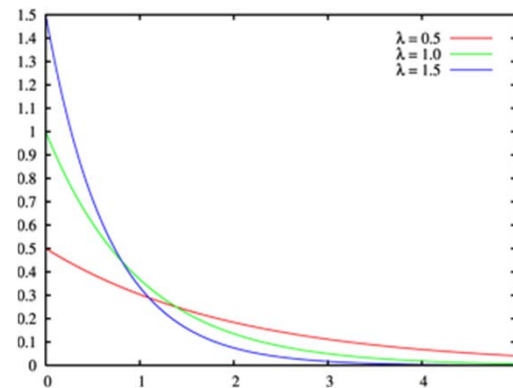
Si concretizza quando la variabile aleatoria continua x è caratterizzata dalla seguente densità di probabilità:

$$D(x) = \lambda e^{-\lambda x} \quad \text{con } \lambda > 0 \quad (25)$$

ed il valore atteso e la varianza sono date dalle seguenti espressioni:

$$\bar{x} = \frac{1}{\lambda} \quad (26)$$

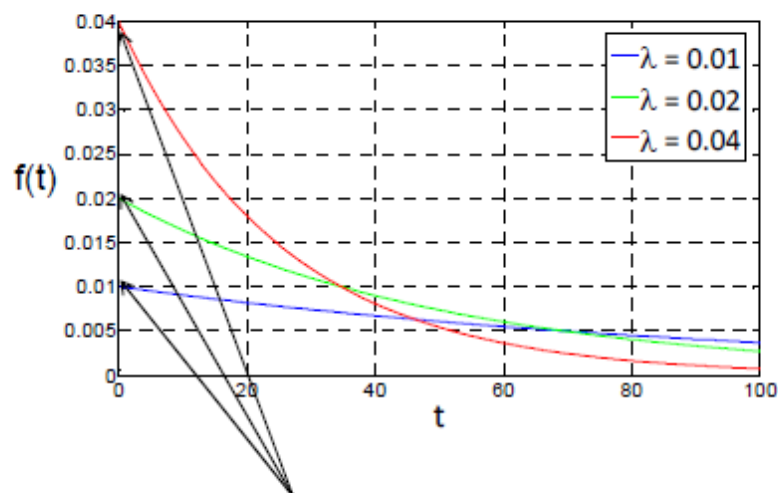
$$\sigma^2 = \frac{1}{\lambda^2} \quad (27)$$



Elementi di Statistica

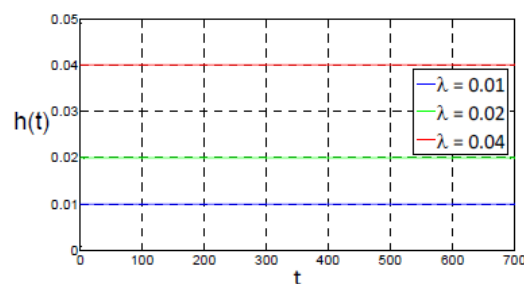
Distribuzione di Laplace (esponenziale)

Variazione della forma di *pdf* al variare di λ

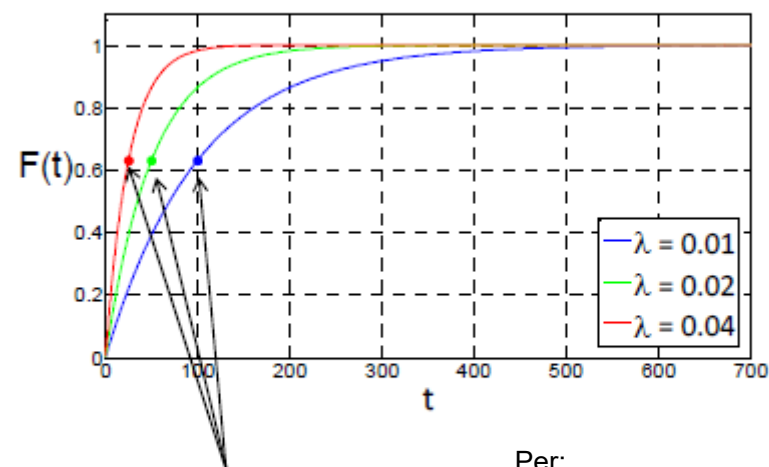


$$f(t) = \lambda \cdot e^{-\lambda t} \rightarrow f(0) = \lambda$$

Variazione della forma della funzione tasso di guasto $h(t)$ al variare di λ



Variazione della forma della *cdf* al variare di λ



$$F\left(\frac{1}{\lambda}\right) = 0.632$$

Per:

$$- t = 1/\lambda = 1/0,01 = 100$$

$$- t = 1/\lambda = 1/0,02 = 50$$

$$- t = 1/\lambda = 1/0,04 = 25$$

$1/\lambda$ = durata media

Elementi di Statistica

Distribuzione di Laplace (esponenziale)

Esempio 1:

La vita di una tipologia di lampadina è descritta da una esponenziale negativa con parametro $\lambda = 9.5 \cdot 10^{-4} \text{ h}^{-1}$. Si chiede di calcolare:

1. La probabilità di rottura tra 0 e 500 ore
2. La probabilità di rottura tra 500 e 1000 ore
3. La probabilità di rottura tra 1000 e 1500 ore
4. La probabilità di rottura tra 1500 e 2000 ore
5. L'affidabilità condizionata tra 1000 e 1500 ore
6. La probabilità che una lampadina duri meno di 800 ore
7. La probabilità che una lampadina viva per più di 1600 ore
8. Il percentile 10% e 90%

$$f(t) = \lambda \cdot e^{-\lambda \cdot t}$$

$$F(t) = 1 - e^{-\lambda \cdot t}$$

t	f	F	Diff F	
0	9.500E-04	0	-	
500	5.908E-04	0.378115	0.378115	→ 37.8 %
1000	3.674E-04	0.613259	0.235144	→ 23.5 %
1500	2.285E-04	0.759492	0.146233	→ 14.6 %
2000	1.421E-04	0.850431	0.09094	→ 9.1 %

5. L'affidabilità condizionata tra 1000 e 1500 ore

$$R(1000,500) = \frac{R(1500)}{R(1000)} = \frac{1 - F(1500)}{1 - F(1000)} = 62.2\%$$

6. La probabilità che una lampadina duri meno di 800 ore

$$F(800) = 1 - e^{-\lambda \cdot 800} = 1 - e^{-0.76} = 53.2\%$$

7. La probabilità che una lampadina viva per più di 1600 ore

$$F(1600) = 1 - e^{-\lambda \cdot 1600} = 1 - e^{-1.52} = 78.1\%$$



$$R(1600) = 1 - 0.781 = 21.9\%$$

Elementi di Statistica

Distribuzione di Laplace (esponenziale)

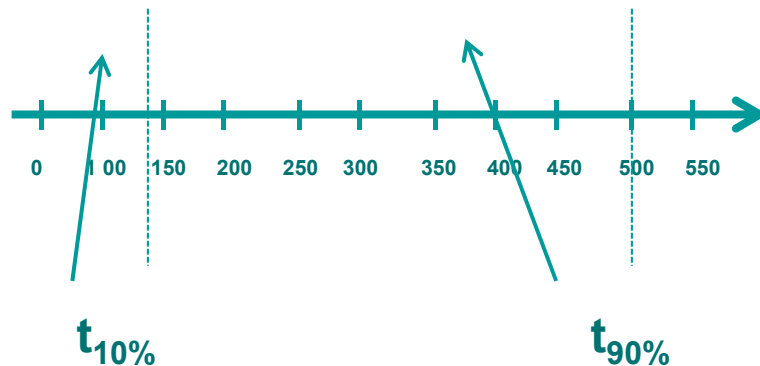
Percentile: il valore h che divide un assegnato insieme di n dati, cioè di n valori, supposti ordinati in senso non decrescente, in modo che il numero dei valori inferiori a h costituisca una data percentuale di n .

Per esempio, il primo, il secondo ecc., percentile è il valore $h_1, h_2, \dots$ tale che l'1%, il 2%, ... dei dati sia inferiore ad $h_1, h_2, \dots$

Il venticinquesimo, il cinquantesimo, il settantacinquesimo percentile si dicono rispettivamente primo, secondo e terzo quartile.

Il confronto tra i vari percentili fornisce un criterio per misurare la forma della distribuzione considerata.

8. Il percentile 10% e 90%



$$F(t) = p = 1 - e^{-\lambda t_p}$$

$$\downarrow$$
$$-\lambda t_p = \ln(1 - p)$$

$$\downarrow$$
$$t_p = -\frac{\ln(1 - p)}{\lambda}$$

Per cui si ha:

$$t_{10\%} = -\frac{\ln(1 - 0.1)}{0.00095} = 111 h$$

$$t_{90\%} = -\frac{\ln(1 - 0.9)}{0.00095} = 2424 h$$

Elementi di Statistica

Distribuzione di Laplace (esponenziale)

Esempio 2:

Sia X la durata in anni della batteria di un cellulare.

Si supponga che X segua una distribuzione esponenziale con durata media 4 anni.

a) Determinare la probabilità che una batteria di durata 2 anni duri, complessivamente, più di 4 anni.

b) Quando una batteria si guasta, viene sostituita da una nuova dello stesso tipo, che funziona indipendentemente dalla precedente.

Calcolare la probabilità che usando 2 batterie la durata complessiva sia superiore a 6 anni. Calcolare media e deviazione standard della durata complessiva.

Elementi di Statistica

Distribuzione di Laplace (esponenziale)

a) X è una variabile aleatoria di legge esponenziale di parametro $\theta = 1/4 = 0.25$

Devo determinare

$$P\{X > 4 | X > 2\} = \frac{P\{X > 4, X > 2\}}{P\{X > 2\}} = \frac{P\{X > 4\}}{P\{X > 2\}} = \frac{1 - \int_0^4 \theta e^{-\theta x} dx}{1 - \int_0^2 \theta e^{-\theta x} dx} = \frac{1 + e^{-\theta x} \Big|_0^4}{1 + e^{-\theta x} \Big|_0^2} = \frac{e^{-4\theta}}{e^{-2\theta}} = e^{-2\theta} = \frac{1}{e^{1/2}} =$$
$$= 0.607$$

b) Sia $Y = X_1 + X_2$ la variabile casuale durata complessiva, con X_1, X_2 indipendenti di legge esponenziale di parametro $\theta = 1/4$

Y è una variabile casuale di legge gamma di parametri $\alpha = 2, \theta = 1/4$.

$$\varphi_Y(x) = \begin{cases} \frac{\theta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\theta x} = \frac{1}{16\Gamma(2)} x e^{-\frac{1}{4}x} & \text{per } 0 < x < \infty \\ 0 & \text{altrove} \end{cases}$$

$$P\{Y > 6\} = 1 - P\{Y \leq 6\} = 1 - \theta^2 \int_0^6 x e^{-\theta x} dx = 1 - \theta^2 \left[-\frac{1}{\theta} x e^{-\theta x} \Big|_0^6 + \int_0^6 \frac{1}{\theta} e^{-\theta x} dx \right] =$$
$$= 1 - \theta^2 \left[-\frac{6}{\theta} e^{-6\theta} + \left(-\frac{1}{\theta^2} e^{-\theta x} \right) \Big|_0^6 \right] = 1 - \theta^2 \left[-\frac{6}{\theta} e^{-6\theta} - \frac{1}{\theta^2} e^{-6\theta} + \frac{1}{\theta^2} \right] = 1 + (6\theta + 1)e^{-6\theta} - 1 =$$
$$= (6\theta + 1)e^{-6\theta} = 2.5e^{-3/2} = 0.558$$

(Si ricordi che la funzione $\Gamma(\alpha)$ è tale che $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$ per $\alpha > 1$. Inoltre per valori di α interi ($\alpha = n$) si ha $\Gamma(n) = (n - 1)!$)

$$E(X) = \frac{\alpha}{\theta} = 8$$

$$Var(X) = \frac{\alpha}{\theta^2} = 32$$

Elementi di Statistica

Distribuzione di Laplace (esponenziale)

Esempio 3:

Una fabbrica possiede due diverse linee di produzione che producono elementi esteriormente indistinguibili, ma di qualità diversa.

I pezzi prodotti dalla linea A sono il 70% del totale e hanno un tempo di vita distribuito secondo una legge esponenziale di parametro λ .

I pezzi prodotti dalla linea B sono il rimanente 30% e hanno un tempo di vita distribuito secondo una legge esponenziale di parametro $\mu > \lambda$

a) Un pezzo viene scelto a caso e indichiamo con T il suo tempo di vita. Qual è la probabilità che $T \leq t$?

b) Calcolare il valore atteso di T .

Elementi di Statistica

Distribuzione di Laplace (esponenziale)

a)

Siano X_A, X_B i tempi di vita dei pezzi provenienti dalla linea A e dalla linea B rispettivamente.

$X_A \sim \text{Esp}(\lambda)$, $X_B \sim \text{Esp}(\mu)$

$$P\{X_A \leq t\} = F_{X_A}(t) = 1 - e^{-\lambda t}$$

$$P\{X_B \leq t\} = F_{X_B}(t) = 1 - e^{-\mu t}$$

$$\begin{aligned} P\{T \leq t\} &= P\{T \leq t | A\}P\{A\} + P\{T \leq t | B\}P\{B\} = P\{X_A \leq t\}P\{A\} + P\{X_B \leq t\}P\{B\} = \\ &= (1 - e^{-\lambda t})0.7 + (1 - e^{-\mu t})0.3 = 0.7 + 0.3 - 0.7e^{-\lambda t} - 0.3e^{-\mu t} = 1 - 0.7e^{-\lambda t} - 0.3e^{-\mu t} \end{aligned}$$

per $t \geq 0$.

b)

Bisogna determinare la densità di T. Al punto a) abbiamo calcolato la funzione di ripartizione.

Derivando si ottiene la densità:

$$\varphi_T(t) = \begin{cases} 0.7\lambda e^{-\lambda t} + 0.3\mu e^{-\mu t} & \text{per } 0 < t < \infty \\ 0 & \text{altrove} \end{cases}$$

$$E(T) = \int_{-\infty}^{\infty} t \varphi_T(t) dt = \int_0^{\infty} t (0.7\lambda e^{-\lambda t} + 0.3\mu e^{-\mu t}) dt = \frac{0.7}{\lambda} + \frac{0.3}{\mu}$$

Elementi di Statistica

Distribuzione di Weibull

La **distribuzione di Weibull** è una distribuzione di probabilità continua definita sui numeri reali positivi e descritta da due parametri (parametro di scala o vita caratteristica, λ) e (parametro di forma, k).

La distribuzione fornisce un'interpolazione tra la **distribuzione esponenziale** (per $k = 1$), la **distribuzione di Rayleigh** (per $k = 2$) e la **distribuzione normale** (per $k = 3$).

Viene impiegata per descrivere **sistemi con tasso di guasto variabile nel tempo**, **come estensione della distribuzione esponenziale** che prevede tassi di guasto costanti nel tempo.



Ernst Hjalmar Waloddi Weibull
(Vittskövle, 18 giugno 1887 – Annecy, 12 ottobre 1979)
è stato un ingegnere e statistico svedese

$$\left. \begin{aligned} f(x) &= \frac{k}{\lambda^k} x^{k-1} e^{-\left(\frac{x}{\lambda}\right)^k} \\ F(x) &= 1 - e^{-\left(\frac{x}{\lambda}\right)^k} \end{aligned} \right\} \begin{aligned} x &> 0 \\ k &> 0 \\ \lambda &> 0 \end{aligned}$$

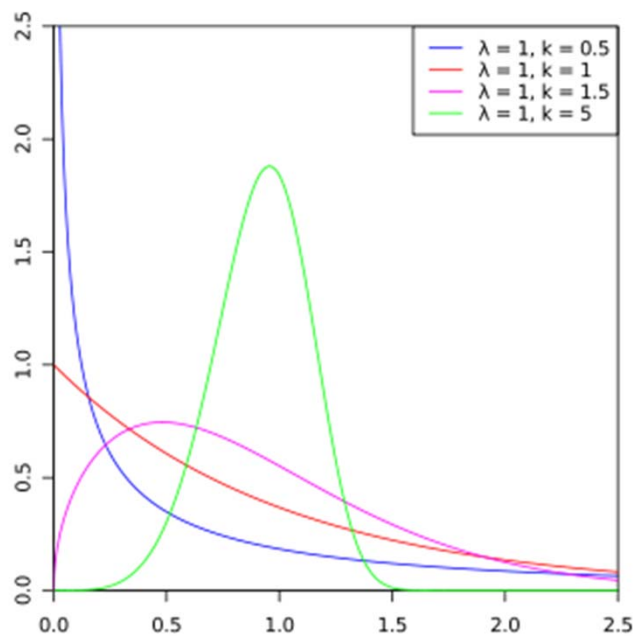
Elementi di Statistica

Distribuzione di Weibull

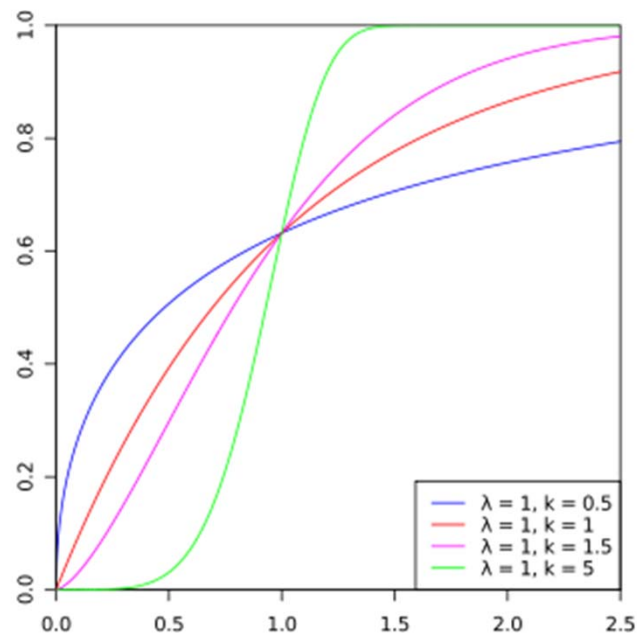
La distribuzione fornisce un'interpolazione tra la distribuzione esponenziale (per $k = 1$), la distribuzione di Rayleigh (per $k = 2$) e la distribuzione normale (per $k = 3$).

$\lambda = \text{cost.} \rightarrow$

Vita caratteristica costante $\rightarrow$ Componente costante ?



$$f(x) = \frac{k}{\lambda^k} x^{k-1} e^{-\left(\frac{x}{\lambda}\right)^k}$$

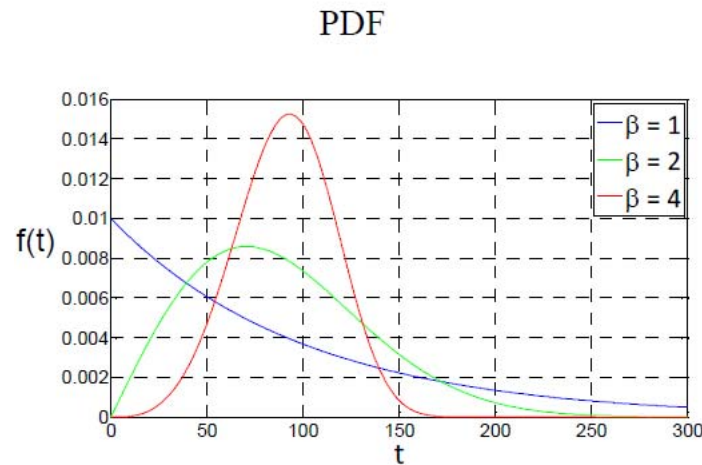


$$F(x) = 1 - e^{-\left(\frac{x}{\lambda}\right)^k}$$

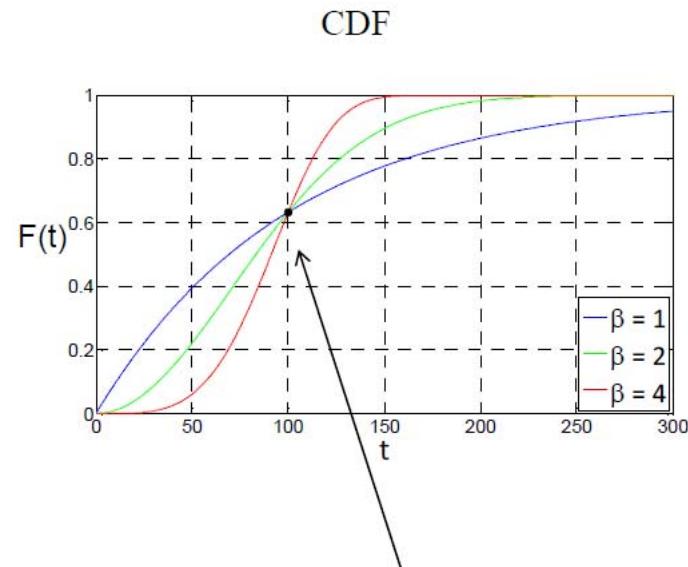
Elementi di Statistica

Distribuzione di Weibull

Variazione della forma di *pdf* e *cdf* al variare di β ($\alpha = 100$ s cost)



β è detto **parametro di forma**



Da notare che $F(\alpha) = 63.2\%$

La distribuzione di Weibull è caratterizzata anche dal cosiddetto **effetto scala**:

Se si ha un sistema costituito da n componenti identici in serie, la cui affidabilità è descritta da una Weibull di parametri α e β
allora

l'affidabilità dell'intero sistema è una Weibull, descritta ancora da una Weibull, di parametri β e $\alpha_{tot} = \alpha/n^{1/\beta}$

Elementi di Statistica

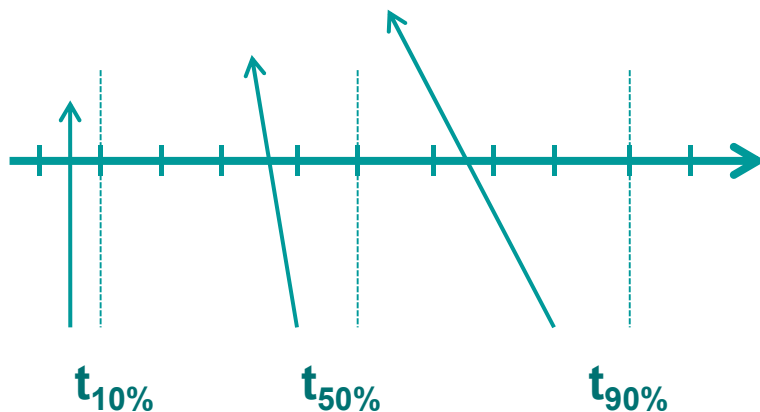
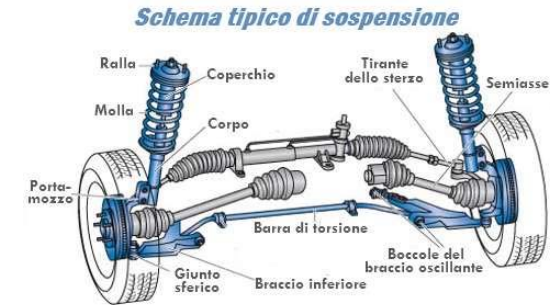
Distribuzione di Weibull

Esempio 1:

Si considerino delle molle di sospensione per auto, la cui vita a fatica è descritta da una legge di Weibull con $\alpha = 300.000$ cicli e $\beta = 2$.

Si calcolino:

1. I percentili 10%, 50% e 90%;
2. La percentuale di pezzi che si rompe prima di 100.000 cicli;
3. Il tasso di guasto a 200.000 e a 300.000 cicli;
4. Se un'auto montasse queste stesse molle su tutti e due gli assi, a quale numero di cicli corrisponderebbe la probabilità di cedimento del 10% per il sistema complessivo?



$$F(t) = 1 - \exp\left[-\left(\frac{t}{\alpha}\right)^\beta\right]$$

$$t_p = \alpha \cdot [-\ln(1-p)]^{1/\beta}$$



$$t_{10\%} = 300000 \cdot [-\ln(1-0.1)]^{1/2} = 97378$$

$$t_{50\%} = 300000 \cdot [-\ln(1-0.5)]^{1/2} = 249766$$

$$t_{90\%} = 300000 \cdot [-\ln(1-0.9)]^{1/2} = 455228$$

Elementi di Statistica

Distribuzione di Weibull

Esempio 1:

Si considerino delle molle sospensione per auto, la cui vita a fatica è descritta da una legge di Weibull con $\alpha = 300.000$ cicli e $\beta = 2$.

Si calcolino:

1. I percentili 10%, 50% e 90%;
2. **La percentuale di pezzi che si rompe prima di 100.000 cicli;**
3. Il tasso di guasto a 200.000 e a 300.000 cicli;
4. Se un'auto montasse queste stesse molle su tutti e due gli assi, a quale numero di cicli corrisponderebbe la probabilità di cedimento del 10% per il sistema complessivo?

$$F(t) = 1 - \exp\left[-\left(\frac{t}{\alpha}\right)^\beta\right]$$

$$F(100000) = 1 - \exp\left[-\left(\frac{100000}{300000}\right)^2\right] = 0.1052 = 10.52\%$$

Elementi di Statistica

Distribuzione di Weibull

Esempio 1:

Si considerino delle molle sospensione per auto, la cui vita a fatica è descritta da una legge di Weibull con $\alpha = 300.000$ cicli e $\beta = 2$.

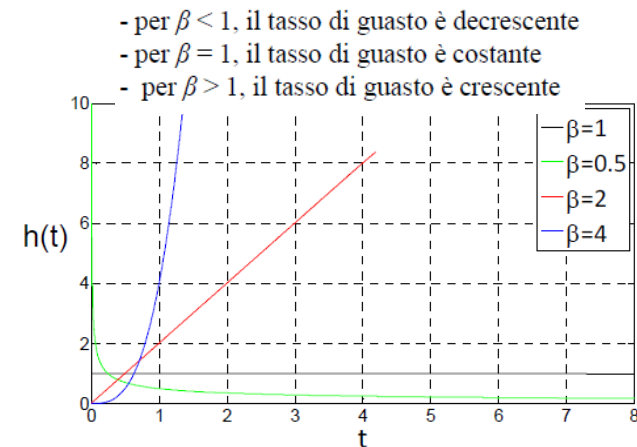
Si calcolino:

1. I percentili 10%, 50% e 90%;
2. La percentuale di pezzi che si rompe prima di 100.000 cicli;
3. **Il tasso di guasto a 200.000 e a 300.000 cicli;**
4. Se un'auto montasse queste stesse molle su tutti e due gli assi, a quale numero di cicli corrisponderebbe la probabilità di cedimento del 10% per il sistema complessivo?

$$h(t) = \frac{f(t)}{R(t)} = \frac{f(t)}{1-F(t)} = \frac{\beta}{\alpha^\beta} t^{\beta-1}$$

$$h(200000) = \frac{2}{300000^2} 200000^{2-1} = 4.44E-06 \text{ guasti/ciclo}$$

$$h(300000) = \frac{2}{300000} \left(\frac{300000}{300000} \right)^2 = \frac{2}{300000} = 6.66E-06 \text{ guasti/ciclo}$$



Elementi di Statistica

Distribuzione di Weibull

Si considerino delle molle sospensione per auto, la cui vita a fatica è descritta da una legge di Weibull con $\alpha = 300.000$ cicli e $\beta = 2$.

Si calcolino:

1. I percentili 10%, 50% e 90%;
2. La percentuale di pezzi che si rompe prima di 100.000 cicli;
3. Il tasso di guasto a 200.000 e a 300.000 cicli;
4. **Se un'auto montasse queste stesse molle (4) su tutti e due gli assi, a quale numero di cicli corrisponderebbe la probabilità di cedimento del 10% per il sistema complessivo?**

Considerando inservibile il sistema quando uno solo dei suoi componenti va fuori uso



I vari componenti possono essere ricondotti ad un unico sistema in serie

$$\alpha_{\text{tot}} = \frac{\alpha}{n^{\frac{1}{\beta}}} = \frac{300000}{4^{\frac{1}{2}}} = \frac{300000}{\sqrt{4}} = 150000 \text{ cicli}$$

$$t_{10\%} = 150000 \cdot [-\ln(1-0.1)]^{1/2} = 48689 \text{ cicli}$$