

Concetti basilari dell'analisi probabilistica

L'analisi probabilistica può essere costruita a partire da alcuni termini il cui significato è riportato qui di seguito in maniera assai semplificata, ma sufficiente a farsi una prima idea. Alla base dell'analisi probabilistica è posto il concetto di **esperimento aleatorio**, inteso come una qualsivoglia **azione** (ovvero osservazione fisica) il cui esito non possa essere stabilito a priori con certezza. Con certezza è solo possibile descrivere tutti gli **esiti** (o risultati o eventi) ammissibili dell'esperimento stesso; dunque, è naturale porsi il problema di quantificare la misura della possibilità relativa che il singolo esito o una combinazione di esiti abbiano di verificarsi. La soluzione a tale problema è appunto l'analisi probabilistica.

- **Esperimento aleatorio**

Lancio di una moneta; estrazione del numero a tombola;

Osservazione di un sistema; accesso ad una risorsa.

- **Spazio dei risultati**

Testa e croce; 1, 2, 3, ..., 88, 89, 90;

Numero di posti occupati, ..., buffer pieno / buffer vuoto;

Risorsa libera / risorsa occupata, ..., quanto si aspetta?

- **Evento**

Singolo risultato o insieme di risultati. E' una "condizione di stato" e non un "passaggio di stato", a differenza del linguaggio comune.

Ad esempio:

Numero di posti occupati o risorsa libera oppure occupata

- **Probabilità**

Quantifica la possibilità (relativa) che si realizzi un evento, tra tutti quelli individuati come possibili. Improbabile ≠ impossibile.

- **Variabile aleatoria**

Codifica uno o più eventi con numeri dell'asse reale, perché si vuole lavorare con i numeri. Introduce il concetto di durata di un evento

- **Processo stocastico**

Introduce il concetto di tempo come contenitore di eventi che hanno una certa durata.

Modella la ripetizione nel tempo di un esperimento aleatorio e, dunque, riproduce la dinamica degli stati di un sistema.

I passi dell'analisi probabilistica

1. IDENTIFICAZIONE dello "spazio" dei risultati elementari possibili;
2. ASSEGNAZIONE delle "probabilità" ai risultati possibili;
3. DEFINIZIONE degli "eventi" di interesse;
4. CALCOLO delle "probabilità di eventi" (di interesse)

dove:

- un risultato "elementare", ovvero un evento elementare, è tale perché è direttamente osservabile ed è facile assegnargli una probabilità; (lancio una moneta) (quante volte esce testa se lancio la moneta)
- un evento o risultato "d'interesse" corrisponde, in generale, ad un sottoinsieme dei risultati elementari, cioè alla combinazione logica di eventi elementari;
- l'evento di interesse deve essere posto in relazione agli eventi elementari mediante l'ALGEBRA DEGLI EVENTI (che sarà introdotta nel seguito).

La probabilità di un evento di interesse, si calcola sommando la probabilità degli eventi elementari che lo compongono.

ESEMPIO: il docente del corso si avvicina all'ultima fila di studenti in aula per osservare se "dormono più della metà". Il risultato elementare corrisponde allo stato di sonno del singolo studente, mentre l'esito "dormono più della metà" è un risultato d'interesse che non è elementare.

L'assegnamento dei valori di probabilità P agli eventi elementari deve rispettare i seguenti **assiomi**:

- (A1) per ogni evento $A \in \Omega$ (*Spazio degli eventi*) si ha $P(A) > 0$; prob non è nulla
- (A2) $P(\Omega) = 1$; per noi tutta la probabilità è pari a 1
- (A3) $P(A \cup B) = P(A) + P(B)$.

La generalizzazione di detti assiomi ad eventi non elementari è la seguente

(A1') per ogni evento $A \in \Omega$ si ha $P(A) \geq 0$;

(A2') $P(\Omega) = 1$;

(A3') $P(A \cup B) = P(A) + P(B)$ nell'ipotesi che $A \cap B = \emptyset$ (eventi disgiunti). Si vedrà tra molto il perché

Il principio di induzione consente di estendere l'assioma (A3') al caso dell'unione di "n" eventi disgiunti:

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = \sum_{i=1}^n P(A_i).$$

Ma per trattare uno spazio di risultati numerabili all'infinito occorre sostituire l'assioma (A3') con il seguente:

$$(A3'') \quad P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i) \leq 1$$

In tal modo, se l'unione all'infinito degli eventi A_i produce un altro evento allora il valore (compreso fra 0 e 1) a cui converge la serie delle $P(A_i)$ rappresenta la probabilità di quell'evento.

Esempio sull' unione all'infinito di eventi

Sia A_i l'evento seguente: "i-1 oggetti sono presenti in un buffer", con $i-1 = 0, 1, 2, \dots, n, \dots$

E si assegnino le seguenti:

$$P(A_i) \hat{=} (1-\rho) \cdot \rho^{i-1}, \quad : 0 < \rho < 1, \quad i = 1, 2, \dots, n, \dots$$

Allora:

$$\begin{aligned} P(\bigcup_{i=1}^{\infty} A_i) &= \sum_{i=1}^{\infty} P(A_i) \\ &= \sum_{i=1}^{\infty} (1-\rho) \rho^{i-1} = (1-\rho) \cdot \sum_{j=0}^{\infty} \rho^j \quad \text{La serie geometrica converge} \\ &= (1-\rho) \cdot \frac{1}{1-\rho} = 1 \quad \text{a } 1/(1-p) \quad \text{infatti: } \bigcup_{i=1}^{\infty} A_i \equiv \Omega \end{aligned}$$

Serie
geometrica!

8

Continua esempio sull' unione all'infinito di eventi

Si consideri adesso la seguente:

$$P(\bigcup_{i=n}^{\infty} A_i), \quad n \geq 2$$

Allora:

$$\begin{aligned} &= \sum_{i=n}^{\infty} (1-\rho) \rho^{i-1} = \sum_{j=0}^{\infty} (1-\rho) \cdot \rho^{n-1} \cdot \rho^j \quad (j \hat{=} i-n) \\ &= (1-\rho) \cdot \frac{1}{1-\rho} \cdot \rho^{n-1} = \rho^{n-1} < 1 \end{aligned}$$

$$\text{infatti: } \bigcup_{i=n}^{\infty} A_i \subset \Omega \quad \text{se } n \geq 2$$

e corrisponde all'evento: "almeno $n-1$ oggetti nel buffer"

9

Esempio sui passi dell'AP

Lo studente Giuseppe Rossi deve sostenere l'esame del corso "Analisi Probabilistica". L'esame è composto da due prove in sequenza, senza il meccanismo di ammissione alla seconda: finita la prima prova si passa sempre alla seconda. Il docente valuta le due prove insieme, dunque solo alla fine della seconda. Nella valutazione di ciascuna prova, il docente attribuisce un punteggio da 1 a 5 e lo studente, dal canto suo, **deve superare entrambe le prove con almeno 2 punti per essere promosso.** (quindi ho un vincolo)

Invece, per il prossimo appello d'esame, il docente ha deciso:

- di assegnare il punteggio di ciascuna prova (1..5) in **maniera completamente casuale;**
- di promuovere lo studente se viene fuori un punteggio cumulativo di **almeno 6** punti sull'insieme delle due prove.

Calcolare la probabilità che Rossi venga promosso:

- proprio col minimo punteggio (6); quindi calcolo le combinazioni che danno 6
- proprio col massimo punteggio (10);
- **o col minimo punteggio (6) o col massimo punteggio (10).** da notare l'OR logico

Svolgimento

Passo 1)

Identificazione dello spazio dei risultati possibili Ω

Lo spazio dei risultati Ω è dato da tutte le possibili coppie (i, j) con i il valore di punteggio ottenuto al primo compito $i = 1..5$ e j il valore di punteggio ottenuto al secondo compito $j = 1..5$. Quindi:

$$\Omega = \left\{ \begin{matrix} (1,1) & (1,2) & (1,3) & (1,4) & (1,5) \\ (2,1) & (2,2) & (2,3) & (2,4) & (2,5) \\ (3,1) & (3,2) & (3,3) & (3,4) & (3,5) \\ (4,1) & (4,2) & (4,3) & (4,4) & (4,5) \\ (5,1) & (5,2) & (5,3) & (5,4) & (5,5) \end{matrix} \right\}.$$

Passo 2)

Assegnazione delle probabilità ai risultati possibili

Poiché il docente ha deciso di assegnare il punteggio di ciascuna prova in maniera completamente casuale, ogni evento elementare (i, j) è equiprobabile per cui

$$P(i, j) = \frac{1}{25} \quad \forall (i, j).$$

Passo 3) Definizione degli eventi di interesse

Stabilito che la promozione si ottiene conseguendo un punteggio complessivo sulle due prove di valore pari ad almeno 6, gli eventi di interesse si riferiscono alla promozione dello studente Rossi proprio col minimo punteggio (6), proprio col massimo punteggio (10) o col minimo punteggio (6) o col massimo punteggio (10).

Risoluzione) Calcolo delle probabilità di eventi di interesse

Definiamo i seguenti eventi:

A $\hat{=}$ promozione con un valore di punteggio pari proprio al minimo (6)

B $\hat{=}$ promozione con un valore di punteggio pari proprio al massimo (10)

C $\hat{=}$ promozione con un valore di punteggio pari al minimo (6) o al massimo (10)

allora

$$P(A) = \frac{3}{25} = 0,12 \text{ (i.e. 3 casi su 25 } \{(2,4),(3,3),(4,2)\}\text{)}$$

$$P(B) = \frac{1}{25} = 0,04 \text{ (i.e. 1 caso su 25 } \{(5,5)\}\text{)}$$

$$P(C) = P(A) + P(B) = \frac{4}{25} = 0,16.$$

Evento limite e sua probabilità

Data una sequenza di eventi $\{E_n, n \geq 1\}$ si dirà che essa è una **sequenza crescente** se risulta $E_1 \subset E_2 \subset \dots \subset E_n \subset E_{n+1} \subset \dots$. Se, invece, risulta $E_1 \supset E_2 \supset \dots \supset E_n \supset E_{n+1} \supset \dots$ allora si dirà che la **sequenza è decrescente**.

E1 sarebbe superset ovvero insieme che contiene E2.

Facendo riferimento a una sequenza crescente si definisce **evento limite** e si indica con $\supset E$ il seguente:

$$\supset E \hat{=} \lim_{n \rightarrow \infty} E_n = \lim_{n \rightarrow \infty} \bigcap_{i=1}^n E_i$$

Viceversa, facendo riferimento a una sequenza decrescente si definisce analogamente:

$$\subset E \hat{=} \lim_{n \rightarrow \infty} E_n = \lim_{n \rightarrow \infty} \bigcup_{i=1}^n E_i$$

Proposizione:

Se $\{E_n, n \geq 1\}$ è una sequenza di eventi crescente o decrescente e si indica semplicemente con E l'evento limite, allora risulta:

$$P(E) \hat{=} P\left(\lim_{n \rightarrow \infty} E_n\right) = \lim_{n \rightarrow \infty} P(E_n)$$

si vedrà in futuro cosa vuol dire

PROVA (per una sequenza crescente):

$$\begin{aligned} P(E) &\hat{=} P\left(\lim_{n \rightarrow \infty} E_n\right) = P\left(\lim_{n \rightarrow \infty} \bigcup_{i=1}^{n-1} E_i\right) = P\left(\bigcup_{i=1}^{\infty} E_i\right) = P\left(E_1 \cup \bigcup_{i=1}^{\infty} E_{i+1} \setminus E_i\right) \\ &= P(E_1) + \sum_{i=1}^{\infty} [P(E_{i+1}) - P(E_i)] = P(E_1) + \lim_{n \rightarrow \infty} \sum_{i=1}^n [P(E_{i+1}) - P(E_i)] = \lim_{n \rightarrow \infty} P(E_n) \end{aligned}$$

La proposizione formalizza un concetto di continuità dell'assegnamento probabilistico che sarà utile in seguito, quando si lavorerà con funzioni continue che effettuano l'assegnamento probabilistico sulla semiretta reale e si porrà il problema di attribuire una probabilità tanto ad un intervallo di reali quanto ad un unico punto. Allora, l'intervallo e il punto saranno, rispettivamente, l'evento e l'evento limite.

Esempio su evento limite e sua probabilità

Sia E_i l'evento seguente:

“ al più i oggetti sono presenti nel buffer”,
con $i = 1, 2, 3, \dots$ e quindi E_1, E_2, \dots una sequenza crescente.

Poiché:

$$P(E_i) = 1 - \rho^i, \quad : 0 < \rho < 1, \quad i = 1, 2, \dots \quad \text{come visto in precedenza}$$

Allora:

$$\begin{aligned} P(E) &= P(\bigcup_{i=1}^{\infty} E_i) = P(E_1) + \sum_{i=1}^{\infty} [P(E_{i+1}) - P(E_i)] \\ &= (1 - \rho) + \sum_{i=1}^{\infty} [(1 - \rho^{i+1}) - (1 - \rho^i)] \\ &= \lim_{n \rightarrow \infty} P(E_n) = \lim_{n \rightarrow \infty} (1 - \rho^n) = 1 \end{aligned}$$

11

Costruzione di un modello di assegnazione

$$P : [0,1,2,\dots,n] \rightarrow [0,1] \in \Re$$

$$\sum_{n=0}^{\infty} \frac{x^n}{n!} = e^x, \quad x > 0 \quad (\text{MODELLO DI PARTENZA})$$

$$\sum_{n=0}^{\infty} \frac{x^n}{n!} e^{-x} = 1, \quad x > 0 \quad (\text{PASSAGGIO INTERMEDI})$$

$$0 \leq \frac{x^n}{n!} e^{-x} \leq 1, \quad n = 0, 1, 2, \dots \quad (\text{VERIFICA FINALE})$$

$P(N = n) = \frac{x^n}{n!} e^{-x}, \quad n = 0, 1, 2, \dots$	MODELLO DI POISSON
--	---------------------------

12

Il modello di Poisson è un modello che nasce dalla teoria che vedremo molto più in la nel corso. Nel mondo reale si utilizza per le chiamate in un call center, cioè o la chiamata ha vita breve e dura pochissimi secondi, oppure la chiamata dura un tempo ragionevolmente lungo (non si sta in chiamata per sempre, cioè all'infinito)

11

Problemi combinatori

Quando lo spazio dei risultati è costituito da un numero finito di eventi elementari equiprobabili, la probabilità di un "tipo di risultato" (evento di interesse) si calcola come rapporto tra il numero di risultati elementari che definiscono quel evento d'interesse e il numero di risultati (eventi) elementari che definiscono tutto lo spazio dei risultati.

Calcolare la probabilità di un evento d'interesse è un problema combinatorio (beccare il 6 al Superenalotto, ad esempio) che può essere risolto a partire dai seguenti calcoli, basilari, di numerosità:

- La numerosità del campione ordinato (disposizione) di "k" risultati estratti da uno spazio di "n", col meccanismo del rimpiazzo dell'oggetto estratto è pari a:

$$\underbrace{n \cdot n \cdots n}_{k \text{ volte}} = n^k \quad \begin{array}{l} \text{esperimento con} \\ \text{risultati ripetibili} \end{array}$$

Esempio

Si consideri l'esperimento "lancio di una moneta" per il quale il possibile spazio dei risultati è $\Omega = \{T, C\}$ ($n = 2$). Posto che il rimpiazzo è automatico per la natura dell'esperimento, se il lancio della moneta viene effettuato tre volte ($k = 3$), allora le possibili disposizioni sono:

T, T, T

T, T, C

T, C, T

C, T, T ovvero $n^k = 2^3 = 8$

C, C, T

C, T, C

T, C, C

C, C, C

- La numerosità del campione ordinato (disposizione) di "k" risultati estratti da uno spazio di "n", senza rimpiazzo è pari a:

$$n(n-1)(n-2) \cdots (n-(k-1)) = \frac{n!}{(n-k)!} \quad \begin{array}{l} \text{esperimento} \\ \text{risultati non ripetibili} \end{array}$$

Fissando $k \hat{=} n$ (permutazione) si ottiene: $n!$ quindi estraggo tutti gli elementi
Esempio

Per ricoprire 2 posizioni in un'azienda ci sono 6 concorrenti. Dal momento che tutti i concorrenti hanno lo stesso bagaglio di esperienze, si decide di estrarre a caso i due nominativi dei futuri promossi. Ovviamente se un concorrente viene estratto, non può essere candidato anche per l'altra posizione. Quindi, le possibili coppie estraibili (disposizioni senza rimpiazzo) sono:

$$\frac{n!}{(n-k)!} = \frac{6!}{(6-2)!} = 30 \quad \begin{array}{l} \text{se fossero 6 posizioni allora} \\ \text{avrei 720 permutazioni possibili} \end{array}$$

- Campione non ordinato (combinazione) di "k" risultati estratti da uno spazio di "n", senza il meccanismo del rimpiazzo.

$$\frac{n!}{k!(n-k)!} \hat{=} \binom{n}{k}.$$

Nel Superenalotto la combinazione vincente del "6" si ottiene dal primo numero estratto su 6 ruote diverse. Il meccanismo di estrazione garantisce che uno stesso numero non possa essere estratto come primo numero su più di una delle 6 ruote. In altre parole, si tratta di un meccanismo di estrazione casuale senza rimpiazzo.

Dunque, si possono estrarre 90 numeri sulla prima ruota, 89 sulla seconda, 88 sulla terza, 87 sulla quarta, 86 sulla quinta e 85 sulla sesta, ovvero:

$$n \cdot (n-1) \cdot (n-2) \cdot (n-3) \cdot (n-4) \cdot (n-(k+1)) = \frac{n!}{(n-k)!}$$

Se si pensa alla sestina vincente, l'ordine con cui vengono estratti i numeri non è importante, quindi, quella sestina può uscire in $k!$ modi diversi, ovvero la numerosità si riduce di $k!$ risultando:

$$\frac{n!}{k!(n-k)!} = \binom{90}{6}.$$

mentre la probabilità aumenta di $k!$, visto che è pari al reciproco della numerosità.

Il precedente è noto come *coefficiente binomiale* (e si legge n su k).

Ipotizzando di avere 1024 thread resi disponibili dalla CPU, separati per tipologie ad esempio thread I/O, thread di rete, thread di calcolo e thread di scrittura. Supponiamo che un processo per essere completato debba avere un thread di ogni tipo, perciò 4. Allora si ha che un processo è completato con coefficiente binomiale 1024 su 4

Il problema dei compleanni

E' dato un gruppo di «k» persone aggregate a caso e si vuole calcolare la probabilità che almeno due di esse siano nate nello stesso giorno di anni sia pur differenti.

Soluzione:

(calcoliamo la probabilità dell'evento d'interesse come il complemento ad 1 della probabilità dell'evento complementare a quello d'interesse)

$$= 1 - \frac{n(n-1)(n-2) \cdots (n-(k-1))}{n^k} \quad \text{con } n = 365$$

In particolare,

con k=23 risulta = 0,5073 e con k=50 risulta = 0,9704

4

Quesito finale

Quanto vale la probabilità (P) che uno qualsivoglia di «n» risultati (elementari ed equiprobabili) si realizzi proprio «h» volte, in corrispondenza della ripetizione dell'esperimento aleatorio «k» volte?

(ovviamente deve essere $k \geq h$)

$$P = \frac{\binom{k}{h} (n-1)^{k-h}}{n^k} \quad h = 0, 1, \dots, k$$

5

Piccolo schema per ricordarsi:

- campione ordinato -> disposizione (con rimpiazzo: esponenziale, senza: rapporto di fattoriali)
- campione non ordinato -> combinazione (coefficiente binomiale)

Una applicazione del calcolo combinatorio

Sia S un insieme di n elementi, di cui d difettosi ($d \leq n$). Si vuole calcolare la probabilità di ottenere r elementi difettosi, estraendo k ($k \leq n$) elementi a caso e senza rimpiazzo dall'insieme S .

Il numero di possibili campioni casuali di k elementi, estratti da un insieme di n senza il rimpiazzo è:

$$\binom{n}{k} \quad \text{appunto coefficiente binomiale}$$

Analogamente, i campioni di r elementi difettosi che possono capitare, dai d elementi difettosi di S , ammontano a:

$$\binom{d}{r}$$

Infine, il numero di campioni non difettosi di $k-r$ elementi, appartenenti al campione dei k elementi estratti ammonta a:

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

(Si noti che i $k-r$ elementi appartengono all'insieme degli $n-d$ elementi non difettosi in S .)

Poiché gli eventi elementari sono equiprobabili, è possibile calcolare la probabilità di ottenere r elementi difettosi dai k elementi estratti come:

prodotto tra elementi difettosi e non difettosi in rapporto ad elementi generici $\longrightarrow P(r | k, d, n) = \frac{\binom{d}{r} \binom{n-d}{k-r}}{\binom{n}{k}}$ Ipergeometrica!

La precedente probabilità è nota come *probabilità Ipergeometrica*. Un'ulteriore considerazione riguarda il numero r di elementi difettosi dal campione estratto di k elementi:

$$\max\{0, d + k - n\} \leq r \leq \min\{d, k\}.$$

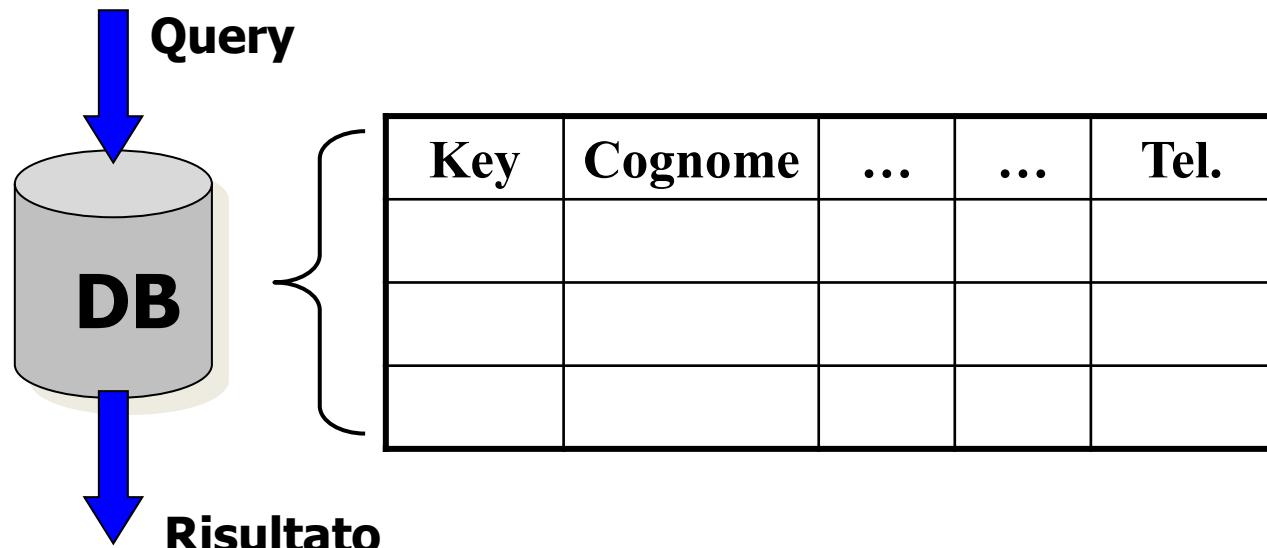
Esempio sulla probabilità ipergeometrica

Una tabella di database contiene informazioni di 10 persone, tra cui i rispettivi numeri di telefono. Però 4 numeri sono inconsistenti.

Se viene estratto un campione di 6 numeri telefonici, qual è la probabilità che 2 dei 6 numeri telefonici estratti risultino inconsistenti?

RISPOSTA:

$$\begin{aligned} P(2|6,4,10) \\ = \binom{4}{2} \cdot \binom{6}{4} / \binom{10}{6} \\ = 0.43 \end{aligned}$$



Algebra degli eventi

È completamente definita dalle seguenti 5 leggi:

Legge Comutativa: $A \cup B = B \cup A$ $A \cap B = B \cap A$

Legge Associativa: $A \cup (B \cup C) = (A \cup B) \cup C$ $A \cap (B \cap C) = (A \cap B) \cap C$

Legge Distributiva: $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

Legge dell'Identità: $A \cup \emptyset = A$ $A \cap \Omega = A$

Legge del Complemento: $A \cup \bar{A} = \Omega$ $A \cap \bar{A} = \emptyset$

Ogni relazione fra eventi di interesse ed eventi elementari può essere stabilita usando opportunamente le leggi appena esposte.

∩ intersezione negli insiemi, e (congiunzione appunto probabilità congiunta) nelle probabilità

Relazioni valide:

1. $A \cup A = A$; $A \cap A = A$; $A \cup \Omega = \Omega$; $A \cap \emptyset = \emptyset$;
2. $A \cap (A \cup B) = A$; $A \cup (A \cap B) = A$;
3. $\overline{A \cup B} = \bar{A} \cap \bar{B}$; $\overline{A \cap B} = \bar{A} \cup \bar{B}$;
4. $\overline{\overline{A}} = A$; $A \cup (\bar{A} \cap B) = A \cup B$.

A titolo di esempio:

$$(A \cap B) \cup (\bar{A} \cap B) = B$$

$$(A \cap B) \cup (\bar{A} \cap B)$$

$$(B \cap A) \cup (B \cap \bar{A}) \quad \text{(legge commutativa)}$$

$$B \cap (A \cup \bar{A}) \quad \text{(legge distributiva)}$$

$$B \cap \Omega \quad \text{(legge del complemento)}$$

$$B \quad \text{(legge dell'identità)}$$

Vedremo nelle prossime pagine come:

$$P(A \cap B) = P(A) \cdot P(B) \quad \text{se e solo se } A \text{ e } B \text{ sono indipendenti}$$

$$P(A \cup B) = P(A) + P(B) \quad \text{se } A \text{ e } B \text{ sono eventi disgiunti (intersezione nulla, NON POSSONO accadere contemporaneamente!)}$$

Esempio evento congiunto: da un sacco di palline colorate e numerate estraggo una pallina rossa e pari (evento congiunto perchè può accadere contemporaneamente)

Arrivano datagrammi dal protocollo IP, il datagramma che prende il sistema è quello id 1 e destinazione proprio il sistema.

Esempi eventi disgiunti: in un database vado a leggere mentre ho in corso una scrittura con lock esclusivo, non può accadere! Tiro un dado o esce un numero pari o dispari entrambi no. Tiro una moneta non può uscire testa e croce contemporaneamente. Devo allocare un processo all'I/O non posso allocarne due contemporaneamente.

Diagrammi di Venn

Un tipo di rappresentazione grafica degli eventi, molto utile per illustrare le leggi che definiscono l'algebra degli eventi, sono i diagrammi di Venn. Lo spazio degli esiti Ω viene rappresentato da un grande rettangolo. Gli eventi da prendere in considerazione, invece, sono rappresentati da cerchi o altre curve chiuse disegnate all'interno del rettangolo. A questo punto, tutti gli eventi non elementari di interesse possono essere evidenziati colorando opportune regioni del diagramma. Ad esempio nei tre diagrammi di Venn illustrati nella figura a seguire, le regioni scurite rappresentano, nell'ordine, gli eventi $A \cup B$, $A \cap B$ ed \bar{A} . Il diagramma di Venn della seconda figura, invece, mostra che $B \subset A$.

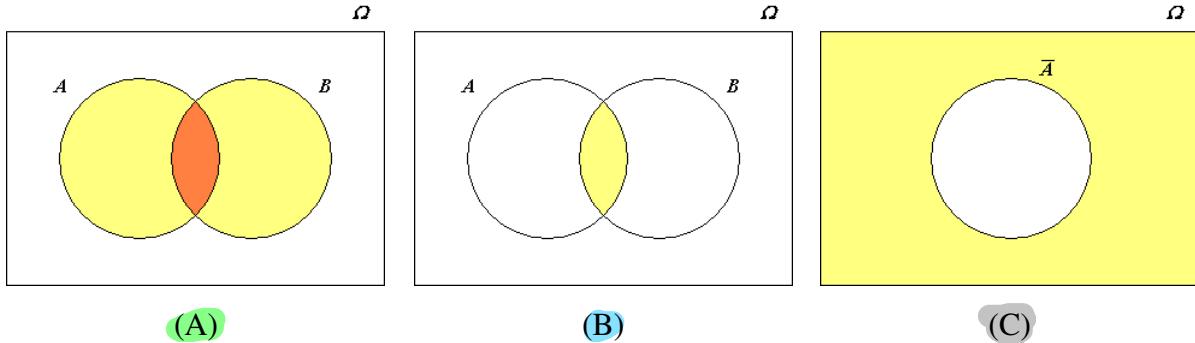


Figura – Diagrammi di Venn per l'unione, l'intersezione e il complemento

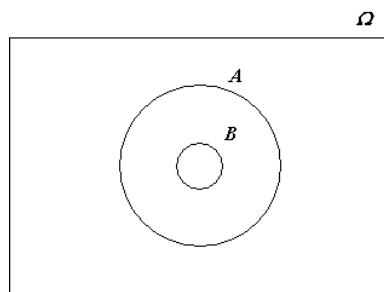


Figura – Diagramma di Venn che illustra la relazione $B \subset A$

Nella terza figura i diagrammi di Venn sono impiegati per verificare la seconda delle due proprietà distributive: $(C \cap A) \cup (C \cap B) = C \cap (A \cup B)$.

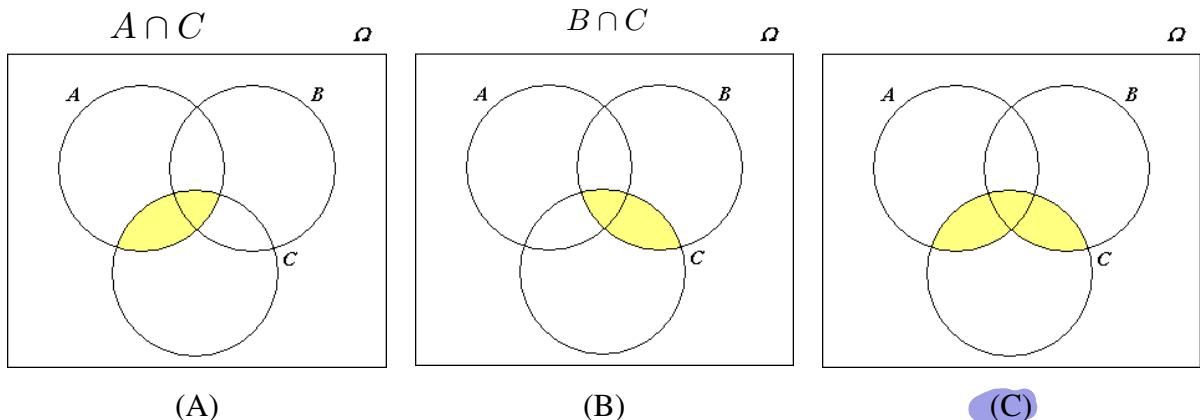


Figura – Illustrazione di una proprietà distributiva per mezzo dei diagrammi di Venn

Per esercizio, si propone di verificare mediante l'uso dei diagrammi di Venn le seguenti relazioni note come ***Leggi di De Morgan***:

$(A \cup B)^c = \bar{A} \cap \bar{B}$ risultato tutto lo spazio del rettangolo esclusi A e B

$(A \cap B) = \overline{A} \cup \overline{B}$ risultato solo l'intersezione tra A e B

DEFINIZIONI

- Gli eventi A_1, \dots, A_n sono detti **mutuamente esclusivi** se e solo se risulta qualsiasi coppia ha intersezione nulla*
$$A_i \cap A_j = \begin{cases} A_i & \text{se } i = j \\ \emptyset & \text{altrimenti} \end{cases} \quad (i = 1 \dots n, j = 1 \dots n)$$
 - Gli eventi A_1, \dots, A_n sono detti **collettivamente esaustivi** se e solo se risulta:*
$$A_1 \cup A_2 \cup \dots \cup A_n = \Omega$$
 - Gli eventi A_1, \dots, A_n costituiscono una **partizione di Ω** se e solo se risultano mutuamente esclusivi e collettivamente esaustivi.*

Probabilità dell'unione di 2 eventi:

Di seguito l'unione della probabilità tra 2 eventi ed il suo caso generale con n elementi

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

Probabilità dell'unione di eventi

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

(DUE EVENTI)

Prova:

Sia $A \cup B = A \cup (\bar{A} \cap B)$ e $B = (A \cap B) \cup (\bar{A} \cap B)$ sostituendo $P(B)$

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

da cui ottengo che

$$P(A \cap B) \geq P(A) + P(B) - 1$$

$$P(A \cap B) \geq P(A) + P(B) - 1$$

$$P(\bar{A} \cap B) = P(B) - P(A \cap B)$$

Disuguglianaza di Bonferroni

$$P\left(\bigcup_{i=1}^n A_i\right) = P(A_1 \cup A_2 \cup \dots \cup A_n) =$$

probabilità dell'unione di N eventi
NON disgiunti

(n EVENTI)

$$= \sum_{\forall i} P(A_i) - \sum_{1 \leq i < j \leq n} P(A_i \cap A_j) +$$

$$+ \sum_{1 \leq i < j < k \leq n} P(A_i \cap A_j \cap A_k) + \dots + (-1)^{n-1} P(A_1 \cap A_2 \cap \dots \cap A_n)$$

La probabilità dell'unione di n eventi è minore uguale alla sommatoria delle probabilità di n elementi

$$P\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n P(A_i)$$

Disuguaglianza di Boole

Disuguaglianze di Bonferroni

Nel caso di collezioni finite di eventi, la precedente disuguagliazione può venire generalizzata nelle cosiddette **disuguaglianze di Bonferroni** le quali forniscono estremi superiori e inferiori alla probabilità per l'unione di tali eventi.

Introduciamo le seguenti quantità:

$$S_1 := \sum_{i=1}^n P(A_i) ,$$

$$S_2 := \sum_{i < j} P(A_i \cap A_j) ,$$

e per $2 < k \leq n$,

$$S_k := \sum P(A_{i_1} \cap \cdots \cap A_{i_k}) ,$$

dove si intende che la sommazione sia da effettuare sopra tutte le k -uple di interi i_1, i_2, \dots, i_k soddisfacenti $i_1 < i_2 < \cdots < i_k$.

Per gli interi dispari $k \geq 1$ si dimostra che

$$P\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{j=1}^k (-1)^{j+1} S_j ,$$

mentre per gli interi pari $k \geq 2$

$$P\left(\bigcup_{i=1}^n A_i\right) \geq \sum_{j=1}^k (-1)^{j+1} S_j .$$

Probabilità condizionata

Si supponga che A sia un evento di interesse (del quale, cioè, si vuole calcolare la probabilità, $P(A)$) e che esso abbia intersezione non vuota con un altro, B, che ha probabilità, $P(B)$, già nota e non nulla. Dunque $A \cap B \neq \emptyset$

La probabilità condizionata $P(A|B)$ misura la probabilità che, come risultato dell'esperimento aleatorio, si realizzi (anche) l'evento A, sotto la condizione che si realizzerà B con certezza. Nel senso che si pone come ipotesi l'esito seguente dell'esperimento: il risultato è uno degli eventi elementari (ω) la cui unione definisce B e vale $P(B)$.

Si capisce che A ha una sorta di dipendenza da B. B accade ed influenza A!

Definizione:

$$P(A|B) \triangleq \sum_{\omega \in A \cap B} P(\omega | B \text{ certo}) = ?$$

Da B certo $\implies P(B) = 1$

Si pone:

$$P(\omega | B \text{ certo}) \equiv P(\omega) = \frac{P(\omega)}{P(B)} \quad \forall \omega \in B$$

$$P(\omega | B \text{ certo}) \equiv P(\omega) = 0 \quad \forall \omega \in \overline{B} \cap \Omega$$

E banalmente si verifica che:

$$\sum_{\omega \in B} \hat{P}(\omega) = \sum_{\omega \in B} \frac{P(\omega)}{P(B)} = 1$$

Dunque:

$$P(\omega | B \text{ certo}) = \frac{P(\omega)}{P(B)} \quad \omega \in A \cap B$$

e si ottiene:

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

la condizionata è pari al rapporto tra probabilità congiunta e probabilità condizionante

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

utile nelle manipolazioni

quindi la congiunta è pari alla condizionata per la condizionante! (dimostrazione nelle prossime pagine)

PROBABILITA' CONGIUNTA E "DIPENDENZA"

CONGIUNTA

$$P(A_1 \cap A_2 \cap A_3) = ?$$

Cattura la dipendenza
fra gli eventi.

Basta porre: $C \hat{=} A_2 \cap A_3$

$$P(A_1 \cap A_2 \cap A_3) = P(A_1 \cap C) = P(A_1 | C) \cdot P(C)$$

$$= P(A_1 | C) \cdot P(A_2 \cap A_3) \stackrel{\text{esplicito nuovamente}}{=} P(A_1 | C) \cdot P(A_2 | A_3) \cdot P(A_3)$$

$$= P(A_1 | A_2 \cap A_3) \cdot P(A_2 | A_3) \cdot P(A_3)$$

Assumendo: $P(A_1 | A_2 \cap A_3) \hat{=} P(A_1 | A_2)$

$$= P(A_1 | A_2) \cdot P(A_2 | A_3) \cdot P(A_3)$$

Quindi A1 dipende da A2 che dipende a sua volta da A3

dipendenza a catena

Dipendenza fra eventi generalizzata

È illustrata dalla formula seguente:

$$P(A_1 \cap A_2 \cap \dots \cap A_n) =$$

suggerimento: $B \hat{=} A_2 \cap \dots \cap A_n$

e poi $= P[A_1 \cap B] = P[A_1 | B]P[B]$

Formula finale
detta Chain Rule: $P(A_1 \cap A_2 \cap \dots \cap A_n) = P[A_1 | (A_2 \cap A_3 \cap \dots \cap A_n)] * \dots * P[A_{n-1} | A_n] * P[A_n]$

• Dipendenza a catena generalizzata

È definita nel seguente modo

$$\begin{aligned} P(A_1 \cap A_2 \cap \dots \cap A_n) &= P(A_n \cap A_{n-1} \cap \dots \cap A_1) \hat{=} \\ &\hat{=} P[A_n | A_{n-1}] * P[A_{n-1} | A_{n-2}] * \\ &\dots * P[A_2 | A_1] * P[A_1] \end{aligned}$$

pensare agli indici 1,...,n come una sequenza di istanti temporali ai quali sono agganciati/riferiti gli eventi corrispondenti, A1,...,An

• Indipendenza stocastica di una coppia di eventi

S'immagini di aver calcolato la probabilità di due eventi qualsivoglia ma non disgiunti (disgiunti significa logicamente incompatibili), $A \subset \Omega$ e $B \subset \Omega$, e di averla calcolata sulla base dell'assegnamento originario di probabilità ad eventi elementari di Ω . Siano esse $P(A)$ e $P(B)$, entrambe > 0 . Ciò premesso, A e B potrebbero essere definiti eventi stocasticamente indipendenti (indipendenti in probabilità) quando la modifica della probabilità già calcolata per uno dei due non induce modifiche alla probabilità calcolata per l'altro evento. La modifica di interesse pratico è quella che consiste nell'aggiornare una delle due probabilità $[P(A) \text{ o } P(B)]$ portandola al valore unitario, ovvero ipotizzare che si verifichi con certezza uno dei due eventi $[A \text{ o } B]$. Quindi, ci si chiede se la nuova $P(A|B)$ [ovvero $P(B|A)$] risulterà uguale alla vecchia $P(A)$ [ovvero $P(B)$] oppure no. In caso affermativo si potrà parlare di indipendenza stocastica. D'altra parte, tenendo presente che - grazie alla formula della probabilità congiunta - i risultati $P(A|B)=P(A)$ e $P(B|A)=P(B)$ sono entrambi implicati dall'assumere che risulti $P(A \cap B) = P(A)P(B) = P(B \cap A)$ si conviene di proporre la seguente definizione di indipendenza stocastica per la coppia di eventi A e $B \subset \Omega$:

Gli eventi A e $B \subset \Omega$ e non incompatibili, né di probabilità individuale nulla, sono indipendenti se e solo se risulta: $P(A \cap B) = P(A) \cdot P(B) > 0$

dimostrazione

$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A) = P(A)P(B) = P(B)P(A) > 0$$

Dunque la probabilità congiunta è pari al prodotto delle singole probabilità se e solo se c'è l'ipotesi di indipendenza!

E' interessante osservare che la formula (+) può essere letta alla seguente maniera:

$$\frac{P(A \cap B)}{P(A)} = \frac{P(B)}{P(\Omega)} \quad \text{ovvero} \quad \frac{P(B \cap A)}{P(B)} = \frac{P(A)}{P(\Omega)}$$

Nell'illustrazione grafica proposta, s'immagini per semplicità che l'area dello spazio Ω valga 1, come la probabilità ad esso associata:

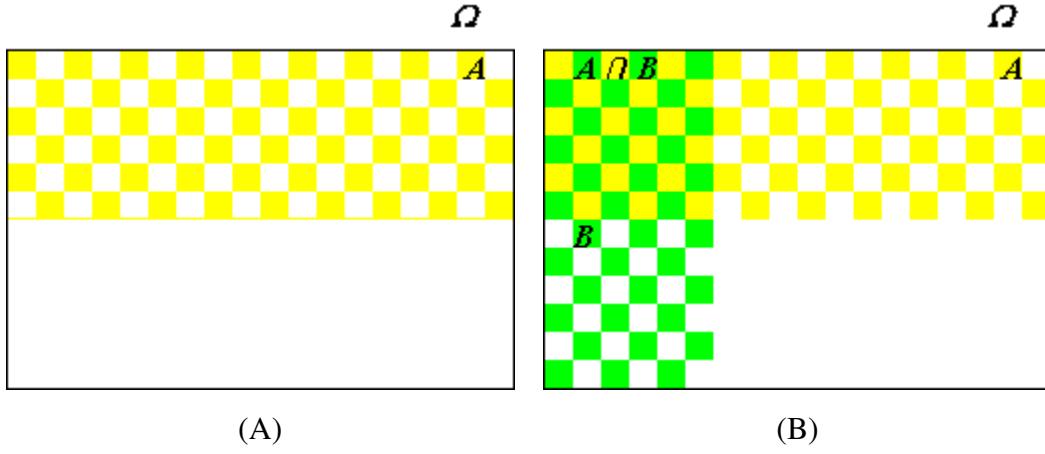


Figura – Illustrazione grafica della proprietà di indipendenza

Allora, visto che l'area di A appare proprio come la metà di quella di Ω e quella di B come $1/3$, risulta :

$$P(A) = \frac{1}{2}, \quad P(B) = \frac{1}{3}, \quad \text{ma} \quad P(A|B) = \frac{1}{2}$$

e quindi:

$$\frac{P(A \cap B)}{P(A)} = \frac{P(B)}{P(\Omega)} \quad \text{ovvero} \quad \frac{1/6}{1/2} = \frac{1/3}{1}$$

Dunque a e B sono stocasticamente indipendenti.

L'illustrazione grafica potrebbe servire a convincersi che il concetto di indipendenza stocastica è una proprietà che scaturisce dalla specifica rappresentazione topologica dei due eventi nello spazio campionario. Infatti, ruotando B di 90 gradi in senso orario esso diventa logicamente incompatibile con A, altro che indipendente!
Esempio

Dato un mazzo di carte da poker di 52 carte a 4 semi di cui due rossi (i.e. cuori e quadri) e due neri (i.e. fiori e picche), si considerino i seguenti eventi

$A \hat{=} \text{estrazione di un asso}; \quad B \hat{=} \text{estrazione di una carta di cuori}.$

La probabilità di questi eventi è data rispettivamente da

$$P(A) = \frac{4}{52} = \frac{1}{13}, \quad P(B) = \frac{13}{52} = \frac{1}{4}.$$

D'altra parte la probabilità di estrarre l'asso di cuori, ovvero $P(A \cap B)$ è $1/52$, come ogni altra carta, perché il mazzo non è truccato. E questo permette di verificare la validità della relazione:

$$\frac{P(A \cap B)}{P(A)} = \frac{P(B)}{P(\Omega)}. \quad \text{Infatti, risulta: } \frac{1/52}{1/13} = \frac{1/4}{1}.$$

Dunque abbiamo provato che è vera la seguente: $P(A \cap B) = P(A) * P(B)$, ovvero che A e B sono eventi stocasticamente indipendenti, senza invocare la probabilità condizionata!

Viceversa, è sufficiente trasformare B in evento certo e invocare la probabilità condizionata per ristabilire l'indipendenza stocastica fra A e B.

Infatti, si ha: $\Omega \rightarrow B$ e quindi: $P(A \cap B | B) \equiv P(A | B) = 1/13$, perché c'è solo un asso di cuori fra le 13 del seme cuori. A questo punto, il semplice confronto del valore appena ottenuto per $P(A|B)$ con $P(A)=1/13$ completa la prova.

Il "motivo" dell'indipendenza è che il rapporto tra il numero di eventi elementari che compongono B (13) e il totale degli eventi elementari (52) che compongono Ω è uguale al rapporto tra il numero di eventi elementari (1) che sono "asso \cap di cuori" ($A \cap B$) e il numero di eventi elementari che compongono A (4). Sempre $1/4$.

Si supponga adesso di lavorare con un mazzo di carte speciale, che contiene sempre 13 carte per seme ma che abbia 2 assi di cuori (buttando via il 2 di cuori, ad esempio). In tal caso, lo studente può verificare che cade l'indipendenza stocastica tra gli eventi A e B.

Il risultato sarebbe $\frac{2}{5} = \frac{1}{4}$

Indipendenza stocastica fra 3 eventi

Dati gli eventi A, B e C devono valere le seguenti condizioni di indipendenza mutua e indipendenza a coppie:

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

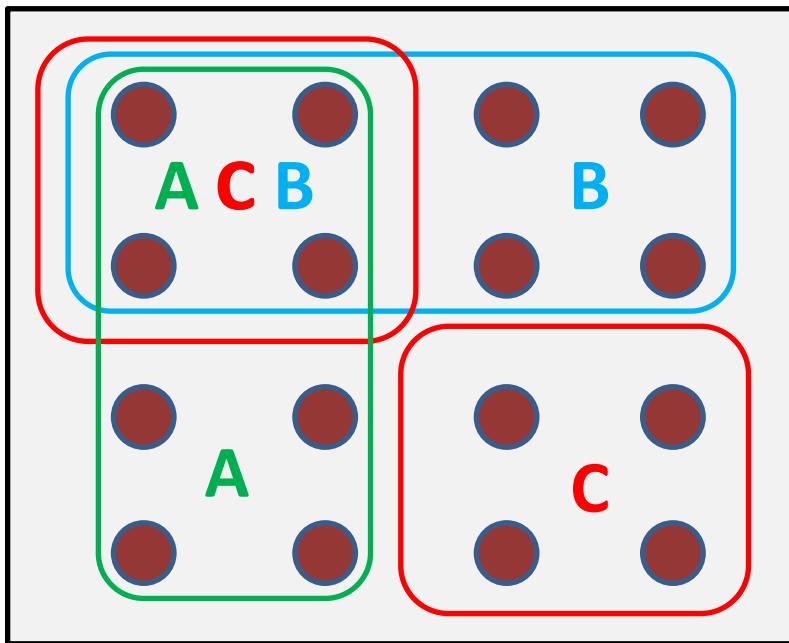
+

$$\left\{ \begin{array}{l} P(A \cap B) = P(A) \cdot P(B) \\ P(A \cap C) = P(A) \cdot P(C) \\ P(B \cap C) = P(B) \cdot P(C) \end{array} \right.$$

Se non valgono tutte, l'indipendenza stocastica non sussiste!

Generalizzare al caso di n eventi.

Pairwise independence and mutual independence



$$P(X \subseteq \Omega) \hat{=} \frac{\#\text{risultati} \in X}{\#\text{risultati} \in \Omega}$$

↓

$$\begin{aligned} P(A) &= P(B) = P(C) \\ &= 8/16 = 1/2 \end{aligned}$$

Pairwise independence and mutual independence

P $P(A \cap B) = P(A | B) \cdot P(B) = P(A) \cdot P(B)$
A $= (1/2) \cdot (1/2) = 1/4$
I

R $P(A \cap C) = P(A | C) \cdot P(C) = P(A) \cdot P(C)$
W $= (1/2) \cdot (1/2) = 1/4$
I

S $P(B \cap C) = P(B | C) \cdot P(C) = P(B) \cdot P(C)$
E $= (1/2) \cdot (1/2) = 1/4$

NO MUTUAL:

$$\begin{aligned}P(A \cap B \cap C) &= P(A | B \cap C) \cdot P(B \cap C) \\&= P(A | B \cap C) \cdot P(B | C) \cdot P(C) = 1 \cdot (1/2) \cdot (1/2) = 1/4 \\&\neq P(A) \cdot P(B) \cdot P(C) = (1/2) \cdot (1/2) \cdot (1/2) = 1/8\end{aligned}$$

Formula della probabilità totale e corollario di Bayes

Data una partizione $\{B_1, B_2, \dots, B_n\}$ dello spazio Ω e un evento di interesse (A):

$$A = A \cap \Omega = A \cap \bigcup_{i=1}^n B_i = \bigcup_{i=1}^n (A \cap B_i)$$

e da qui

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

grazie all'ipotesi di partizionamento di Ω , che rende disgiunti gli eventi:

$$A \cap B_i, i = 1, \dots, n$$

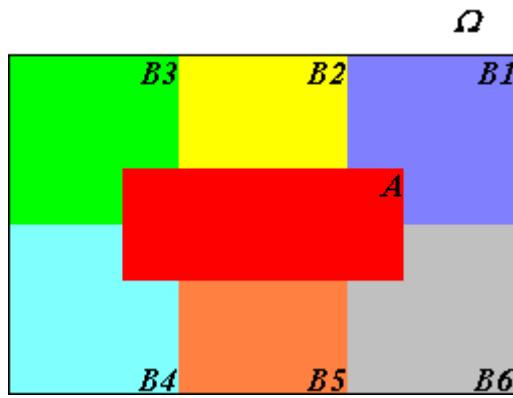


Figura – Illustrazione della formula della probabilità totale per A e $\{B_1, B_2, \dots, B_6\}$

Da notare come la totale sia una SOMMA di tutte le congiunte!

Quindi la totale serve a conoscere la probabilità di un evento a partire da una partizione

Corollario di Bayes: Importante

Scelto a caso l'evento B_j della partizione $\{B_1, B_2, \dots, B_j, \dots, B_n\}$ si ricava la formula

seguente, nota come regola di Bayes:

$$\begin{aligned} P(B_j | A) &= \frac{P(B_j \cap A)}{P(A)} = \frac{P(A \cap B_j)}{P(A)} \\ &= \frac{P(A | B_j)P(B_j)}{\sum_{i=1}^n P(A | B_i)P(B_i)} \end{aligned}$$

dove $P(B_j)$ è detta prior probability di B_j
 $P(B_j|A)$ è detta posterior probability di B_j

Essa consente di calcolare la probabilità che si realizzi un evento specifico della partizione, posto che si è realizzato l'evento di interesse.

In altre parole permette di calcolare la probabilità di un evento tramite informazioni conosciute a priori sull'evento appunto prior probability.

Esempio sulla probabilità totale e sulla regola di Bayes

Consideriamo tre macchine (M_1 , M_2 , M_3) che forniscono pezzi che dovranno essere assemblati dopo un controllo di qualità:

- I pezzi difettosi, espressi in percentuale, prodotti da ogni macchina sono:
 - M_1 ----- 1% (un pezzo su cento lavorati dalla macchina M_1 è difettoso)
 - M_2 ----- 5% (cinque pezzi su cento lavorati dalla macchina M_2 sono difettosi)
 - M_3 ----- 2% (due pezzi su cento lavorati dalla macchina M_3 sono difettosi)
- I pezzi che arrivano al controllore di qualità sono forniti al 15% dalla macchina M_1 al 35% dalla macchina M_2 e al 50% dalla macchina M_3

Definiamo gli eventi:

1. B_i = “provenienza pezzo” $i = M_1, M_2, M_3$
2. A = “pezzo difettoso”

Teorema della PROBABILITÀ TOTALE

$$P(A) = \text{probabilità “pezzo difettoso”}$$

$$P(A) = P(A|B_{M1})P(B_{M1}) + P(A|B_{M2})P(B_{M2}) + P(A|B_{M3})P(B_{M3})$$

$$P(A) = P(0.01)P(0.15) + P(0.05)P(0.35) + P(0.02)P(0.5) = 0.029$$

REGOLA BAYES:

$P(B_{M2}|A) =$ probabilità che il pezzo difettoso, scelto casualmente tra i pezzi difettosi, proviene dalla M_2

$$P(B_{M2} | A) = \frac{P(A | B_{M2})P(B_{M2})}{P(A)} = \frac{0.05 * 0.35}{0.029} \cong 0.51$$

“Conditional Independence” e grafi (di conoscenza)

DEFINIZIONE:

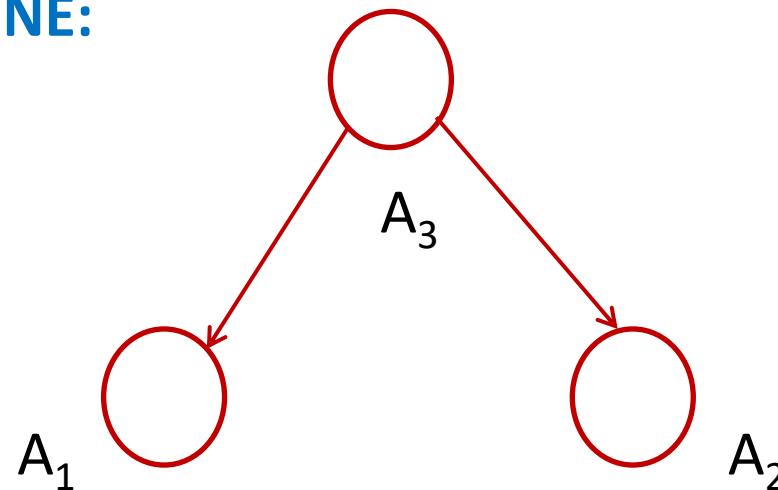
$$P(A_1 \cap A_2 | A_3) \doteq P(A_1 | A_3) \cdot P(A_2 | A_3)$$

DI CONSEGUENZA:

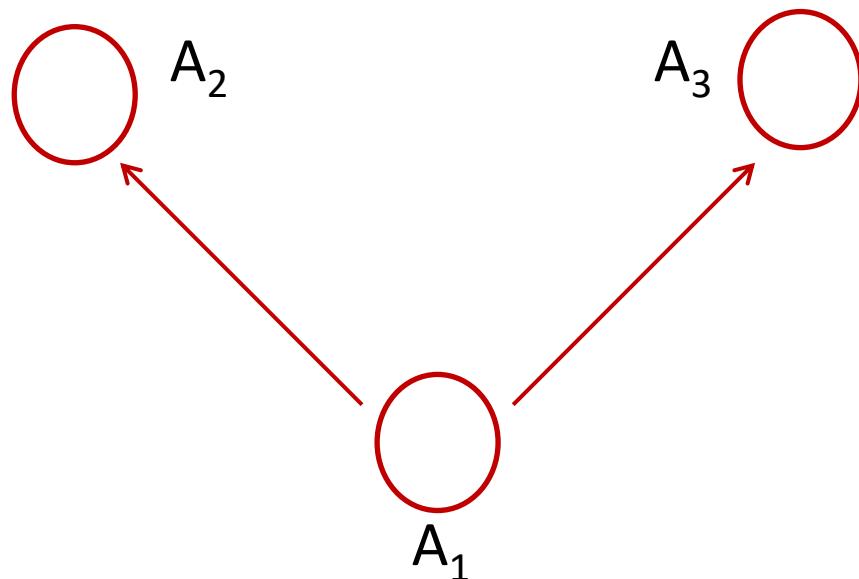
vedi probabilità congiunta e dipendenza

$$\begin{aligned} P(A_1 \cap A_2 \cap A_3) &= P(A_1 \cap A_2 | A_3) \cdot P(A_3) \\ &= P(A_1 | A_3) \cdot P(A_2 | A_3) \cdot P(A_3) \end{aligned}$$

RAPPRESENTAZIONE:



“Conditional Independence” e grafi



OSSERVAZIONE:

NO arco tra
A2 e A3

$$\begin{aligned} P(A_1 \cap A_2 \cap A_3) &= P(A_1 | A_2 \cap A_3) \cdot P(A_2 \cap A_3) \\ &= P(A_1 | A_2 \cap A_3) \cdot P(A_2) \cdot P(A_3) \end{aligned}$$

ALLORA:

$$\begin{aligned} P(A_2 \cap A_3 | A_1) &= P(A_2 \cap A_3) / P(A_1) \\ &= P(A_2) \cdot P(A_3) / P(A_1) \neq P(A_2) \cdot P(A_3) \end{aligned}$$

A2 and A3 are INDEPENDENTS but are NOT CONDITIONAL INDEP

ESEMPIO

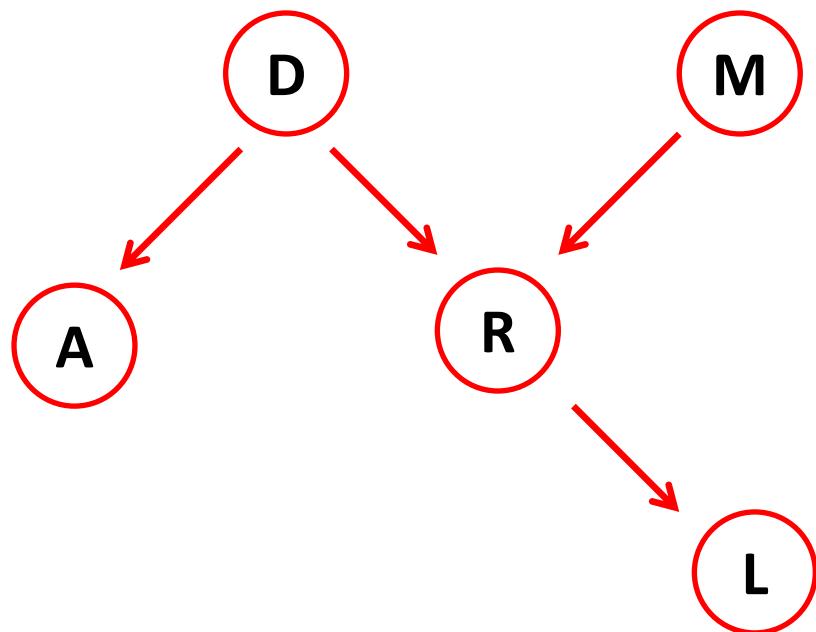
L := “La Lezione comincia alle 10 e 50”

R := “Il docente arriva in Ritardo”

A := “Argomento della lezione è MODELLI su GRAFI.”

M := “Forte Maltempo in atto”

D := “Il Docente è Legato”



ASSUNZIONI:

D INDIP da M

A è DIP da D

R è DIP da M

L è DIP da R

Affidabilità come probabilità di successo

Sulla base dei primi concetti di analisi probabilistica, l'affidabilità (*Reliability*) di un sistema può essere definita come la probabilità che esso risulti funzionante (successo) all'osservatore casuale che "ispeziona" il sistema in un istante di tempo che non ha importanza. L'osservatore casuale è in grado di rilevare lo stato di buono o cattivo funzionamento di tutti i componenti, ripetendo l'osservazione dello stato di ognuno di essi in tempi trascurabili.

Esempio: Controllo del funzionamento del server di prenotazioni a 00:00

Si consideri l'evento

$A_i \hat{=} \text{"componente } i\text{-mo funziona"}$, $i=1..n$,

Affidabilità del sistema seriale

$$R_S \hat{=} P(\text{"sistema funziona"}) \\ = P(\text{"tutti i componenti funzionano"})$$

allora

$$R_S = P(A_1 \cap A_2 \cap \dots \cap A_n) = \prod_{i=1}^n R_i, \quad \text{ove } R_i \hat{=} P(A_i)$$

se aggiungiamo l'ipotesi che i componenti sono indipendenti.

appunto congiunta
che dei singoli
componenti funzionanti.
Se un componente non
funziona, non funziona
tutto il sistema.

Affidabilità del sistema parallelo:

$$R_P = 1 - P(\text{"il sistema non funziona"}) \\ = 1 - P(\text{"tutti i componenti non funzionano"}) \\ = 1 - P(\overline{A}_1 \cap \overline{A}_2 \cap \dots \cap \overline{A}_n) \\ = 1 - \prod_{i=1}^n P(\overline{A}_i) = 1 - \prod_{i=1}^n (1 - P(A_i)) = 1 - \prod_{i=1}^n (1 - R_i)$$

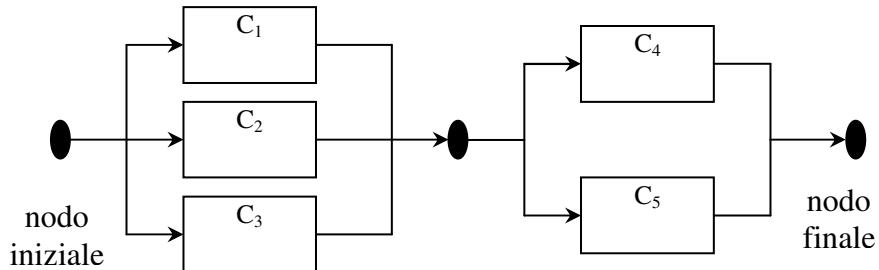
Il sistema smette di funzionare
se non vi è alcun, componente
funzionante.

se aggiungiamo l'ipotesi che i componenti sono indipendenti.

In entrambi i casi si hanno componenti indipendenti!

Diagrammi a Blocchi in Affidabilità

Schematizzazione a blocchi dell'organizzazione di un sistema:



PER DEFINIZIONE:

Il sistema funziona se la numerosità e la disposizione dei componenti rilevati funzionanti dall'osservatore casuale sono tali da garantire almeno un cammino orientato (senza cicli interni) tra nodo iniziale e nodo finale sul quale tutti i componenti sono funzionanti.

Ad esempio: C1, C2 e C5 funzionanti \Rightarrow sistema funziona (due cammini)

Ragionando in termini di cammini, l'affidabilità di un sistema si esprime come segue:

$$P(\text{"Sistema funziona"}) = P\left(\bigcup_{i=1}^n \text{"Tutti funz. sul Camm}_i\text{"}\right)$$

Il fatto che lo stesso componente possa appartenere a più di un cammino esclude la possibilità di trattare gli eventi "Ai" = "Tutti funzionanti sul cammino i", $i=1,\dots,n$ come eventi disgiunti. Quindi per calcolare l'affidabilità del sistema, occorre applicare la formula della probabilità dell'unione di n eventi non disgiunti:

$$\begin{aligned} P\left(\bigcup_{i=1}^n A_i\right) &= P(A_1 \cup A_2 \cup \dots \cup A_n) = \\ &= \sum_{\forall i} P(A_i) - \sum_{1 \leq i < j \leq n} P(A_i \cap A_j) + \sum_{1 \leq i < j < k \leq n} P(A_i \cap A_j \cap A_k) + \dots + (-1)^{n-1} P(A_1 \cap A_2 \cap \dots \cap A_n) \end{aligned}$$

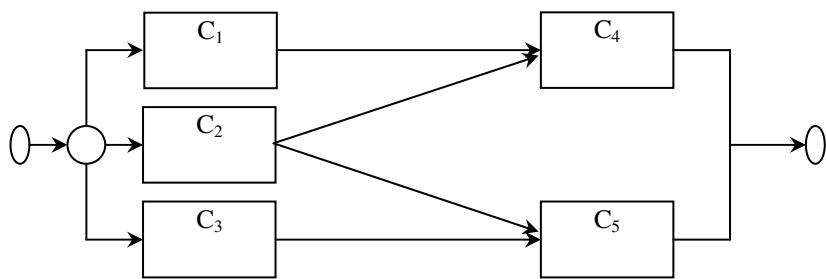
vista a pagina 16

Il metodo dello spazio degli eventi

Quando un sistema è organizzato in una forma che non può essere ricondotta alla composizione di blocchi seriali o paralleli o misti si può fare ricorso al metodo di analisi dello spazio degli eventi. Dove occorre:

- *Individuare tutti i possibili eventi;*
- *Definire gli eventi corrispondenti al funzionamento del sistema;*
- *Sviluppare le formule di probabilità.*

Analisi del sistema:



Considerati i soliti eventi

$$A \triangleq \text{"sistema funziona"}$$

$$A_i \triangleq \text{"componente } i\text{-mo funziona"}$$

l'analisi è basata sulla relazione che lega l'evento A agli eventi cui corrisponde il funzionamento del sistema e che vanno individuati per ispezione del diagramma a blocchi. Per il sistema considerato, risulta:

$$A = (A_1 \cap A_4) \cup (A_2 \cap A_4) \cup (A_2 \cap A_5) \cup (A_3 \cap A_5)$$

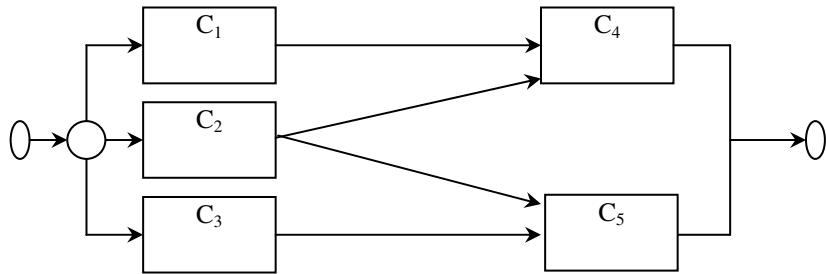
Osservando che i termini dell'unione hanno intersezione non nulla a due a due (A_4 compare nel primo e nel secondo termine, ad esempio), l'affidabilità del sistema può essere calcolata con la formula seguente:

$$\begin{aligned} P(A) &= P(E_1 \cup E_2 \cup E_3 \cup E_4) = \\ &= \sum_{1 \leq i \leq 4} P(E_i) - \sum_{1 \leq i < j \leq 4} P(E_i \cap E_j) + \sum_{1 \leq i < j < k \leq n} P(E_i \cap E_j \cap E_k) - P(E_1 \cap E_2 \cap E_3 \cap E_4) \end{aligned}$$

dove $E_1 \triangleq A_1 \cap A_4$, $E_2 \triangleq A_2 \cap A_4$ e così via.

Esempio: Probabilità totale e analisi di affidabilità

Alternativa al metodo dello spazio degli eventi per il sistema:



Considerati i due eventi:

$B \hat{=} \text{"Componente 2 funziona"}$ e $A \hat{=} \text{"Sistema funziona"}$

una possibile partizione di Ω è $\{B, \bar{B}\}$ e quindi la probabilità che il sistema sia funzionante è:

$$\begin{aligned} &\text{dalla formula della totale} \\ P(A) &= P(A \cap B) + P(A \cap \bar{B}) \\ &= P(A | B)P(B) + P(A | \bar{B})P(\bar{B}) \end{aligned}$$

dopo aver calcolato: $P(A | B)$ e $P(A | \bar{B})$.

Con la condizione di funzionamento certo del componente 2, il sistema si riduce al parallelo dei componenti 4 e 5. Applicando appunto l'affidabilità in parallelo

$$\text{Allora: } P(A | B) = 1 - (1 - R_4)(1 - R_5)$$

Viceversa, assumendo che il componente 2 non funzioni, il sistema si riduce al parallelo di due blocchi; il primo blocco essendo costituito dalla serie dei componenti 1 e 4 e il secondo blocco dalla serie dei componenti 3 e 5. Due blocchi indipendenti tra loro!

$$\text{Allora: } P(A | \bar{B}) = 1 - (1 - R_1R_4)(1 - R_3R_5)$$

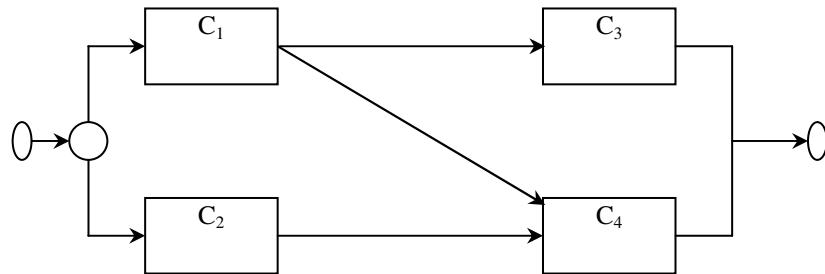
In definitiva:

$$R_{\text{SISTEMA}} = [1 - (1 - R_4)(1 - R_5)]R_2 + [1 - (1 - R_1R_4)(1 - R_3R_5)](1 - R_2)$$

CONGIUNTA

Esempio: Spazio degli eventi, probabilità totale e regola di Bayes

È dato il seguente schema a blocchi dell'organizzazione di un sistema:



ed è nota l'affidabilità dei singoli componenti, intesa come “probabilità di successo”:

$$R_{C1} = 0,95, R_{C2} = 0,85, R_{C3} = 0,92, R_{C4} = 0,98$$

- Calcolare l'affidabilità del sistema applicando il metodo dello spazio degli eventi;
- Ripetere il calcolo dell'affidabilità facendo uso della formula della probabilità totale;
- Calcolare la probabilità che il componente C1 funzioni, posto che il sistema funziona.

Soluzione

Si definisca l'evento $A \hat{=} \text{almeno un cammino funzionante nel sistema}$.

I cammini possibili sono:

$$B_1 = C_1 \cap C_3, B_2 = C_2 \cap C_4, B_3 = C_1 \cap C_4 .$$

Quindi,

$$A = B_1 \cup B_2 \cup B_3 = (C_1 \cap C_3) \cup (C_2 \cap C_4) \cup (C_1 \cap C_4)$$

$$\begin{aligned} P(A) &= P(B_1 \cup B_2 \cup B_3) \\ &= P(B_1) + P(B_2) + P(B_3) - P(B_1 \cap B_2) - P(B_1 \cap B_3) - P(B_2 \cap B_3) + \\ &\quad + P(B_1 \cap B_2 \cap B_3) \\ &= P(B_1) + P(B_2) + P(B_3) - P(B_1 \cap B_2) - P(B_1 \cap B_3) + \\ &\quad + P(B_1 | B_2 \cap B_3) * P(B_2 \cap B_3) \\ &= P(B_1) + P(B_2) + P(B_3) - P(B_1 \cap B_2) - P(B_1 \cap B_3) + \\ &\quad + P(B_1 | B_2 \cap B_3) * P(B_2 | B_3) * P(B_3) \end{aligned}$$

$$\begin{aligned}
&= (R_{C1} * R_{C3}) + (R_{C2} * R_{C4}) + (R_{C1} * R_{C4}) - (R_{C1} * R_{C3} * R_{C2} * R_{C4}) + \\
&\quad - (R_{C2} * R_{C1} * R_{C4}) - (R_{C3} * R_{C1} * R_{C4}) + (R_{C3} * R_{C2} * R_{C1} * R_{C4}) \\
&= 0,874 + 0,833 + 0,931 - 0,728042 - 0,79135 - 0,85652 + 0,728042 \\
&= 0,99013
\end{aligned}$$

Usando la formula della probabilità totale, si considerino i seguenti eventi:

$A \hat{=} \text{il sistema funziona}$

$B \hat{=} \text{il componente } C_1 \text{ funziona}$

allora

$$P(A) = P(A | B) * P(B) + P(A | \bar{B}) * P(\bar{B}) = P(A \cap B) + P(A \cap \bar{B}).$$

Se il componente C_1 non funziona, il sistema si riduce alla serie fra C_2 e C_4

$$P(A | \bar{B}) = R_{C2} * R_{C4}.$$

Se il componente C_1 funziona, il sistema si riduce al parallelo fra i componenti C_3 e C_4

$$P(A | B) = 1 - [(1 - R_{C3}) * (1 - R_{C4})].$$

In conclusione

$$P(A) = [1 - (1 - R_{C3}) * (1 - R_{C4})] * R_{C1} + (R_{C2} * R_{C4}) * (1 - R_{C1}) = 0,99013$$

La probabilità che il componente C_1 funzioni, posto che il sistema funziona, si ottiene applicando la regola di Bayes:

$A \hat{=} \text{il sistema funziona}$

$B \hat{=} \text{il componente } C_1 \text{ funziona}$

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

$$P(A | B) * P(B) = [1 - (1 - R_{C3}) * (1 - R_{C4})] * R_{C1}$$

Quindi

$$P(B | A) = \frac{0,94848}{0,99013} = 0,957934816 \approx 0,96$$

Esempio: Probabilità totale e limite superiore all'Affidabilità

Sia $\{C_1, C_2, \dots, C_n\}$ l'insieme dei cammini individuati su un diagramma a blocchi assegnato e si consideri l'evento:

$$D_i \hat{=} \text{"Almeno un componente in } C_i \text{ è guasto".}$$

In linea di principio si può scrivere:

$$\begin{aligned} 1 - R &= P(D_1 \cap D_2 \cap \dots \cap D_n) \\ &= P(D_n | D_1 \cap D_2 \cap \dots \cap D_{n-1}) \dots P(D_2 | D_1)P(D_1) \end{aligned}$$

ma nella pratica, avendo a che fare con diagrammi di dimensioni non banali, la relazione precedente non è utilizzabile.

Però si potrebbe calcolare un valore limite (superiore), rinunciando al valore vero.

Per ricavare il valore limite vero occorre provare che:

$$P(D_2 | D_1) \geq P(D_2)$$

PROVA:

$$P(D_2) = P(D_2 | D_1)P(D_1) + P(D_2 | \overline{D_1})(1 - P(D_1))$$

e riconoscendo che

$$\begin{aligned} P(D_2 | \overline{D_1}) &= \text{Prob "almeno un componente guasto in } C_2 | \text{ tutti funzionano in } C_1 \\ &\text{risulta} \end{aligned}$$

$$\begin{aligned} P(D_2 | \overline{D_1}) &= 1 - \prod_{\substack{i \in C_2 \\ i \notin C_1}} R_i \leq 1 - \prod_{i \in C_2} R_i = P(\overline{D_2}) \\ \Rightarrow P(D_2) &\leq P(D_2 | D_1)P(D_1) + P(D_2)(1 - P(D_1)) \\ \Rightarrow 0 &\leq P(D_2 | D_1) - P(D_2) \end{aligned}$$

Quindi

$$P(D_i | D_1 \dots D_{i-1}) \geq P(D_i) \Rightarrow (1 - R) \geq \prod_i P(D_i)$$

In conclusione:

$$R \leq 1 - \prod_{i=1}^n \left(1 - \prod_{k \in C_i} R_k \right)$$

IMPORTANTE

Le prove di Bernoulli

Rappresentano un possibile modello della ripetizione di "n" tentativi di accesso di un utente ad una risorsa, permettendo di calcolare: ad esempio richiesta di usare cpu, se fallisco ritento

- La probabilità di "k" successi su "n" tentativi;
- La probabilità che il primo successo si verifichi al "k-mo" tentativo.

Lo spazio dei possibili risultati per una sola prova è: $\Omega_1 = \{0,1\}$;

Per due prove è: $\Omega_2 = \{(0,0), (0,1), (1,0), (1,1)\}$

Per "n" prove è: $\Omega_n = \{2^n \text{ n-più di zero e uno}\}$

Nel modello di Bernoulli, l'assegnamento delle probabilità ai risultati di Ω_1 è:

$$P(0) = q \geq 0, \quad P(1) = p \geq 0 \quad \text{con} \quad p + q = 1$$

perciò successi p, fallimenti q=1-p

Per l'assegnamento delle probabilità a Ω_n è necessario definire:

$A_i \hat{=} \text{"successo alla prova i-ma"}, \quad \overline{A_i} \hat{=} \text{"insuccesso alla prova i-ma"}$.

Si consideri ora un possibile $\omega_n \in \Omega_n$: sequenza di prove

$$\omega_n = A_1 \cap A_2 \cap \dots \cap A_k \cap \overline{A_{k+1}} \cap \dots \cap \overline{A_n}.$$

Nel modello di Bernoulli, le prove sono indipendenti e le probabilità di successo e insuccesso non cambiano all'aumentare delle prove, e l'ordine dei successi/insuccessi è irrilevante. Le prove sono n e devono essere tutte eseguite, quando si vuole che la probabilità di "k" successi su "n" tentativi. Dunque si ottiene:

$$P(\omega_n) = P(A_1 \cap A_2 \cap \dots \cap A_k \cap \overline{A_{k+1}} \cap \dots \cap \overline{A_n}) = p^k q^{n-k}.$$

Come caso particolare si osserva che:

$$q^{k-1} p$$

è la probabilità che il primo successo si realizzi al tentativo "k-simo" (nota anche come probabilità geometrica). Da notare che le prove a disposizione diventano illimitate, quando si vuole la probabilità che il primo successo si verifichi al "k-mo" tentativo

Inoltre, se non interessa l'ordine in cui si ottengono:

$$P("k \text{ successi in } n \text{ prove}") = \binom{n}{k} p^k q^{n-k} \quad k = 0, 1, \dots, n .$$

La precedente è nota come **probabilità binomiale**.

Non interessa l'ordine dei successi

Generalizzazione delle prove di Bernoulli

Consiste nell'associare "k" possibili risultati ad ognuna delle prove.

Esempio: a "k" possibili risorse può essere mirato ognuno degli "n" tentativi di accesso di un utente, oppure l'unico tentativo di accesso di "n" utenti che si susseguono uno dopo l'altro.

Occorre identificare lo spazio dei risultati di n prove generalizzate (Ω_n) .

Sia ω l'evento elementare di Ω_n , allora:

$$\omega \triangleq \left\{ \begin{array}{c} 1,1,\dots,1; 2,2,\dots,2; \dots; k,k,\dots,k \\ n_1 \text{ volte} \quad n_2 \text{ volte} \quad \dots \quad n_k \text{ volte} \end{array} \right\} \text{ e } (n_1 + n_2 + \dots + n_k = n)$$

$$P(n_1, n_2, \dots, n_k) = \binom{n}{n_1} \binom{n-n_1}{n_2} \dots \binom{n-n_1-\dots-n_{k-1}}{n_k} P_1^{n_1} P_2^{n_2} \dots P_k^{n_k}$$

dove

$$Prob\{ \text{"vero il j-mo risultato nella generica prova"} \} \triangleq P_j$$

Dunque in definitiva nelle prove di Bernoulli la probabilità di successo rimane costante e indipendente!

La probabilità Binomiale Negativa

Se si ribaltano i termini dell'esperimento bernoulliano, fissando il numero di successi (k) che si vogliono ottenere e lasciando libero il numero di tentativi (n), si vuole costruire un modello di ripartizione per determinare la probabilità di quanti tentativi bisogna effettuare per ottenere k successi. Si definiscano pertanto i seguenti eventi:

- $A = \text{"n° tentativi per realizzare k successi"}$. Con k fissato a priori e n° tentativi (n) libero, purché sia $n \geq k$
- $B = \text{"esattamente } k-1 \text{ successi in } n-1 \text{ tentativi"}$ successi fissati quindi prove di Bernoulli!
- $C = \text{"il tentativo } n\text{-esimo è un successo"}$

L'evento A può essere calcolato a partire dai due eventi indipendenti B e C. A è verificato se si verificano gli eventi B e C, ossia:

$$A = B \cap C$$

Quindi possiamo calcolare la probabilità che si verifichi l'evento A come:

$$P(A) = P(B \cap C) = P(B | C) * P(C)$$

Poiché gli eventi B e C sono tra loro indipendenti, possiamo riscrivere la predente espressione come segue:

$$P(A) = P(B) * P(C)$$

P(B) si può calcolare con il modello di Bernoulli. Abbiamo infatti visto che, nel caso in cui non sia importante la sequenza dei successi, la probabilità di ottenere esattamente k successi in n prove è data dalla probabilità binomiale:

$$P(\text{"k successi in n prove"}) = \binom{n}{k} p^k q^{n-k} \quad k = 0, 1, \dots, n.$$

P(B) si calcola considerando: una particolare sequenza di n-1 tentativi, in cui i primi k-1 sono successi ed i restanti $n-1-(k-1)=n-k$ sono insuccessi, e moltiplicando la probabilità di quella sequenza per il numero di tutte le possibili sequenze di n-1 tentativi con k-1 successi.

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

La probabilità di successo di ogni singola prova è nota ed è pari a p . La probabilità che si verifichi l'evento C è quindi: $P(C) = p$. A questo punto è possibile calcolare la probabilità dell'evento A per semplice sostituzione delle probabilità calcolate.

$$P(A) = \binom{n-1}{k-1} p^{k-1} q^{n-k}$$

La precedente è nota come *probabilità binomiale negativa* (o di Pascal).

Si può notare come diremo molto in futuro che la binomiale è la somma di prove di Bernoulli.

Prove di Bernoulli in Affidabilità: Sistemi “m-out-of-n”

Per questo tipo di sistemi l'affidabilità è definita come segue

$$R_{m|n} = P(\text{"almeno } m \text{ componenti funzionano"}) = P(A(\geq m))$$

Partizionando lo spazio dei risultati Ω si può scrivere

Cioè il sistema funziona quando almeno “m” componenti su “n” funzionano. Cioè evento A($\geq m$)

$$\begin{aligned} R_{m|n} &= P\left(\bigcup_{i=m}^n \text{"esattamente } i \text{ componenti funzionano"}\right) \\ &= \sum_{i=m}^n P(\text{"esattamente } i \text{ componenti funzionano"}) \end{aligned}$$

Dunque: $R_{m|n} = P(A(\geq m)) = P\left(\bigcup_{i=m}^n A(=i)\right) = \sum_{i=m}^n P(A(=i))$ Rifacendosi alle prove di Bernoulli:

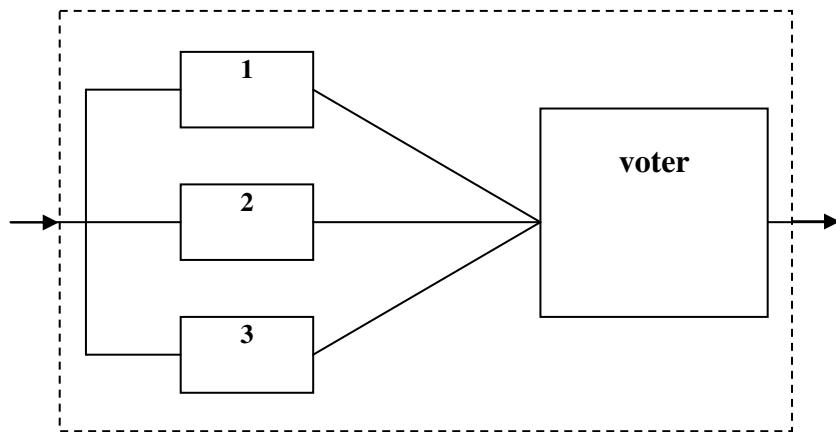
$$R_{m|n} = \sum_{i=m}^n \binom{n}{i} R^i (1-R)^{n-i}$$

con ipotesi che i componenti sono indipendenti ed identici con $AFF=R$.

Ridondanza con voto di maggioranza

Come esempio di sistema “m-out-of-n” si riporta il sistema a ridondanza modulare tripla (TMR). L’organizzazione TMR è classica nei sistemi di elaborazione e di controllo di apparecchiature o impianti. Tre oggetti, solitamente identici, lavorano in parallelo producendo un risultato che viene mandato in ingresso ad un comparatore (voter). Il comparatore ha il compito di rilevare l’eventuale disaccordo fra i risultati, che in condizioni di normalità devono essere identici assicurando che il risultato è corretto. Scoprire che uno degli oggetti produce un risultato diverso dagli altri induce a ritenere che sia proprio quello a funzionare male: in tal caso quell’oggetto viene messo fuori linea. Finché i due oggetti rimanenti continueranno a produrre lo stesso risultato si potrà continuare a ritenere che sia quello corretto. Al primo disaccordo il sistema dovrà essere considerato mal funzionante e quindi disattivato.

Se non sono identici il comparatore non funziona!



L'affidabilità del TMR è:

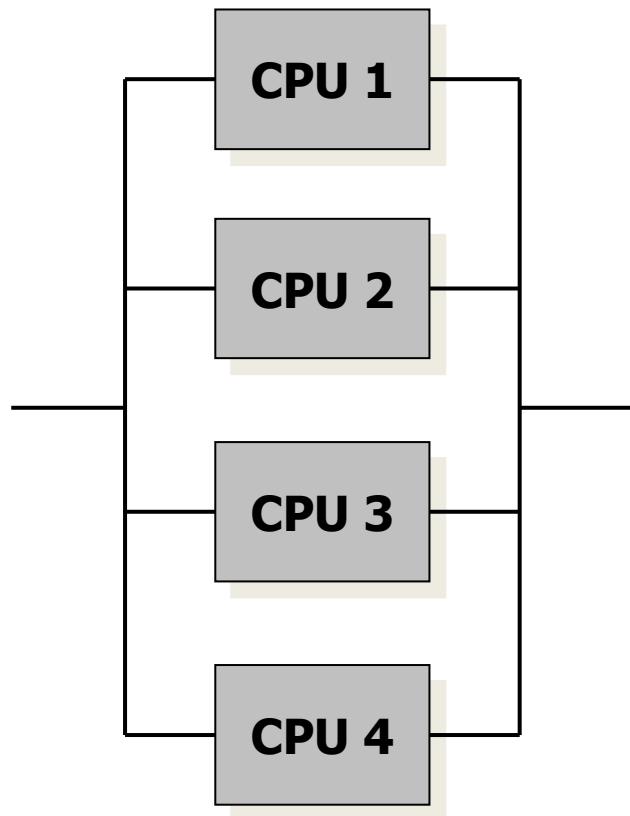
$$R_{m|n} = \sum_{i=m}^n \binom{n}{i} R^i (1-R)^{n-i}$$

dunque sostituendo n=3 ed m=2

$$R_{TMR} = \binom{3}{2} R^2 (1-R)^{3-2} + \binom{3}{3} R^3 (1-R)^{3-3}$$

Esercizio su sistema “m-out-of-n”: descrizione del problema e quesiti (1|2)

- Un sistema di calcolo ad elaborazione parallela, dotato di 4 CPU identiche e una RAM comune, è considerato "funzionante" se risultano funzionanti almeno 2 qualsivoglia delle 4 CPU, a parte la RAM.
- Sapendo che l'affidabilità della singola CPU vale 0.9, proporre un modello di calcolo e calcolare l'affidabilità del sistema.



Esercizio: descrizione del problema e quesiti (2|2)

1. Qual è la probabilità che il sistema lavori al massimo della sua prestazione, ovvero con tutte e quattro le CPU funzionanti?
2. Qual è la probabilità che "funzioni al minimo, cioè con 2 sole CPU", però dando per certo che funziona?
3. Ricalcolare l'affidabilità del sistema nell'ipotesi che debba funzionare una specifica CPU (ad es. la n°1) e poi anche solo un'altra delle rimanenti 3.
4. Ricalcolare l'affidabilità del sistema dando per certo che funziona una specifica CPU (ad es. la la n°1).
5. Rappresentare il diagramma a blocchi del sistema e provare a calcolare l'affidabilità col metodo dello spazio degli eventi.

Esercizio: modello e calcolo dell'affidabilità

Si tratta di un sistema «m-out-of-n» con n=4 e m=2.

Ricordando:

$$R_{m|n} = \sum_{i=m}^n \binom{n}{i} \cdot (R)^i \cdot (1-R)^{n-i}$$

Si calcola:

$$R_{2|4} = \sum_{i=2}^4 \binom{4}{i} \cdot (0.9)^i \cdot (0.1)^{4-i} = 0.0486 + 0.2916 + 0.6561 = 0,9963$$

Osservazione:

Sia C_i l'evento: «esattamente i cpu funzionano»

e sia F l'evento: «il sistema funziona»

Allora:

$$R_{2|4} \doteq P(F) = P\left(\bigcup_{i=2}^{i=4} (F \cap C_i)\right) = \sum_{i=2}^{i=4} P(F \cap C_i)$$

Perciò:

$$P(F \cap C_2) = 0.0486; \quad P(F \cap C_3) = 0.2916; \quad P(F \cap C_4) = 0.6561$$

Esercizio: risposte ai quesiti 1. e 2.

Quesito 1:

$$P(F \cap C_4) = \binom{4}{4} (0.9)^4 (0.1)^0 = 0.6561$$

Quesito 2:

Sia E l'evento: «solo 2 su 4 funzionano | sistema funziona»

Allora:

$$P(E) = \frac{P(F \cap C_2)}{P(F)} = \frac{0.0486}{0.9963} = 0.0487$$

Osservazione:

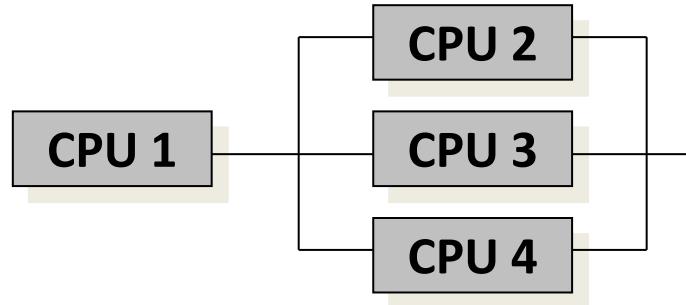
La probabilità (incondizionata) che il sistema funzioni al livello di prestazione 2 su 4 è quasi la stessa della probabilità (condizionata) che il sistema funzioni allo stesso livello quando funziona, solo perché $P(F)$ è quasi =1.

Ma le due probabilità in questione sono concettualmente ben diverse!

Esercizio: risposte ai quesiti 3. e 4.

Quesito 3:

Schema
corrispondente:

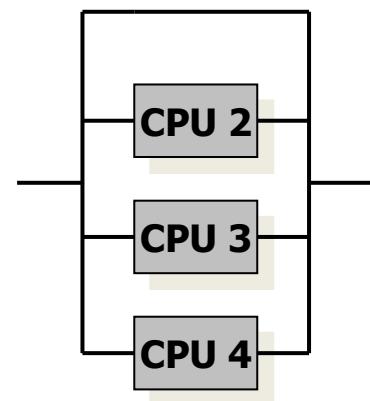


$$\begin{aligned} R &= R_{CPU1} \cdot R_{1|3} = R_{CPU1} \cdot (1 - (1 - R_2)(1 - R_3)(1 - R_4)) \\ &= 0.9 \cdot (1 - 0.9)(1 - 0.9)(1 - 0.9) = 0.8991 \end{aligned}$$

Quesito 4:

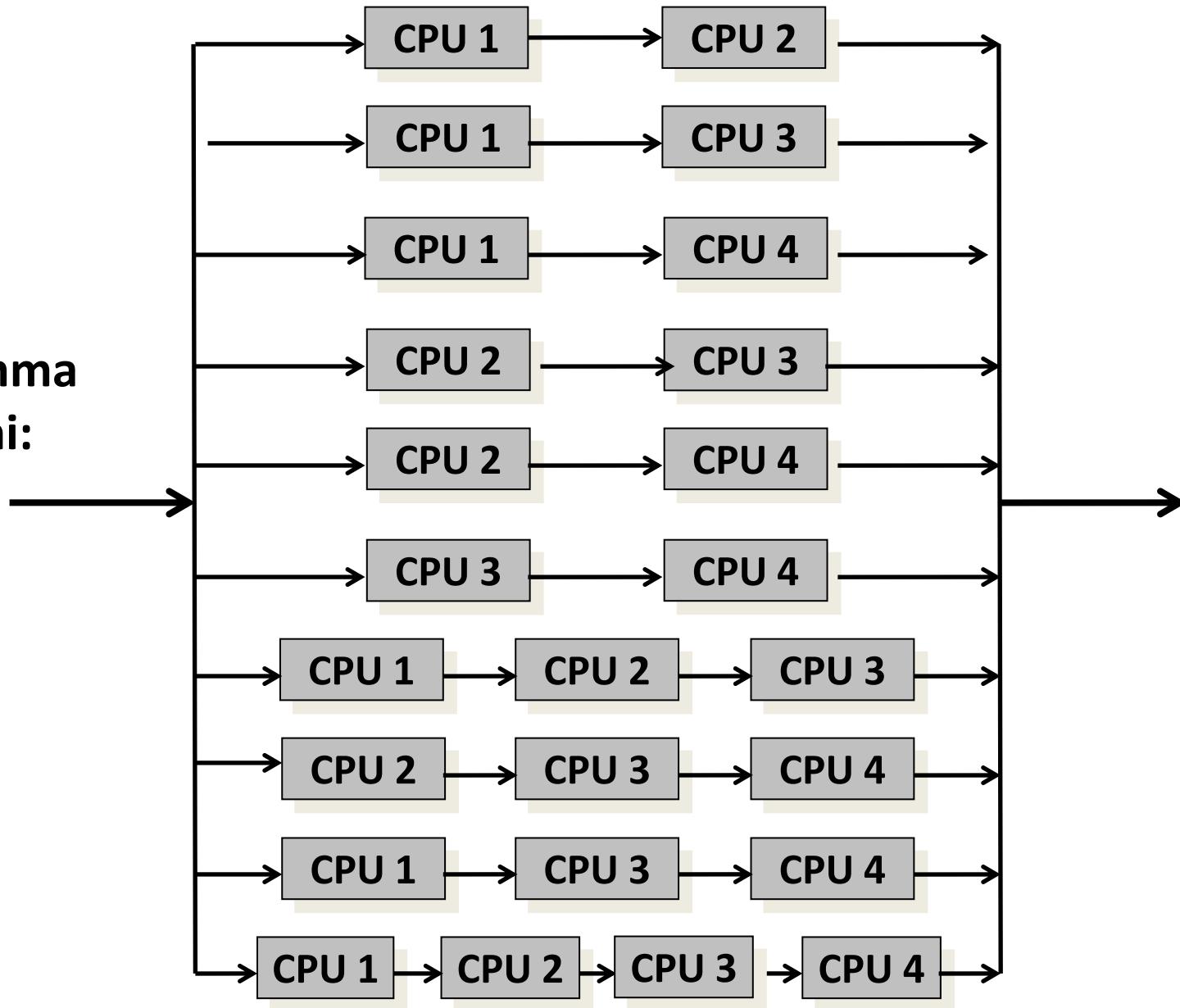
$R = P(2\text{-out-of-4} \mid \text{CPU 1 funziona certamente}) = P(1\text{-out-of-3})$

$$R = R_{2|4} \mid \text{CPU1} = R_{1|3} = 1 - (1 - R_2)(1 - R_3)(1 - R_4) = ???$$



Esercizio: risposta al quesito 5.

Diagramma
a blocchi:



Esercizio 1: descrizione del problema e quesiti (1|2)

Un sistema di elaborazione è composto da 4 processori che possono accedere tramite 1 bus a 4 moduli di memoria.

- Ad ogni ciclo di clock, ognuno dei processori può richiedere il bus per accedere ad uno dei moduli di memoria.
- Se più di uno è richiedente, il conflitto viene risolto assegnando il bus in maniera completamente casuale (ovvero equiprobabile).
- Un processore richiedente, al quale non è stato assegnato il bus, può anche non insistere nella richiesta al successivo ciclo di clock.
OVVERO INDIPENDENZA TRA GLI EVENTI
- La probabilità che un certo numero « n , ($n=0,1,\dots,4$)» di processori richieda di accedere al bus nello stesso ciclo di clock è nota, per ipotesi, dalla tabella seguente:

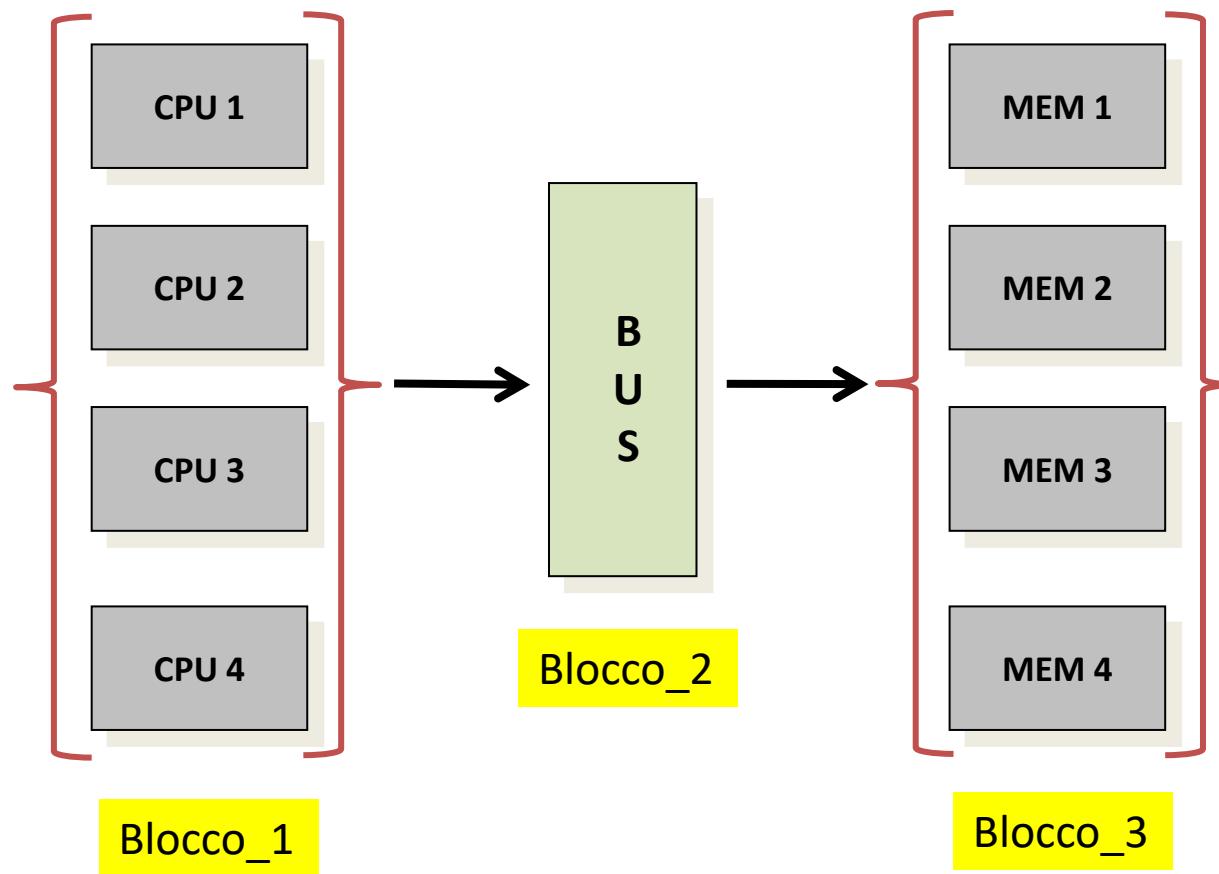
n° proc. richied.	0	1	2	3	4
probabilità	0,2	0,40	0,20	0,12	0,08

Esercizio 1: descrizione del problema e quesiti (2|2)

Indicare il modello di Affidabilità e poi calcolare:

1. la probabilità che si verifichi conflitto per il bus; quindi 2,3 o 4 processori richiedano il bus
2. la **probabilità che uno specifico processore che ne abbia fatto richiesta si veda assegnato il bus;** Bayes ?
3. la probabilità che quel processore ottenga il bus, per la prima volta, alla sua terza richiesta; prove di Bernoulli
4. la probabilità che occorrono più di 3 richieste per ottenere il bus per la prima volta;
5. la **probabilità che lo stesso processore debba fare 4 richieste per ottenere il bus per 2 volte,** sia pure non necessariamente consecutive. 2 successi in 4 tentativi: binomiale negativa

Esercizio 1: modello per il calcolo dell'affidabilità



IPOTESI:

Il sistema funziona quando funzionano almeno 1 CPU, il BUS e almeno 2 MEM

Modello risultante:

$$R_{Sis} = R_{B1} \cdot R_{B2} \cdot R_{B3} = R_{1|4} \cdot R_{BUS} \cdot R_{2|4}$$

Esercizio 1: risposta al quesito 1.

- Sia B_n l'evento: «n ($n=0,1,\dots,4$) processori richiedono il bus nello stesso clock»
- Sia E l'evento d'interesse: «almeno 2 processori nello stesso clock, ovvero conflitto»

Allora, poiché gli eventi B_n sono disgiunti, cioè tra loro incompatibili comunque li si cerchi di raggruppare (a 2 a 2, a 3 a 3 e tutti insieme), risulta:

$$P(E) = P\left(\bigcup_{n=2}^{n=4} B_n\right) = \sum_{n=2}^{n=4} P(B_n) = 0,20 + 0,12 + 0,08 = 0,40$$

Osservazioni:

Dalla tabella fornita con il problema si evince che gli eventi $[B_n, n=0,1,\dots,4]$ sono eventi diciamo elementari, in quanto le relative probabilità sono assegnate!

Gli eventi $[B_n, n=0,1,\dots,4]$ avrebbero potuto non essere elementari, ma calcolabili a partire da altri eventi elementari, cioè direttamente osservabili, ad es. $[C_i, i=1,\dots,4]$, con $C_i = \text{«il processore } i\text{-esimo richiede il bus»}$.

In tal caso, sarebbe sorta una difficoltà: $P(C_i | C_j, C_k, C_l) = ? \dots$

Come si può ridurre una tale difficoltà? Come si può eliminare del tutto?

Esercizio 1: risposta al quesito 2.

Sia A l'evento (d'interesse): «un processore richiedente ottenga il bus»

Ricorrendo al concetto di probabilità condizionata, a partire dalle probabilità note degli eventi B_n , però con $n=1, \dots, 4$, occorre calcolare le seguenti:

$$P(\hat{B}_n) = \frac{P(B_n)}{\sum_{i=1}^{i=4} P(B_i)} \quad n = 1, \dots, 4 \quad \rightarrow \quad \begin{cases} P(\hat{B}_1) = \frac{0.4}{0.8} = 0.50, & P(\hat{B}_2) = \frac{0.2}{0.8} = 0.25, \\ P(\hat{B}_3) = \frac{0.12}{0.8} = 0.15, & P(\hat{B}_4) = \frac{0.08}{0.8} = 0.10 \end{cases}$$

D'altra parte, l'ipotesi di completa casualità nell'assegnazione in caso di conflitto si traduce nelle seguenti:

$$P(A | \hat{B}_2) = 0.5, \quad P(A | \hat{B}_3) = 0.33, \quad P(A | \hat{B}_4) = 0.25 \quad [\text{a parte: } P(A | \hat{B}_1) = 1.0]$$

A questo punto, si può applicare la formula della probabilità totale:

$$\begin{aligned} P(A) &= \sum_{n=1}^{n=4} P(A | \hat{B}_n) \cdot P(\hat{B}_n) = \\ &= 1 \cdot 0.50 + 0.5 \cdot 0.25 + 0.33 \cdot 0.15 + 0.25 \cdot 0.10 = 0.70 \end{aligned}$$

funziona anche Bayes per calcolare la probabilità

Esercizio 1: risposta ai quesiti 3. e 4.

Sia E_3 l'evento (d'interesse): «il bus è ottenuto dal processore richiedente al terzo tentativo»;

Osservazione:

- $P(E_3)$ può essere calcolata come probabilità geometrica sotto le seguenti ipotesi: I tentativi sono indipendenti e ripetibili quanto si vuole; inoltre, la probabilità di successo al generico tentativo è sempre la stessa e pari alla $P(A)=0.7$ calcolata in risposta al quesito 2.

$$\Rightarrow P(E_3) = (1 - P(A)) \cdot (1 - P(A)) \cdot P(A) = 0.3 \cdot 0.3 \cdot 0.7 = 0.063$$

Sia F l'evento (d'interesse): «occorrono più di tre richieste per ottenere il bus per la prima volta»;

- Allora:

$$\begin{aligned} P(F) &= 1 - [P(E_1) + P(E_2) + P(E_3)] \\ &= 1 - [0.7 + 0.21 + 0.063] = 1 - 0.973 = 0.027 \end{aligned}$$

(potendo ripetere la richiesta all'infinito)

Osservazione: E_2 ed E_3 sono eventi tra di loro incompatibili (e pure con E_1) e, comunque, ognuno dei due risulta come «and» di altri eventi indipendenti la cui probabilità si calcola a sua volta come probabilità totale (di eventi dipendenti)!

Esercizio 1: risposta al quesito 5.

Sia $G_{2|4}$ l'evento (d'interesse): «fare 4 richieste per ottenere il bus per 2 volte, sia pure non necessariamente consecutive»;

Osservazione:

- $P(G_{2|4})$ può essere calcolata come una nuova probabilità, detta binomiale negativa, sotto le ipotesi delle prove di Bernoulli, nel caso speciale di potere ripetere le prove all'infinito e per la fortuna che non è stato chiesto un ordine preciso nell'ottenere i 2 successi sulle 4 richieste!

Siano: $G_{k-1|n-1}$ = « $k-1$ successi su $n-1$ prove» e S = «successo alla prova n -sima»

Allora: $P(G_{k|n}) = P(G_{k-1|n-1} | S) \cdot P(S)$, con $P(S) = P(A)$

e con: $P(G_{k-1|n-1} | S) = P(G_{k-1|n-1}) = \binom{n-1}{k-1} P(A)^{k-1} (1 - P(A))^{n-k}$

dunque: $P(G_{n|k}) = \binom{n-1}{k-1} P(A)^k (1 - P(A))^{n-k}$

altro non è che la dimostrazione della binomiale negativa

$$\Rightarrow P(G_{4|2}) = \binom{4-1}{2-1} 0.7^2 (1 - 0.7)^{4-2} = 3 \cdot 0.49 \cdot 0.09 = 0.132$$

Esercizio 2: descrizione del problema

Un'applicazione gestisce l'archivio dei prodotti presenti in un ipermercato e aggiorna i prezzi quotidianamente.

I dati utilizzati dall'applicazione risiedono in 4 diversi database, ognuno dei quali contenente specifiche categorie di prodotti presenti nell'ipermercato.

I database, però, possono essere soggetti a inconsistenza, cioè i prezzi dei vari prodotti non sono sempre aggiornati in modo da riflettere le ultime variazioni occorse.

Database	% di prodotti il cui prezzo non è aggiornato
Db Alimentari	2%
Db Giocattoli	7%
Db Faidate	5%
Db Abbigliam	13%

Esercizio 2: dati, quesiti e risposte (1|2)

La totalità dei prodotti presenti può essere partizionata come segue:

69% alimentari, 6% giocattoli, 2% fai da te e, infine, il 23% abbigliamento.

cosa accade se tutte le percentuali sono al 25%?

Con quale probabilità un “cliente casuale” troverà un prodotto (quasivoglia) il cui prezzo non è aggiornato?

Sia “A” l’evento (d’interesse) : “prezzo prodotto non aggiornato”
E sia “ $A|Db_i$ ”, $i=1,\dots,4$ l’evento condizionato: “prezzo non aggiornato, posto che provenga dal database i-esimo”.

Trattando le percentuali come probabilità, si può calcolare $P(A)$ ricorrendo alla formula della probabilità totale:

$$\begin{aligned} P(A) &= P(A/Db_1)*P(Db_1) + P(A/Db_2)*P(Db_2) + \\ &\quad + P(A/Db_3)*P(Db_3) + P(A/Db_4)*P(Db_4) = 0.049 \end{aligned}$$

Esercizio 2: dati, quesiti e risposte (2|2)

Calcolare la probabilità dei seguenti eventi “Dbi, i=1,...,4” = “prezzo non aggiornato estratto è proveniente dal database i”

$$P(Db1 | A) = \frac{P(A | Db1)P(Db1)}{P(A)} = \frac{0.02 * 0.69}{0.049} = 0.28$$

$$P(Db2 | A) = \frac{P(A | Db2)P(Db2)}{P(A)} = \frac{0.07 * 0.06}{0.049} = 0.086$$

$$P(Db3 | A) = \frac{P(A | Db3)P(Db3)}{P(A)} = \frac{0.05 * 0.02}{0.049} = 0.020$$

$$P(Db4 | A) = \frac{P(A | Db4)P(Db4)}{P(A)} = \frac{0.13 * 0.23}{0.049} = 0.61$$

Osservazione: la formula di Bayes permette di calcolare probabilità dette “a posteriori” (qui significa ad estrazione avvenuta) che indicano “provenienza” in questo caso, ma “colpe”, “meriti”, “responsabilità” ...

Esercizio 2bis: Alle pensiline dei bus Unical

Solo quattro autolinee regionali fanno servizio giornaliero per gli studenti pendolari dell'Unical (Federico, Lirosi, Preite e Romano). Arrivano con frequenze diverse e si può assumere che, su 20 arrivi di autobus al giorno, 8 appartengano a Preite, 6 a Romano, 4 a Lirosi e 2 a Federico.

Si stima che da un (qualsivoglia) bus di Preite scenda alle pensiline il 40% di studenti fuori corso e il 60% di studenti in corso, mentre il 20% di fuori corso scende da un bus di Romano, il 30% da un Lirosi e il 10% da un Federico.

devo prima calcolare la prob totale che sia un fuori corso.
E poi usare sempre Bayes

1. Immaginando di intervistare uno studente pendolare a caso alle pensiline, qual è la probabilità che sia uno dei fuori corso?
2. Accertato che sia fuori corso, qual è la probabilità che provenga dal crotonese, ovvero sia sceso da un bus di Romano?

Example 2.4.4 (Monty Hall Problem) This classic problem is called the “Monty Hall problem” because of the game show host Monty Hall who designed this game. There are three doors A , B , and C , and behind one of them is a car and behind the other two are goats. The contestant is asked to select any door, and the host Monty Hall opens one of the other two doors and reveals a goat. He then offers the choice to the contestant of switching to the other unopened door or keep the original door. The question now is whether the probability of winning the car is improved if she switches or it is immaterial whether she switches or not.

The answer is counterintuitive in the sense that one may be misled into thinking that it is immaterial whether one switches or not. We will analyze this problem in two different ways.

Mathematical Analysis. Let us analyze this problem from the point of view of Bayes' theorem. We shall first assume that the host has prior knowledge of the door behind which the car is located. Let us also assume that the host opens door B given that the contestant's choice is door A . The a priori probabilities of a car behind the doors A , B , and C are

$$P\{A\} = \frac{1}{3}; \quad P\{B\} = \frac{1}{3}; \quad P\{C\} = \frac{1}{3}$$

We can make the following observations. If the contestant's choice is door A and the car is behind door A , then the host will open the door B with probability of $\frac{1}{2}$ (since she has a choice between B and C). On the other hand, if the car is behind door B , then there is zero probability that she will open door B because of her prior knowledge. If the car is behind door C , then she will open door B with probability 1. Thus we can write the following conditional probabilities for the host opening door B :

$$P\{B | A\} = \frac{1}{2}; \quad P\{B | B\} = 0; \quad P\{B | C\} = 1$$

We can now calculate the total probability $P(B)$ of the host opening door B .

$$\begin{aligned}P\{B\} &= P\{B | A\}P\{A\} + P\{B | B\}P\{B\} + P\{B | C\}P\{C\} \\&= \frac{1}{2} \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} = \frac{1}{2}\end{aligned}$$

Using Bayes' theorem [Eq. (2.4.4)] we can now find the a posteriori probabilities of the car behind the doors A or C (B has already been opened by the host) conditioned on B :

$$P\{A | B\} = \frac{P\{BA\}}{P\{B\}} = \frac{P\{B | A\}P\{A\}}{P\{B\}} = \frac{\frac{1}{2} \cdot \frac{1}{3}}{\frac{1}{2}} = \frac{1}{3}$$

$$P\{C | B\} = \frac{P\{BC\}}{P\{B\}} = \frac{P\{B | C\}P\{C\}}{P\{B\}} = \frac{\frac{1}{2} \cdot \frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}$$

Similar analysis holds for other cases of opening the doors A or C . We can see from the result above that the contestant *must* switch if she wants to double her probability of winning the car.

TABLE 2.4.1

First Choice	Host Opens	Second Choice	Probability
A	B	A	$P\{B\} = \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6}$
A	C	A	$P\{C\} = \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6}$
B	C	B	$P\{C\} = 1 \cdot \frac{1}{3} = \frac{1}{3}$
C	B	B	$P\{B\} = 1 \cdot \frac{1}{3} = \frac{1}{3}$

TABLE 2.4.2 For not switching

First Choice	Host Opens	Second Choice	Win or Loss	Probability
A	B	A	Win	$P\{B\} = \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6}$
A	C	A	Win	$P\{C\} = \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6}$
B	C	B	Loss	$P\{C\} = 1 \cdot \frac{1}{3} = \frac{1}{3}$
C	B	B	Loss	$P\{B\} = 1 \cdot \frac{1}{3} = \frac{1}{3}$

TABLE 2.4.3 For switching

First Choice	Host Opens	Second Choice	Win or Loss	Probability
A	B	C	Loss	$P\{B\} = \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6}$
A	C	B	Loss	$P\{C\} = \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6}$
B	C	A	Win	$P\{C\} = 1 \cdot \frac{1}{3} = \frac{1}{3}$
C	B	A	Win	$P\{B\} = 1 \cdot \frac{1}{3} = \frac{1}{3}$

SECONDA PARTE ANALISI PROBABILISTA

Variabili aleatorie e distribuzioni

Dato uno spazio di risultati (anche non numerico), Ω di un esperimento aleatorio una variabile aleatoria X è una funzione che associa ad ognuno dei risultati ω numeri reali. Per ipotesi i risultati dell'esperimento saranno numeri. Preso un insieme di risultati, se si studia il suo andamento temporale andremo a trattare di un processo stocastico.

Variabile aleatoria: $X : \Omega \rightarrow \mathbb{R}$ Processo stocastico: $X_\omega : \mathbb{R} \rightarrow S$ con $\omega \in \Omega, S \subseteq \mathbb{R}$

Una volta eseguito l'esperimento, il valore $X(\omega) = x$ risulta completamente determinato e viene detto **realizzazione** (come definito anche dopo)

Fortunatamente, in questa sede, è sufficiente concentrarsi su risultati che sono già

- 1) numeri interi non negativi: come nel caso di "k successi su n prove" oppure nel caso di "esattamente/al più/almeno k posti occupati in un buffer";
- 2) numeri reali non negativi: come nel caso della "durata del servizio reso da una risorsa ad un utente", oppure della "durata dell'attesa dell'utente prima di ottenere la risorsa", oppure del "tempo al guasto di un componente/sistema".

Insomma si riduce di molto lo spazio dei risultati

Caso 1) → variabili aleatorie discrete (a valori nel discreto).

Caso 2) → variabili aleatorie continue (a valori nel continuo).

Nel caso 2) saranno considerate variabili aleatorie che fanno corrispondere il numero reale x - detto **realizzazione** della variabile aleatoria X - all'evento "la durata del servizio/attesa (o del fenomeno qualsivoglia) è non superiore a x ". Dunque tutti i possibili risultati nel continuo (precise durate del fenomeno) compresi nell'intervallo reale $[0,x]$ verranno posti in corrispondenza alla realizzazione x della X e ciò verrà indicato con la notazione " $X \leq x$ ", intendendo che la X è la variabile aleatoria che rappresenta l'insieme delle possibili durate del fenomeno. L'insieme dei numeri (ovvero le durate) sarà considerato positivo, cioè da 0 a infinito perché durate negative non hanno senso

Tornando al caso discreto, si consideri la variabile aleatoria, K , che rappresenta "il numero di possibili successi su n prove di Bernoulli". Una sua realizzazione, k , corrisponde all'evento "k successi su n prove", del quale è stata già calcolata la probabilità:

$$\binom{n}{k} p^k q^{n-k}$$

È allora naturale definire distribuzione di probabilità della variabile aleatoria discreta K la funzione $P : [0,1,2,\dots,n] \rightarrow [0,1] \in \Re$ che ripartisce i singoli valori di probabilità alle singole realizzazioni, $k=0,1,\dots,n$, della K rispettando gli assiomi della probabilità.

$$P(K = k | n) \equiv P_K(k | n) \triangleq \binom{n}{k} p^k q^{n-k}, \quad k = 0,1,\dots,n;$$

La funzione appena definita (nota come *legge binomiale*, $B(n,p)$) è un assegnamento probabilistico sullo spazio delle realizzazioni della K perché risulta:

- 1) $0 \leq \binom{n}{k} p^k q^{n-k} \leq 1 \quad \forall k = 0,1,\dots,n$ dunque probabilità compresa tra 0 e 1
- 2) $\sum_{k=0}^n \binom{n}{k} p^k q^{n-k} = (p+q)^n = 1 \quad$ sarebbe il binomio di Newton
appunto somma delle probabilità pari a 1

Ad un'altra legge di probabilità ben nota, *la legge geometrica*, si può pervenire introducendo la variabile aleatoria, S, che rappresenta "l'insieme dei numeri di prova in corrispondenza dei quali si può ottenere il primo successo, nell'ambito di una successione indefinita di prove di Bernoulli"

Infatti, partendo dall'evento "il primo successo al k-mo tentativo"
si pone:

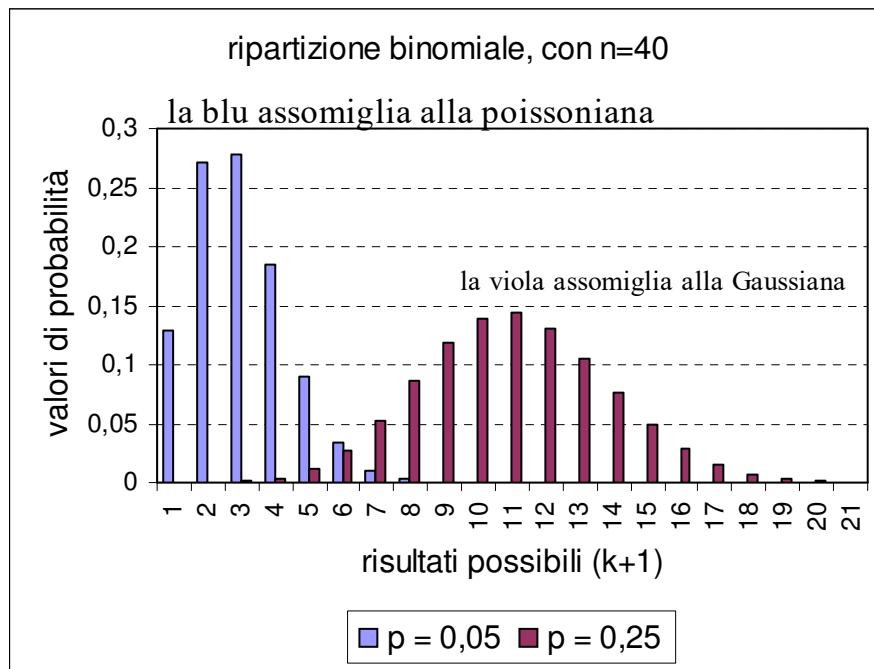
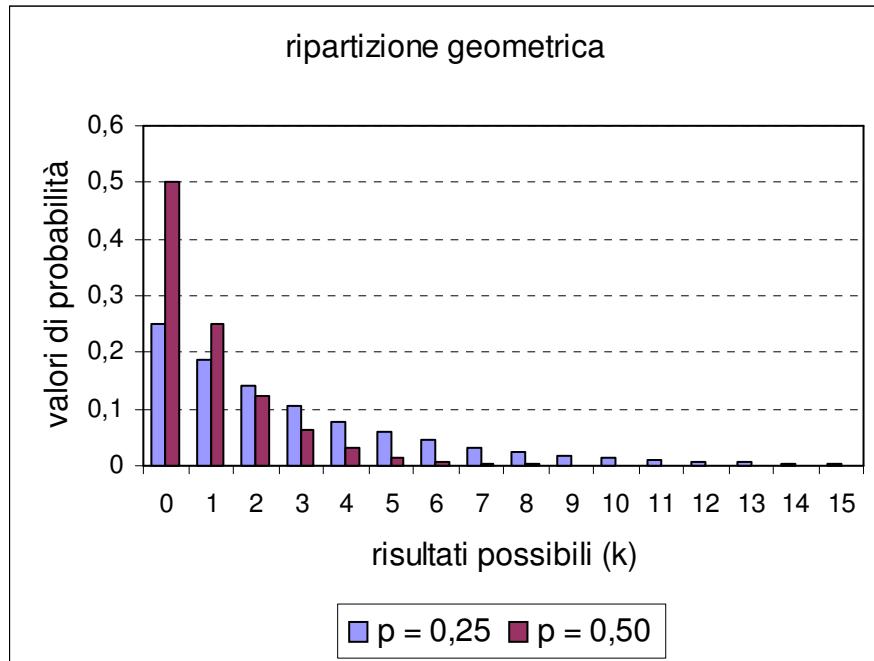
$$P(S = k | \infty) \equiv P_S(k | \infty) \triangleq q^{k-1} p, \quad k = 1, \dots, n, \dots \quad \begin{array}{l} \text{si parte da 1 perché deve} \\ \text{accadere sto successo} \end{array}$$

e si verifica che:

$$0 \leq q^{k-1} p \leq 1 \quad \forall k = 1, \dots, n, \dots \quad \text{e} \quad \sum_{k=1}^{\infty} q^{k-1} p = \dots = p/(1-q) = 1$$

anche qui la probabilità è tra 0 e 1 e la somma di tutte le probabilità ritorna 1

Rappresentazioni di Geometrica e Binomiale



Esempio di costruzione di un modello probabilistico di ripartizione

A partire dalla seguente (lo sviluppo in serie della funzione esponenziale positiva)

$$\sum_{n=0}^{\infty} \frac{x^n}{n!} = e^x \quad x > 0$$

E moltiplicando ambo i membri per e^{-x} si ottiene

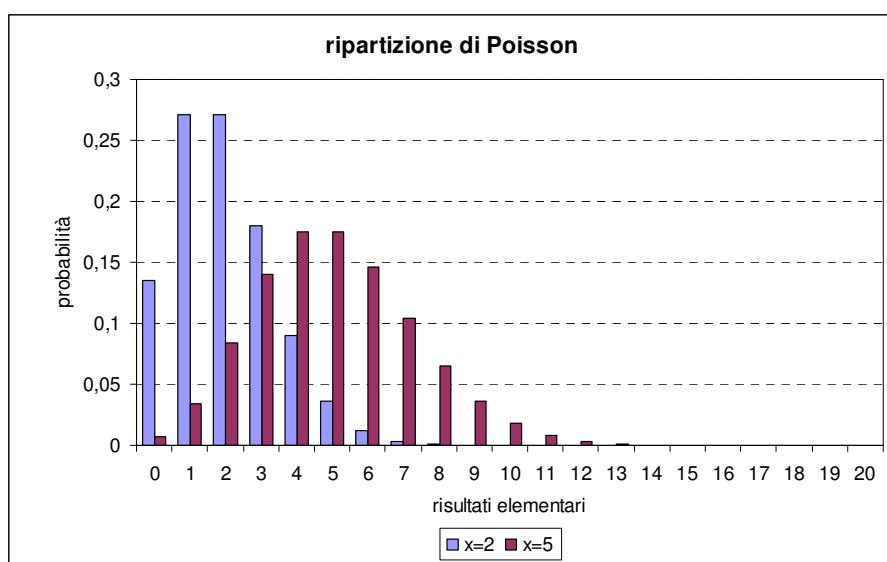
$$\sum_{n=0}^{\infty} \frac{x^n}{n!} e^{-x} = 1 \quad x > 0$$

Osservando che il termine dentro la sommatoria $\frac{x^n}{n!} e^{-x} \geq 0$ rispetta l'assioma (A1) della probabilità (di non negatività), mentre la sommatoria uguale ad 1 può essere letta come la verifica dell'assioma (A3), è immediato concludere che la funzione che scriviamo

$$P(N = n) = \frac{x^n}{n!} e^{-x}$$

dove $n = 0, 1, 2, \dots$ sono i risultati possibili della variabile aleatoria N ed x è un parametro di detta funzione.

La funzione è un nuovo modello di ripartizione (noto come modello di Poisson).



IMPORTANTE
Distribuzione(cumulativa)di probabilità

Per introdurre il concetto e la definizione di distribuzione cumulativa di probabilità per una variabile aleatoria discreta, è utile fare riferimento all'esperimento aleatorio che consiste nell'osservazione di un buffer e al risultato "*al più i primi m posti risultano occupati*", sotto l'ipotesi che il numero di posti disponibili sia praticamente illimitato.

Introducendo la variabile aleatoria K per rappresentare tutti i risultati (elementari) del tipo "*i primi k posti risultano occupati*", con $k = 0, 1, 2, \dots, n, \dots$, si potrebbe adottare la legge geometrica per ripartire valori di probabilità alle possibili realizzazioni:

$$P(K = k | \infty) \equiv P_K(k | \infty) \hat{=} q^k p, \quad k = 0, 1, \dots, n, \dots$$

assumendo che il fattore "p" sia la probabilità di trovare un posto libero dopo k posti trovati tutti occupati (ognuno con la stessa probabilità "q").

A questo punto, per calcolare la probabilità del risultato d'interesse, "*al più i primi m posti risultano occupati*", basta cumulare (sommare) le probabilità ripartite dalla legge geometrica su tutti i casi elementari compresi fra "0" ed "m".

Dunque:

$$\text{Probabilità("al più i primi m posti occupati")} = \sum_{k=0}^m q^k p \hat{=} C(m)$$

L'importanza della probabilità appena calcolata giustifica abbastanza la definizione della funzione C , sul dominio dei possibili valori di m , quale distribuzione cumulativa associata ad una data (qui era la geometrica) distribuzione di probabilità nel discreto.

Affinché la C sia ben posta deve sempre risultare:

$$0 \leq C(m) \leq 1 \quad \forall m \quad \text{e} \quad C(m) \rightarrow 1 \quad \text{se } m \rightarrow \infty$$

$C(m)$ resta una probabilità!

IMPORTANTE

Funzione di distribuzione

Con riferimento al caso continuo, l'utilità di misurare la probabilità che "la durata di un certo fenomeno sia non superiore ad un valore reale positivo, x , fissato" conduce alla definizione di funzione di distribuzione associata alla variabile continua X :

$$F_X : [0 \leq x < \infty] \rightarrow [0,1] \in \Re \quad \begin{array}{l} \text{codominio da 0 a 1} \\ \text{perché è una probabilità} \end{array}$$

La probabilità che la durata sia al più pari a x è pari a

$$P(X \leq x) \equiv F_X(x) \hat{=} \int_{t=0}^x f_X(t) dt \quad \begin{array}{l} \text{l'estremo superiore è } x, \text{ dunque} \\ f_X \text{ è una funzione integrale} \end{array}$$

Proprietà di F_X

l'integrale sarebbe la somma delle probabilità

$$0 \leq F_X(x) \leq 1, \quad 0 \leq x < \infty$$

Banalmente la cumulativa deve essere nulla all'inizio perché non hai accumulato nulla!

$$F_X(x_2) \geq F_X(x_1), \quad x_2 \geq x_1 \quad \begin{array}{l} \text{è monotona} \\ \text{se } x=0 \text{ allora la funzione è pari a 0} \end{array}$$

"alla fine hai terminato"

$$F_X(x \rightarrow 0) \rightarrow 0, \quad F_X(x \rightarrow \infty) \rightarrow 1 \quad \begin{array}{l} \text{ovvero probabilità è} \\ \text{pari 1} \end{array}$$

La f_X è detta funzione densità di probabilità ed ha il compito di ripartire valori di probabilità alle realizzazioni, continue e non negative, della X .

Dunque detta in parole povere, densità di probabilità significa: quanta probabilità c'è in punto? Ad esempio 0.4 a $t=1$, 1.0 a $t=0$.

A livello elementare, si può assumere che la densità di probabilità sia una funzione continua e quindi integrabile secondo Jordan, nonché ottenibile per derivazione in base alla seguente $f_X : [0 \leq x < \infty] \rightarrow [0, M] \in \Re^+, \quad M > 1$

M può essere molto maggiore di 1
 $f_X(x)dx$ è una probabilità
 che il fenomeno duri proprio x a
 meno di un dx . Esempio probabilità
 che un download arriva a 30s
 e finisce tra 30s e 31s. Quindi
 $dx = 1s$.

$$0 \leq f_X(x) \mid \int_{x=0}^{\infty} f_X(x) dx = 1$$

la derivata si usa perché non si può sapere l'istante esatto

$$\frac{d}{dx} F_X(x) = \frac{d}{dx} \left(\int_{t=0}^x f_X(t) dt \right) = f_X(x)$$

in pratica dalla distribuzione
 passo alla densità tramite derivata

la derivata sarebbe il teorema fondamentale

$$P(x_1 \leq X \leq x_2) = \int_{t=x_1}^{t=x_2} f_X(t) dt, \quad \geq 0 = F(x_2)$$

data la probabilità tra x_1 e x_2

$$P(x \leq X \leq x + \Delta x) = F_X(x + \Delta x) - F_X(x)$$

$$\cong f_X(x)\Delta x, \quad \Rightarrow P(X = x) = 0 \quad \begin{array}{l} \text{risulta nulla la probabilità} \\ \text{associabile ad un punto} \end{array}$$

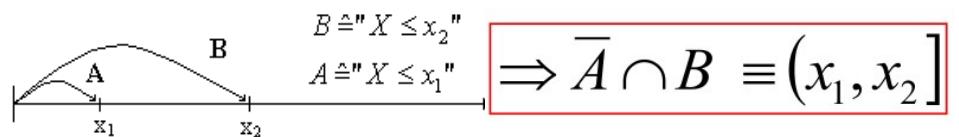
probabilità che la
 variabile aleatoria
 X ha realizzazione
 comprese tra: $x, x + \Delta x$



Punti di discontinuità nella F_X

Si farà riferimento a due risultati, " $X \leq x_1$ " e " $X \leq x_2$ " con $x_1 \leq x_2$, di un esperimento aleatorio inteso come l'osservazione della durata di un certo fenomeno d'attesa, per illustrare il significato e l'utilità della presenza di punti di discontinuità di (salto) prima specie nella classe delle funzioni reali usate per modellare distribuzioni (cumulative) di probabilità nel continuo.

A tal fine, è conveniente far corrispondere gli eventi A e B, rispettivamente, ai due suddetti risultati di durata nel continuo e ragionare sulla seguente figura:



Partendo dalle seguenti relazioni:

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

si riconosce che:

$$P(\bar{A} \cap B) \equiv P(x_1 < X \leq x_2)$$

mentre

$$P(B) - P(A) = P(X \leq x_2) - P(X \leq x_1) = F_X(x_2) - F_X(x_1)$$

quindi

$$P(x_1 \leq X \leq x_2) = F_X(x_2) - F_X(x_1)$$

A questo punto, il fatto che $x_2 \rightarrow x_1$ può essere visto in termini dell'evento $(\bar{A} \cap B)$ - corrispondente ad un intervallo - che tende ad un evento limite corrispondente ad un solo punto della semiretta reale, (Ω è lo spazio dei reali non negativi).

Dunque, l'ipotesi di continuità della F_X nel punto x_1 implica che è nulla la probabilità associata a quel punto (nota anche come *evento limite*) e, viceversa, rinunciando all'ipotesi di continuità della F_X in x_1 , il salto di discontinuità corrisponderebbe alla probabilità finita e non nulla che si vuole associare al punto x_1 .

Per cogliere l'importanza pratica della questione discussa, si può pensare all'esigenza di valutare la probabilità finita di un tempo d'attesa nullo: in tal caso la funzione di distribuzione è continua a destra dello zero.

Prima distribuzione e la più semplice

La distribuzione esponenziale

E' assai importante nell'analisi probabilistica perché **della durata "completamente casuale"** di un fenomeno.

Prima di ricavare l'espressione formale della distribuzione esponenziale, occorre premettere che per "durata completamente casuale" s'intende quella rappresentata da una variabile aleatoria, X, per la quale valga la seguente:

$$Pr\{X \leq t + \Delta t | X \geq t\} = Pr\{X \leq \Delta t\}$$

probabilità che l'evento finisce prima di $t + \Delta t$
sapendo che arriva a t sicuramente è pari
alla probabilità che finisce prima di Δt

Proprietà di assenza di memoria perché NON DIPENDE dalla sua storia

sia $X =$ vita del fenomeno, $Y = X - t$ durata residua del fenomeno e $t =$ "età" del fenomeno

$$Pr\{Y \leq \Delta t | X \geq t\} = Pr\{X \leq \Delta t\}$$

cioè che **la durata residua del fenomeno sia indipendente dall'età, ma distribuita identicamente alla durata vera e propria**. Per un essere umano questo purtroppo non è vero.

Dimostrazione del come arrivare alla distribuzione esponenziale. Ipotesi assenza di memoria Partendo dalla ben nota formula della probabilità congiunta ed applicando la condizionata

$$Pr\{t \leq X \leq t + \Delta t\} = Pr\{X \leq t + \Delta t | X \geq t\} Pr\{X \geq t\}$$

il complemento al minore
uguale è maggiore uguale
per questo ho uguale
imponendo la (*)

$$= Pr\{X \leq \Delta t\} Pr\{X \geq t\}$$

Parto dal rapporto incrementale
per arrivare all'equazione differenziale

dunque:

$$\frac{F(t + \Delta t) - F(t)}{\Delta t} = \frac{F(\Delta t)}{\Delta t} [1 - F(t)] = \frac{F(0 + \Delta t) - F(0)}{\Delta t} [1 - F(t)]$$

$F(0) = 0$ data la proprietà vista
in precedenza

e passando a $\lim_{\Delta t \rightarrow 0}$ si ottiene una semplice equazione differenziale:

$$\frac{d}{dt} F(t) = \frac{d}{dt} F(t)|_{t=0} [1 - F(t)]$$

è un'equazione differenziale

che ammette come soluzione appunto la legge esponenziale:

$$F(t) = 1 - e^{-\lambda t}, \quad \text{dove} \quad \lambda = \frac{d}{dt} F(t)|_{t=0}$$

lambda è la derivata nell'origine
ovvero che la probabilità finisce subito

Si può vedere l'assenza di memoria come assenza di usura (fattore endogeno), ed il fatto che l'esponenziale termini per qualche fattore esogeno

Un esempio di un fenomeno che non invecchia (privo di memoria) è una richiesta di un servizio.

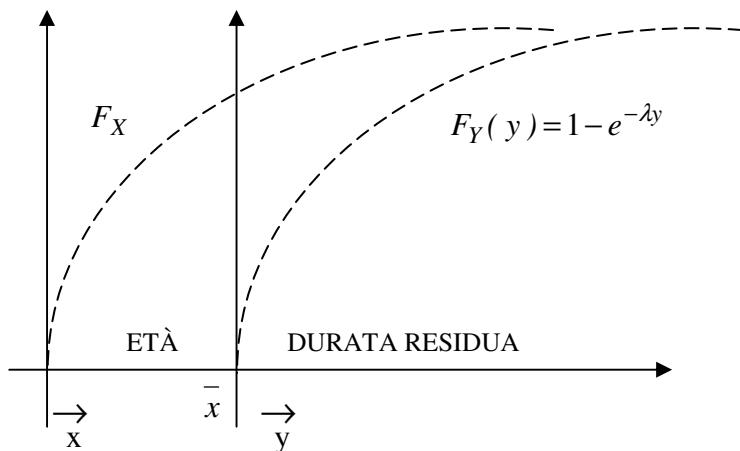
Il ragionamento appena sviluppato porta a definire la distribuzione esponenziale:

$$F_X(x) \hat{=} \begin{cases} 1 - e^{-\lambda x} & x \geq 0 \\ 0 & \text{altrimenti} \end{cases} \quad \text{Non è simmetrica}$$

con la seguente densità esponenziale:

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

e si dice che per essa vale la proprietà di "assenza di memoria" illustrata dalla seguente figura:



L'esponenziale è l'UNICO che gode di assenza di memoria, che può essere visto come trade-off

DIMOSTRAZIONE DELLA PROPRIETÀ DI ASSENZA DI MEMORIA

Ipotesi: X ha una distribuzione esponenziale

Tesi: $Pr\{Y \leq y | X \geq \bar{x}\} = Pr\{X \leq y\}$ dove $Y = X - \bar{x}$

Prova:

applicando congiunta e condizionata

$$\begin{aligned} P(Y \leq y | X \geq \bar{x}) &= P(X - \bar{x} \leq y | X \geq \bar{x}) = \frac{P(\bar{x} \leq X \leq \bar{x} + y)}{P(X \geq \bar{x})} = \frac{F(\bar{x} + y) - F(\bar{x})}{1 - F(\bar{x})} = \\ &= \frac{1 - \exp\{-\lambda(\bar{x} + y)\} - (1 - \exp\{-\lambda\bar{x}\})}{\exp\{-\lambda\bar{x}\}} \\ &= 1 - \exp\{-\lambda y\} = F_X(y) \hat{=} P(X \leq y) \end{aligned}$$

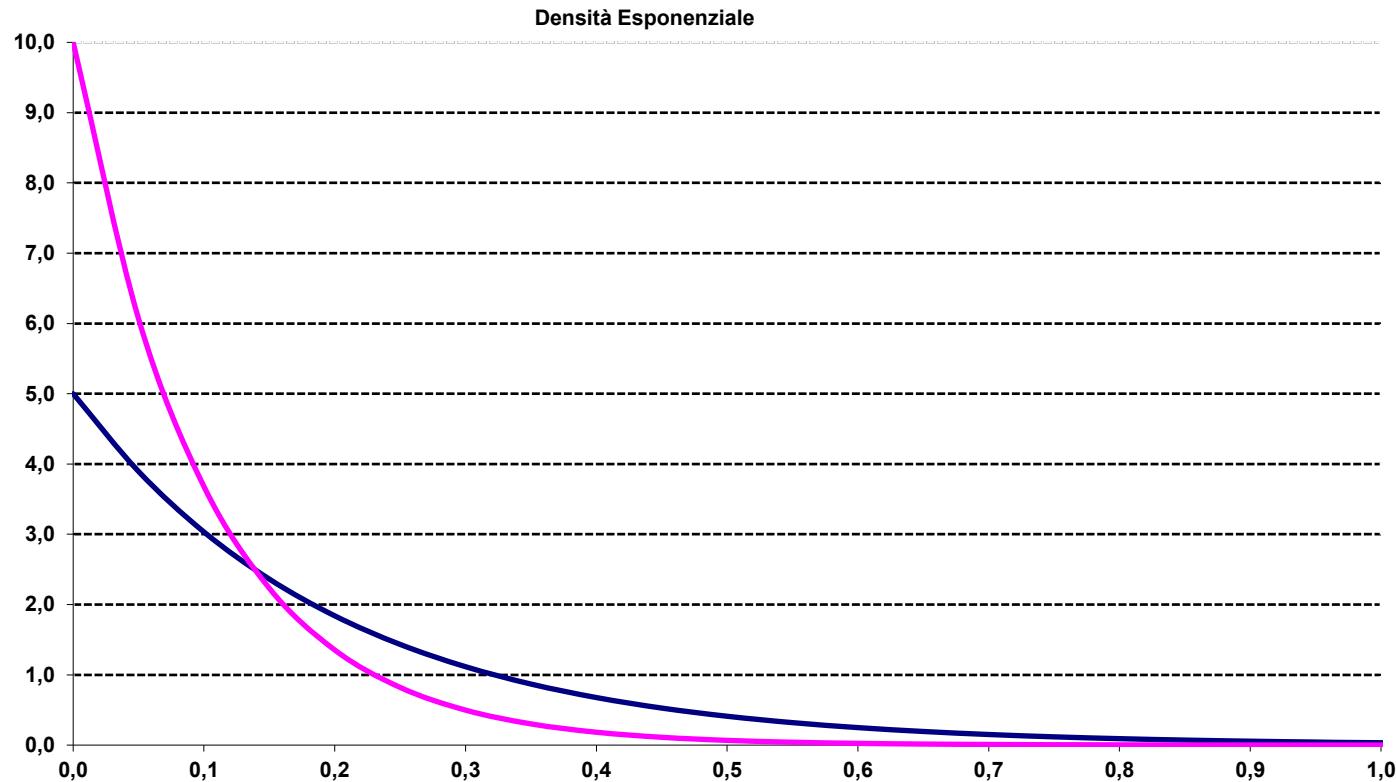
La funzione di distribuzione della vita residua coincide con quella della vita intera, però traslata di una quantità pari all'età.

La funzione densità esponenziale

la blu dura un po' di più della viola
ma comunque tende a finire

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

esponente negativo
altrimenti sarebbe
crescente all'infinito
e non tendente a 0



l'area sottesa dalla
curva è la probabilità

Affidabilità come misura di durata o di sopravvivenza

Un'immediata applicazione del concetto di funzione di distribuzione nell'analisi probabilistica di sistemi è quella che conduce alla definizione della funzione Affidabilità come misura cumulativa riferita all'intervallo $[0, t]$. (il modello esponenziale si utilizza per l'affidabilità)

A partire dalla variabile aleatoria $X = "tempo al guasto"$, si può definire la funzione affidabilità, $R(t)$, di un componente (o di un sistema):

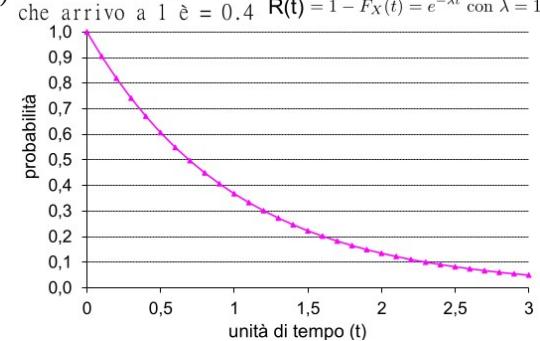
$$R(t) \doteq P(X \geq t) = 1 - P(X \leq t) = 1 - F_X(t) \quad \begin{array}{l} F(t) \text{ è una cumulativa, dunque } R(t) \text{ è} \\ \text{una cumulativa} \\ \Leftarrow \text{perché continua} \end{array}$$

come la probabilità che il componente/sistema abbia una durata di funzionamento che copra l'intervallo $[0, t]$ senza interruzioni, posto che l'istante "0" è l'istante di attivazione del funzionamento. (vuol dire componente collaudato) Ad esempio la probabilità che arrivo a 1 è = 0,4 $R(t) = 1 - F_X(t) = e^{-\lambda t}$ con $\lambda = 1$

Per ipotesi, dunque, deve risultare:

$$R(t \rightarrow 0) \rightarrow 1$$

Proprietà caratteristica è invece:



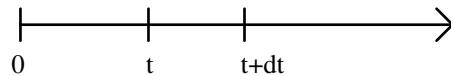
Probabilità del guasto

Faccio la derivata di $R(t)$ così vedo cosa accade in un punto, dunque da

$$\frac{d}{dt} R(t) = \frac{d}{dt} (1 - F_X(t)) = -f_X(t) \quad \begin{array}{l} (\text{essendo l'affidabilità decrescente}) \\ (\text{mi aspetto il meno nel}) \\ (\text{risultato della derivata ricordalo}) \end{array}$$

ottengo che il differenziale di $R(t)$ $f_X(t)dt = -dR(t)$

Tale risultato $-dR(t) = f_X(t)dt$ è noto come "probabilità di un guasto in ($t, t+dt$)".



Più interessante è avere una misura della:

"probabilità di guasto in ($t, t+dt$) | funzionamento fino a t "

che si ottiene introducendo il concetto di tasso di guasto.

A differenza della funzione affidabilità, il tasso di guasto è una misura istantanea e non di tutto l'intervallo precedente il punto considerato. Precisamente, vuole quantificare la

tendenza al guasto che è possibile associare ad un determinato componente/sistema

funzionante in quello stesso istante. (Probabilità che al prossimo istante dopo t si guasti)

Al variare di "t", è rappresentato dalla funzione $h(t)$, definita nel modo seguente:

Dunque, in $h(t)$ per ipotesi si

raggiungerà sicuramente

l'istante t, in particolare $h(t) \triangleq \lim_{\Delta t \rightarrow 0} \frac{\text{Prob}(t < X \leq t + \Delta t | X > t)}{\Delta t}$

ignorando delta t
e sostituendo con
congiunta e condizionante

che si raggiunge si chiama età

$$\text{risulta che } h(t) = \lim_{\Delta t \rightarrow 0} \frac{\text{Prob}(t < X < t + \Delta t)}{\Delta t P(X > t)} = \lim_{\Delta t \rightarrow 0} \frac{F_X(t + \Delta t) - F_X(t)}{\Delta t R(t)}$$

rapporto incrementale

$$\text{ovvero } f_X(t) = \lim_{\Delta t \rightarrow 0} \frac{F_X(t + \Delta t) - F_X(t)}{\Delta t (1 - F_X(t))} = \frac{f_X(t)}{1 - F_X(t)} = \frac{\lambda e^{-\lambda t}}{1 - (1 - e^{-\lambda t})} = \lambda$$

λ del esponenziale, è il rischio del guasto ovvero il parametro che dice quanto velocemente finisce una cosa. Sostituendo correttamente $1 - F_X(t) = R(t)$

si ottiene che $h(t) = \lim_{\Delta t \rightarrow 0} \frac{R(t) - R(t + \Delta t)}{\Delta t R(t)} = \frac{-R'(t)}{R(t)}$ che messa assieme a $f_X(t)dt = -dR(t)$
h(t) sarebbe una velocità porta a $R(t) \cdot h(t)dt = f_x(t)dt$
h(t) dt è probabilità Dunque la prob. che X duri proprio x a meno di un dx
è la congiunta tra affidabilità e tasso di guasto

Da quest'ultima è possibile ottenere una relazione che lega il tasso stesso all'affidabilità,

integrando il primo e il secondo membro tra 0 e u :

$$\int_0^u h(t)dt = \int_0^u -\frac{R'(t)}{R(t)}dt = -\int_{R(0)}^{R(u)} \frac{dR(t)}{R(t)}$$

ovvero, poiché $R(0) = 1$,

$$-\ln R(u) = \int_{t=0}^u h(t)dt,$$

ed in conclusione

funzione di rischio cumulato

$$\Rightarrow R(u) = \exp \left[- \int_{t=0}^u h(t)dt \right] \Rightarrow 1 - F_X(u) = \exp \left[- \int_0^u h(t)dt \right]$$

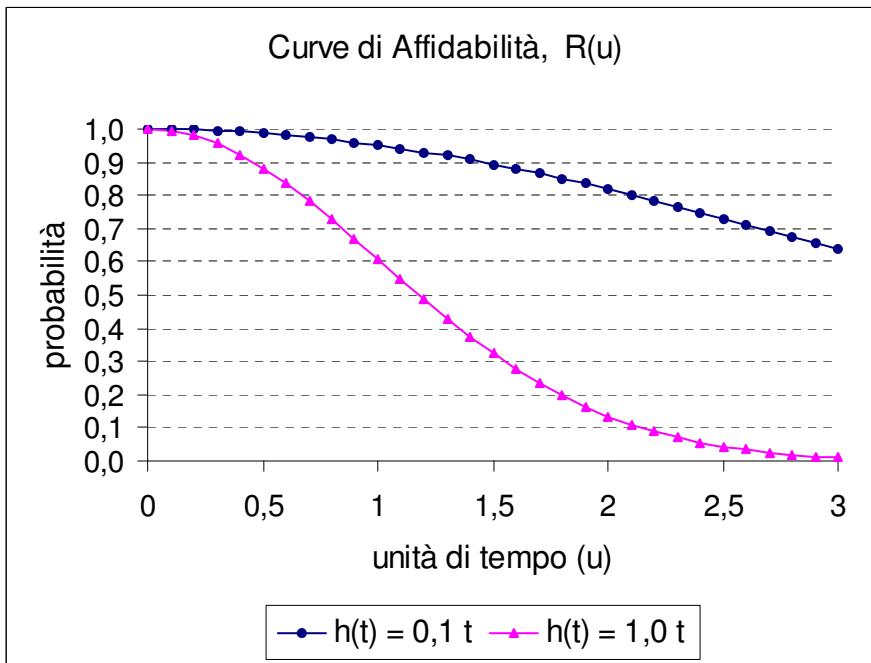
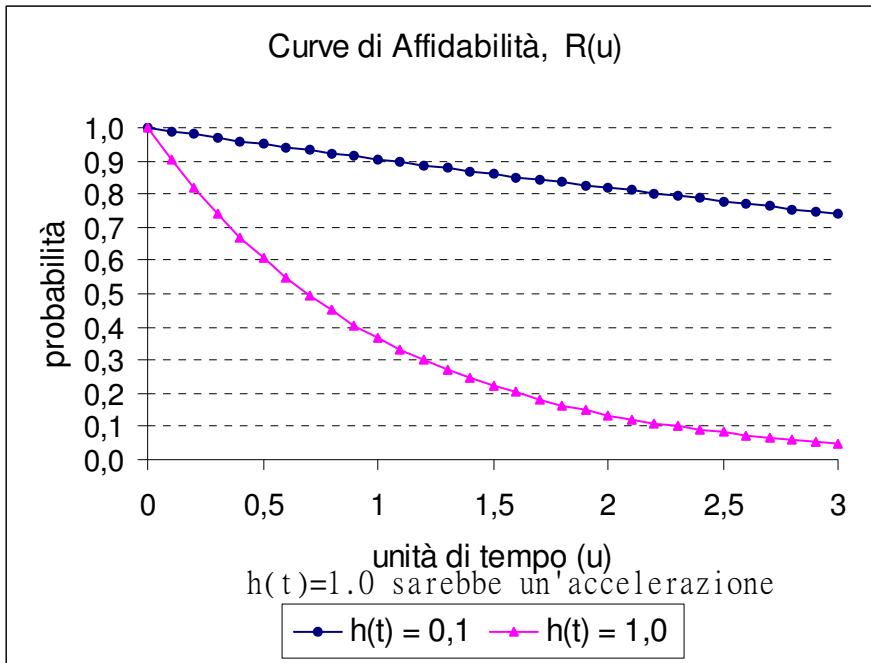
Caso particolare: $h(t) = \lambda \Rightarrow R(u) = \exp\{-\lambda \cdot u\}$

La funzione $h(t)$ è il "modello di guasto" alla base della funzione affidabilità.

(se mi dai la velocità con cui si consuma un fenomeno io so dove ti trovi)

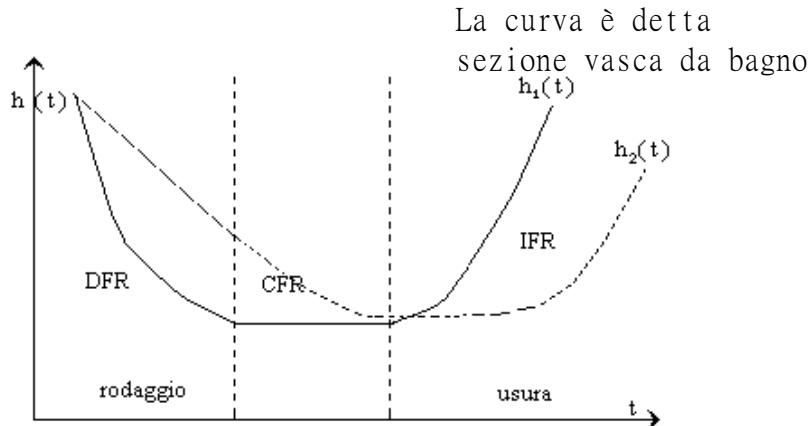
Il rischio di guasto può quindi essere o meno costante. Se è costante l'affidabilità si modella come una semplice esponenziale altrimenti al posto dell'esponente si trova l'integrale della funzione di rischio

Tasso di guasto costante e tasso lineare



Modelli di guasto

DFR=decreasing failure rate
 CFR=constant failure rate
 IFR= increasing failure rate



(DECREASING-CONSTANT-INCREASING FAILURE RATE)

Il grafico raffigura andamenti verosimili che possono essere attribuiti alla funzione $h(t)$

$h_1(t)$: limitato periodo di rodaggio seguito da un periodo a tasso di guasto costante, detto vita utile, e poi dal periodo di usura;
 $h_2(t)$: periodo di rodaggio più lungo e poi vita utile e usura.

Weibull è un'estensione del caso precedente particolare $h(t) = \lambda$

Il modello in qualche maniera rappresenta la curva del grafico

Il modello di Weibull : $h(t) \doteq \lambda \alpha t^{\alpha-1}$, $\lambda > 0, \alpha > 0$.

Proprietà: $(\alpha > 1 \Rightarrow IFR, \alpha < 1 \Rightarrow DFR, \alpha = 1 \Rightarrow CFR)$

Grazie alla relazione: $R(u) = \exp \left[- \int_{t=0}^u h(t) dt \right] = 1 - F_X(u)$

Porta alla definizione delle funzioni densità e distribuzione di Weibull :

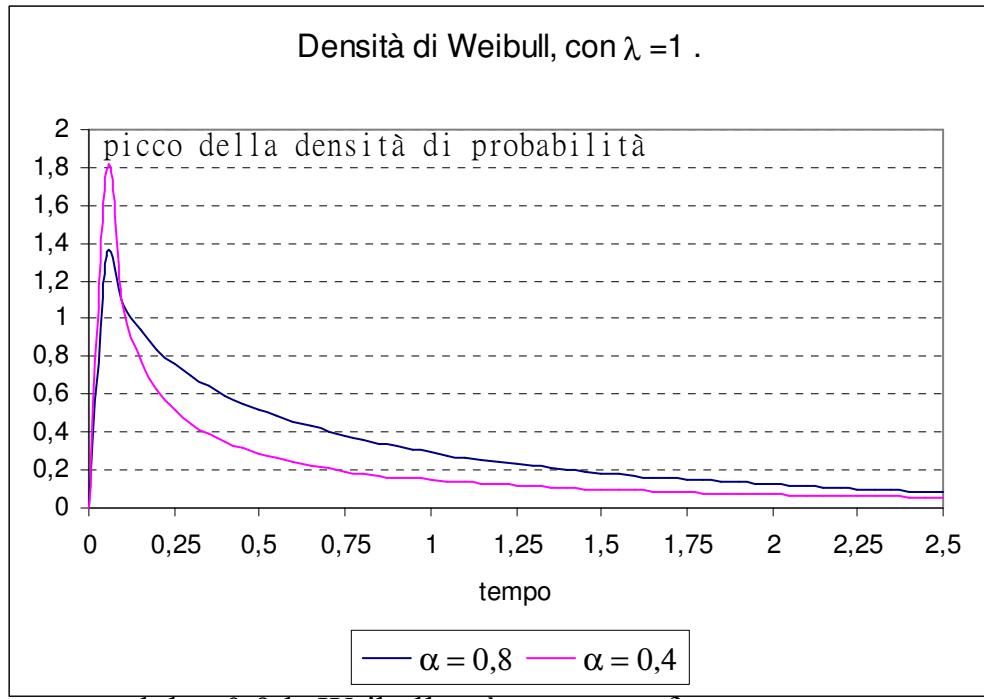
$$f_X(t) \doteq \lambda \alpha t^{\alpha-1} \exp \left\{ - \lambda t^\alpha \right\}, \quad \lambda > 0, \alpha > 0, \quad t \geq 0$$

$$F_X(t) \doteq 1 - \exp \left\{ - \lambda t^\alpha \right\}, \quad \lambda > 0, \alpha > 0, \quad t \geq 0$$

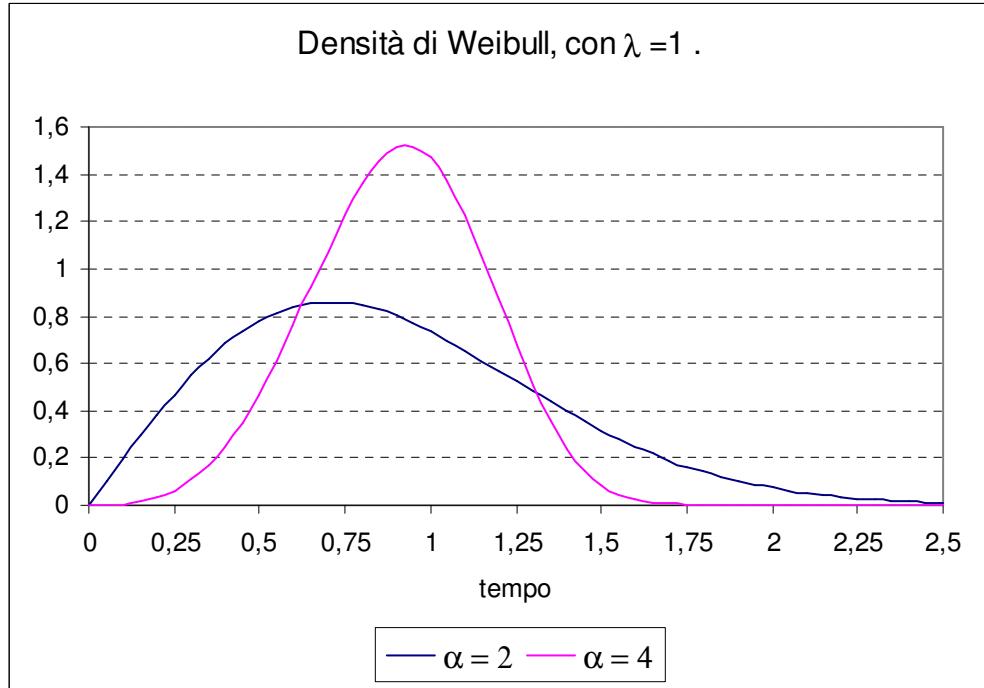
Weibull può descrivere la durata di vita per un fenomeno la cui "probabilità di morire" può variare nel tempo, a differenza della distribuzione esponenziale che prevede tassi di guasto costanti nel tempo, Weibull viene usato per affidabilità con tasso di guasto non costante

Rappresentazione grafica del modello di Weibull

utilizzato nella mortalità infantile



con alpha=0.9 la Weibull può essere confusa con un esponenziale



con $\alpha > 1$ si creano
curve totalmente diverse

Gli istogrammi con Weibull sono complicati e se $\alpha \leq 1$, risultano ancora più complicati

Media, varianza e momenti di una variabile aleatoria

Con il concetto di media di una variabile aleatoria si vuole associare un singolo numero alla "forma" secondo la quale la probabilità è ripartita fra le singole realizzazioni della variabile stessa.

A tal fine, con riferimento ad una variabile aleatoria discreta e non negativa, ad es. N a valori in $[0,1,2,\dots,n,\dots]$ si definisce momento del primo ordine, o valore atteso, o speranza matematica, o, più semplicemente media e si indica con $E[N]$ il valore finito della seguente serie, quando esiste:

Il valore atteso di una v.a. è la media pesata dei valori x_i che può assumere la variabile, moltiplicati per la probabilità che si verifichi

$$\sum_{n=0}^{\infty} n P_N(n) \triangleq E[N]$$

Una media aritmetica stima un valore atteso ecco perchè sono sinonimi

Con riferimento a X non negativa e continua a valori in $[0, \infty)$, si definisce alla stessa maniera e si indica con $E[X]$ il valore numerico finito del seguente integrale, quando esiste:

L'integrale è la somma pesata del prodotto tra realizzazione e probabilità che si verifichi

$$\int_{x=0}^{\infty} x f_X(x) dx \triangleq E[X]$$

deve convergere

$E[X]$ è un numero!

Proprietà:

$E[\text{numero}] = \text{numero}$ quindi

$E[E[X]] = \text{numero!}$

Generalizzando, sono detti momenti di ordine "k" della X e della N, rispettivamente, i seguenti:

$$E[X^k] \triangleq \int_{x=0}^{\infty} x^k f_X(x) dx, \quad E[N^k] \triangleq \sum_{n=0}^{\infty} n^k P_N(n)$$

Con il concetto di varianza, invece, si vuole introdurre una misura numerica dell'entità della dispersione delle realizzazioni possibili attorno alla media.

Dunque si definisce: è lo scostamento dal valore atteso: appunto valore atteso x meno qualcosa

Una variabile aleatoria con varianza nulla non ha senso in quanto perderebbe l'aleatorietà

$$VAR[X] \triangleq E[(x - E[X])^2] = \int_{x=0}^{\infty} (x - E[X])^2 f_X(x) dx$$

non si preoccupa dove si discostano gli scarti della differenza

$$\text{nel caso discreto } VAR[N] \triangleq E[(n - E[N])^2] = \sum_{n=0}^{\infty} (n - E[N])^2 P_N(n)$$

Si capisce anche che la varianza è la media pesata degli scarti

P.S. calcolare il valore atteso di Weibull è complicato dato che devi fare un integrale (esula dal corso)

La varianza di una somma di variabili aleatorie è pari alla somma delle varianze di ciascuna se assumiamo che quelle variabili siano indipendenti mentre per il valore atteso di una somma di variabili aleatoria non deve valere l'indipendenza.

$$VAR[X + Y] = VAR[X] + VAR[Y]$$

$$E[X + Y] = E[X] + E[Y]$$

Formule alternative per il calcolo del valore atteso

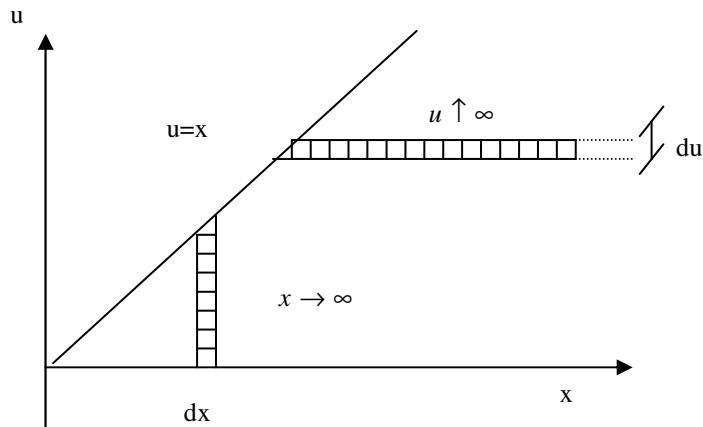
Per il calcolo del valore atteso di una variabile aleatoria non negativa, vale la seguente formula alternativa:

$$E[X] \triangleq \int_{x=0}^{\infty} [1 - F_X(x)] dx \quad \text{utilizzando l'affidabilità } R(t)$$

PROVA:

$$\begin{aligned} \int_{x=0}^{\infty} xf_X(x) dx &= \int_{x=0}^{\infty} \int_{u=0}^x du f_X(x) dx = \\ &= \int_{u=0}^{\infty} \int_{x=u}^{\infty} f_X(x) dx du = \int_{u=0}^{\infty} [1 - F_X(u)] du \end{aligned}$$

integrale della curva di sopravvivenza ritorna il valore atteso!



Per il calcolo del valore atteso della N, vale la seguente formula alternativa:

$$E[N] \triangleq \sum_{n=0}^{\infty} n P(N=n) = \sum_{n \geq 1} P(N \geq n)$$

Infatti:

$$\begin{aligned} \sum_{n=0}^{\infty} n P_N(n) &= \\ P_N(1) + P_N(2) + P_N(3) + \dots &= \sum_{n \geq 1} P(N \geq n) \end{aligned}$$

Calcolo del valore atteso della legge esponenziale

Si possono seguire due metodi alternativi:

$$A) \quad E[X] \hat{=} \int_{t=0}^{\infty} t \lambda e^{-\lambda t} dt$$

ovvero risolvendo per parti

$$\frac{1}{\lambda} \int t \lambda e^{-\lambda t} d(\lambda t) = \frac{1}{\lambda} \int x e^{-x} dx \quad (\text{avendo posto: } x \hat{=} \lambda t)$$

Poiché:

$$d(xe^{-x}) = xd(e^{-x}) + e^{-x}dx = -xe^{-x}dx + e^{-x}dx$$

$$\int d(xe^{-x}) = - \int xe^{-x}dx + \int e^{-x}dx$$

$$\Rightarrow \int xe^{-x}dx = - \int d(xe^{-x}) + \int e^{-x}dx =$$

$$[-xe^{-x}]_0^\infty + [-e^{-x}]_0^\infty = 0 + 1$$

si ottiene:

$$\frac{1}{\lambda} \int_0^\infty xe^{-x}dx = \frac{1}{\lambda} = E[X]$$

IN PARTICOLARE $\frac{1}{\lambda}$ È DETTO MTTF (MEAN TIME TO FAILURE) TEMPO MEDIO AL GUASTO

OSSERVAZIONE:

Il reciproco del valore atteso della variabile aleatoria X è quell'unico parametro, λ , che compare nella definizione della distribuzione esponenziale $F_X(x) = 1 - \exp\{-\lambda x\}$ e nell'espressione della densità $f_X(x) = \lambda \exp\{-\lambda x\}$.

Calcolo del momento del secondo ordine e della varianza

È stato appena dimostrato che la media di una variabile aleatoria distribuita con legge esponenziale di parametro λ è pari a $1/\lambda$. Dunque, è bene puntualizzare che se X rappresenta il tempo di vita di un componente non soggetto ad usura, allora il tempo medio al guasto è:

$$E[X] = \int_0^\infty R(t)dt = \frac{1}{\lambda}, \quad \text{con } R(t) = \exp\{-\lambda t\}$$

e, nella notazione anglosassone dell'affidabilità, è noto con l'acronimo **MTTF** (mean time to failure).

Dalla definizione:

$$E[X^2] = \int_0^\infty t^2 f_X(t)dt$$

ricordando che:

$$f_X(t)dt = dR(t) = R'(t)dt$$

si ottiene:

$$\begin{aligned} E[X^2] &= - \int_0^\infty t^2 R'(t)dt \\ &= -t^2 R(t) \Big|_0^\infty + \int_0^\infty 2t^{2-1} R(t)dt = \int_0^\infty 2t R(t)dt \end{aligned}$$

e di nuovo, con $R(t) = \exp\{-\lambda t\}$, risulta: $E[X^2] = \frac{2}{\lambda^2}$.

Per calcolare la varianza (del tempo di vita) si può usare la formula seguente, che sarà dimostrata più avanti:

$$\begin{aligned} VAR[X] &= E[X^2] - E[X]^2 && \text{(Utile in futuro)} \quad \text{(varianza è momento del secondo ordine meno momento del primo ordine al quadrato)} \\ &= \int_0^\infty 2t R(t)dt - \left(\int_0^\infty R(t)dt \right)^2 \end{aligned}$$

e per la legge esponenziale risulta: $VAR[X] = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$

La varianza dell'esponenziale è molto grande

EXTRA - MEDIA E VARIANZA DI WEIBULL

ALPHA E BETA SONO DETTI RISPETTIVAMENTE
PARAMETRI DI FORMA E DI SCALA

Proposizione :

Sia X una var. al. di Weibull con parametri α e β ($\beta \triangleq \frac{1}{\Gamma(1/\alpha)}$) secondo notazione di Excel.

Allora risultano:

$$\text{valore atteso} \triangleq \mu = \bar{\alpha}^{1/\beta} \Gamma(1+1/\beta)$$

$$\text{varianza} \triangleq \sigma^2 = \bar{\alpha}^{-2/\beta} \Gamma(1+2/\beta) - \mu^2$$

$$(\text{ricordare: } \sigma^2 = E[X^2] - E[X]^2)$$

Prova : Partiamo dalla espressione della densità di Weibull usata da EXCEL

$$f_X(x) \triangleq \alpha \cdot \beta \cdot x^{\beta-1} \cdot \exp(-\alpha x^\beta)$$

Calcoliamo il valore atteso:

$$E[X] = \int_{x=0}^{\infty} x \cdot \alpha \beta x^{\beta-1} \cdot e^{-\alpha x^\beta} \cdot dx$$

$$= \int_{x=0}^{\infty} \alpha \beta x^{\beta} \cdot e^{-\alpha x^\beta} \cdot dx$$

$$\text{Poniamo } z = \alpha x^\beta$$

$$\text{quindi} \Rightarrow x = (z/\alpha)^{1/\beta} \quad dx = \frac{1}{\alpha \beta} \left(\frac{z}{\alpha}\right)^{\frac{1}{\beta}-1} dz$$

$$\begin{aligned}
 \text{Da qui } \Rightarrow E[X] &= \int_{z=0}^{\infty} \alpha \beta \frac{z}{\alpha} \cdot e^{-z} \cdot \frac{1}{\alpha \beta} \cdot \left(\frac{z}{\alpha}\right)^{\frac{1}{\beta}-1} \cdot dz \\
 &= \int_0^{\infty} \left(\frac{z}{\alpha}\right)^{\frac{1}{\beta}} \cdot e^{-z} \cdot dz \\
 &= \alpha^{-\frac{1}{\beta}} \int_{z=0}^{\infty} z^{\frac{1}{\beta}} e^{-z} \cdot dz
 \end{aligned}$$

Ebbene, l'integrale indefinito che vedete è la funzione "gamma" di Euler (precisamente: il II integrale di Euler) che si riporta sui libri con la seguente notazione:

$$M(\alpha) \stackrel{\Delta}{=} \int_{z=0}^{\infty} z^{\alpha-1} \cdot e^{-z} \cdot dz \quad \text{con } \alpha > 0$$

Il calcolo di $M(\alpha)$ per specifici valori di α è detto "quozio di Euler" ☺

Nel nostro caso, ci limitiamo ad osservare quanto segue:

$$\begin{aligned}
 \int_{z=0}^{\infty} z^{\frac{1}{\beta}} \cdot e^{-z} \cdot dz &\equiv M\left(1 + \frac{1}{\beta}\right) \Rightarrow \text{conclusione:} \\
 &\Rightarrow E[X] = \alpha^{-\frac{1}{\beta}} \cdot M\left(1 + \frac{1}{\beta}\right)
 \end{aligned}$$

Statistiche dell'Ordinamento

Date X_1, X_2, \dots, X_n variabili aleatorie indipendenti ed identicamente distribuite, sono utili

nell'analisi probabilistica le variabili aleatorie Y_1, \dots, Y_n definite alla seguente maniera:

Presi n studenti che fanno uno scritto che partono allo stesso momento $Y_1 \hat{=} \min\{X_1, X_2, \dots, X_n\}$ Y_1 è la statistica del primo ordine e corrisponde al primo studente che consegna

e così via $Y_2 \hat{=} \min[\{X_1, X_2, \dots, X_n\} - \{Y_1\}] \dots Y_{n-1} \hat{=} \min[\{X_1, X_2, \dots, X_n\} - \{Y_1, Y_2, \dots, Y_{n-2}\}]$

Dunque Y_n corrisponde $Y_n = \max\{X_1, X_2, \dots, X_n\}$ sarebbero durate che partono all'ultimo che consegna tutte allo stesso momento perciò

Y_i sono funzioni di X_i note come **statistiche dell'ordinamento**. Infatti, detta Y_k la statistica di ordine k di variabili aleatorie

($k=1, \dots, n$), è immediato riconoscere che, se X_i rappresenta il tempo al guasto del componente i ($i=1, \dots, n$), allora la statistica di ordine $k = n-m+1$ rappresenta il tempo al guasto di un sistema "*m out of n*". $\min/\max\{X_1, X_2, \dots, X_n\}$ è una funzione di variabili aleatorie come si vedrà

In linea di principio, per calcolare l'affidabilità del sistema occorrerebbe determinare la funzione di distribuzione della statistica corrispondente, dato che:

Y_i può essere anche visto come

l' i -esima richiesta ad un server

$$R_{m|n}(t) = 1 - F_{Y_{n-m+1}}(t)$$

In pratica, note le funzioni di affidabilità $R_1(t), \dots, R_n(t)$ di tutti i componenti, è possibile ricavare direttamente l'affidabilità del sistema con le prove di Bernoulli.

Il sistema "*m out of n*" risulterà funzionante all'istante t se risulterà funzionante *almeno uno* dei campioni di *almeno m* componenti che è possibile estrarre dalla popolazione di numerosità pari a n . Dunque, riconoscendo che: in pratica invece di calcolare la singola affidabilità la separo nelle singolo affidabilità dei componenti

- $\prod_{i=1}^{n-j} [1 - R_i(t)] \prod_{i=1}^j R_i(t)$ è la probabilità che risultino funzionanti all'istante t
In particolare probabilità congiunta
un campione di j componenti indipendenti ; tra $n-j$ non funzionano e j che funzionano
- $\binom{n}{j}$ è il numero di campioni di numerosità j che devono essere considerati ;
- j deve variare da m a n , in modo da considerare i campioni di tutte le dimensioni utili ;

si ottiene. Il caso più semplice dove i componenti sono identici:

$$R_{m/n}(t) = \sum_{j=m}^n \binom{n}{j} [1 - R(t)]^{n-j} R(t)^j$$

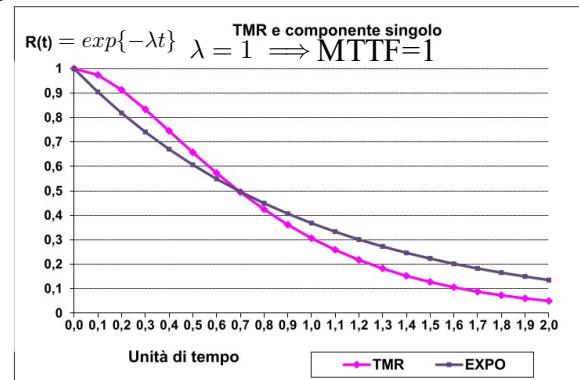
Casi particolari

- Il TMR con componenti identici ($R_1(t) = \dots = R_n(t) \hat{=} R(t)$). "perfetto"

Assumendo che il comparatore abbia probabilità di guasto trascurabile :

$$R_{TMR}(t) = R_{2/3}(t) = \binom{3}{2} [1 - R(t)][R(t)]^2 + \binom{3}{3} [R(t)]^3 = 3[R(t)]^2 + 2[R(t)]^3$$

La probabilità in blu è una congiunta ed è la probabilità che il sistema funziona a con 2 su 3 mentre la verde è che il modello funzioni 3 su 3. Il grafico mostra che la TMR interseca l'esponenziale



Affidabilità sistema parallelo= "1-out-of n":

$$R_{1/n}(t) = \sum_{j=1}^n \binom{n}{j} [1 - R(t)]^{n-j} R(t)^j = 1 - F_{Y_n}(t) \quad \text{Con } Y_n = \max\{X_1, X_2, \dots, X_n\} \quad \text{procedendo con la dimostrazione si ha che}$$

$$R_{1/n}(t) = \sum_{j=1}^n \binom{n}{j} [1 - R(t)]^{n-j} R(t)^j = \sum_{j=0}^n \binom{n}{j} [1 - R(t)]^{n-j} R(t)^j - \binom{n}{0} [1 - R(t)]^n = [1 - R(t) + R(t)]^n - 1 \cdot [1 - R(t)]^n$$

$$= 1 - [1 - R(t)]^n \quad \text{esplicitando le mortalità: } [1 - R(t)]^n \implies 1 - \Pr\{X_1 \leq t \cap \dots \cap X_n \leq t\} = 1 - \prod_{i=1}^n F_{X_i}(t) \quad \text{Eguagliando i due risultati arancioni } F_{Y_n}(t) = \prod_{i=1}^n F_{X_i}(t) \quad \text{perciò la funzione tempo al guasto del ultimo è dato dal prodotto delle distribuzioni.}$$

Affidabilità sistema seriale= "n-out-of n":

$$R_{n/n}(t) = \Pr\{X_1 > t, \dots, X_n > t\} = \prod_{i=1}^n R(t) = \binom{n}{n} [1 - R(t)]^{n-n} R(t)^n = 1 - F_{Y_1}(t) \quad \text{con } Y_1 = \min\{X_1, X_2, \dots, X_n\}$$

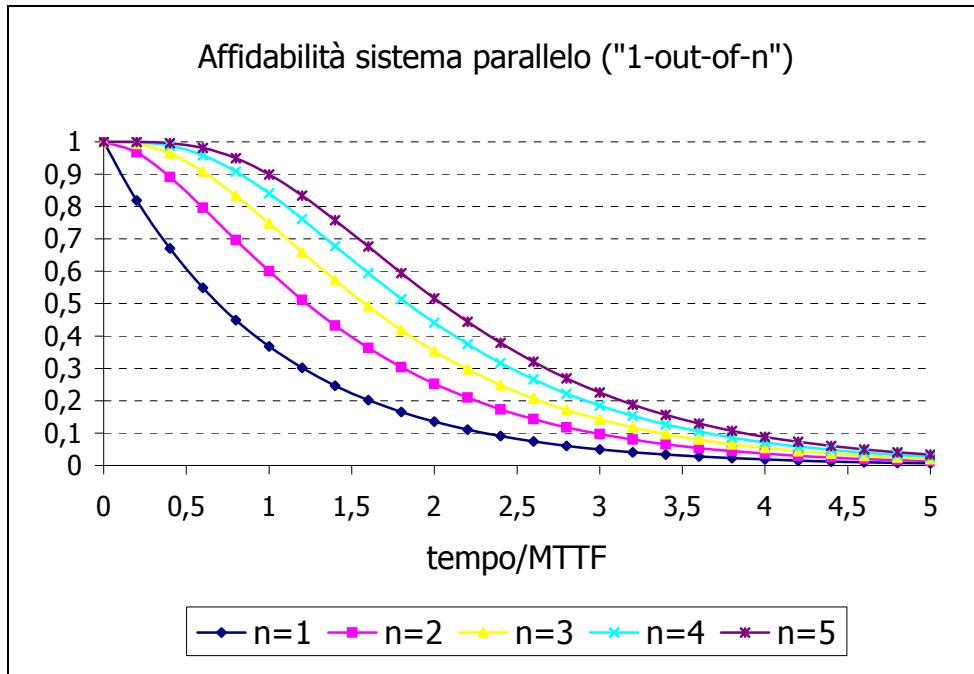
Eguagliando i risultati in fucsia, ed esplicitando R(t) si ottiene che: $\prod_{i=1}^n [1 - F_{X_i}(t)] = 1 - F_{Y_1}(t)$ ovvero che la produttoria fra le sopravvivenze è pari alla sopravvivenza.

Come corollario al calcolo dell'affidabilità del sistema seriale, si può ricavare un risultato utile in generale:
una funzione di distribuzione definita come min di altre n esponenziali indipendenti è anch'essa esponenziale, con parametro pari alla somma degli n parametri. Infatti, se:

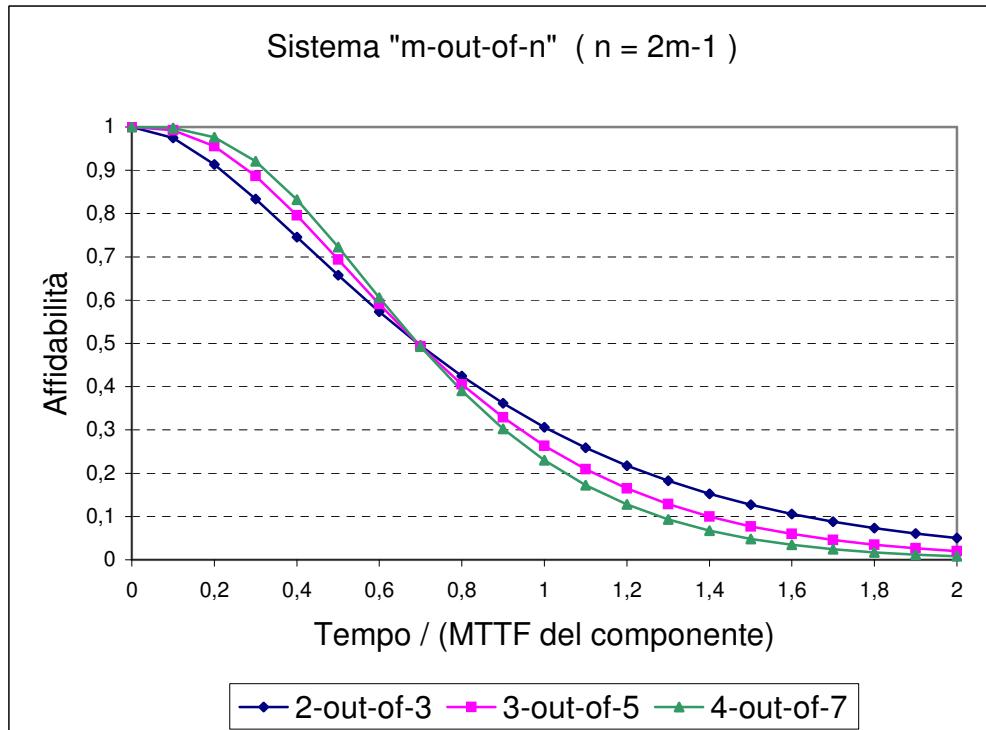
$$F_{X_i}(t) = 1 - e^{-\lambda_i t} \implies F_{Y_1}(t) = 1 - \prod_{i=1}^n e^{-\lambda_i t} = 1 - \exp[-\sum_{i=1}^n \lambda_i]t$$

Il min rappresenta nell'affidabilità il tempo minimo per un guasto. Nel caso di componenti in serie il min rappresenta il tempo al guasto del sistema!

Grafici di affidabilità sistemi "m-out-of-n"



dopo 4 componenti in parallelo l'affidabilità che si guadagna non è poi così tanta

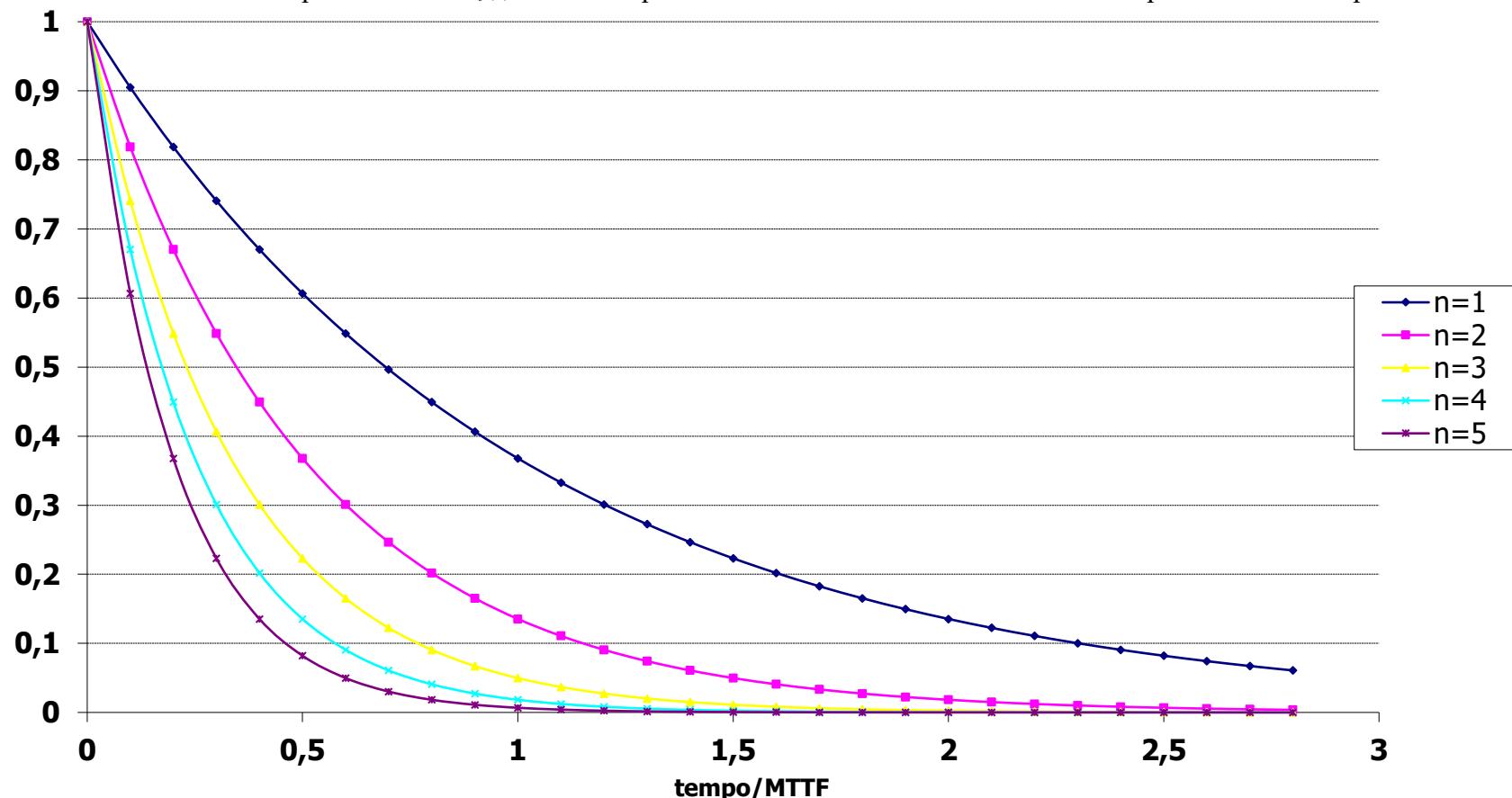


Il 2 su 3 a lungo andare è quello che ha affidabilità più alta anche se all'inizio è quello che perde affidabilità più velocemente.

AFFIDABILITA' DEL SISTEMA SERIALE A COMPONENTI IDENTICI

Affidabilità sistema seriale ("n-out-of-n")

In questo caso $F_{X_i}(t) = 1 - e^{-\lambda t}$ quindi mostra come il modello seriale con l'esponenziale si va a perdere subito



Esercizio

Sistema “m-out-of-n”

- L'elaborazione di alcuni dati acquisiti da un aeromobile durante il volo viene fatta da quattro processori indipendenti che lavorano in parallelo.
- Ad ognuno viene dato lo stesso input e si suppone che i risultati siano corretti se almeno tre dei quattro processori restituiscono lo stesso risultato.
- Il comparatore dei risultati può essere considerato "perfetto" e la legge di guasto, identica per i quattro componenti, è quella esponenziale, con un MTTF pari a 10.000 ore:

Esercizio – Quesiti (1_di_2)

1. Calcolare la probabilità che nessuno dei quattro componenti si guasti prima del 5% del proprio MTTF.
Quindi che supera il 5%
2. E la probabilità che il sistema superi le *500 ore di vita?*
(affidabilità)
3. Qual è, invece, la probabilità che *dopo le 500 ore* di vita risultino funzionanti *non più di due* dei quattro componenti?
Quindi al massimo due processori qualunque 0 e 1 sono compresi!
4. E il tasso di guasto del sistema alle *500 ore di vita?*
5. Ricalcolare il tasso di guasto del sistema supponendo, però, che alle 500 ore il sistema si sia appena ridotto a 3 componenti, per effetto di uno ed un solo guasto.

Esercizio

Soluzione del quesito 1 (1_di_2)

Sia X_i il tempo al guasto del componente i-esimo ($i=1,2,3,4$)

Sappiamo che:

$$\lambda_i = \frac{1}{10000} = 10^{-4} \text{ ore}^{-1} \quad \forall i = 1, \dots, 4$$

ossia, $\lambda_i = \lambda$ ($i=1, \dots, 4$).

Pertanto le funzioni di distribuzione saranno:

$$F_{X_i}(t) = 1 - e^{-\lambda t}$$

Esercizio

Soluzione del quesito 1 (2_di_2)

Definita la v.a.

$$X = \min\{X_1, X_2, X_3, X_4\}$$

Possiamo calcolare la probabilità che nessuno dei quattro componenti si guasti prima del 5% del proprio MTTF come:

$$P(X > t) = 1 - F_X(t) = \exp\left[-t \sum_{i=1}^4 \lambda_i\right]$$

Sostituendo i valori noti e ricordando che $\lambda_i = \lambda$ ($i=1, \dots, 4$) si ottiene:

$$P(X > 500) = e^{-4 \cdot 10^{-4} \cdot 500} = 0.819$$

Esercizio

Soluzione del quesito 2

L'affidabilità per $t=500$ ore di un singolo componente è:

$$R(t = 500) = P(X > 500) = e^{-10^{-4} \cdot 500} = 0.951$$

Allora la probabilità che il sistema risulti funzionante alle 500 ore di vita è:

$$\begin{aligned} R_{3/4}(t = 500) &= \sum_{j=3}^4 \binom{4}{j} \left[e^{-10^{-4} \cdot 500} \right]^j \left[1 - e^{-10^{-4} \cdot 500} \right]^{4-j} \\ &= 4 \cdot [R(t = 500)]^3 - 3 \cdot [R(t = 500)]^4 \\ &= 0.986 \end{aligned}$$

Esercizio

Soluzione del quesito 3

L'affidabilità per $t=500$ ore di un singolo componente è:

$$R(t = 500) = P(X > 500) = e^{-10^{-4} \cdot 500} = 0.951$$

Dunque, la probabilità che a 500 ore (da inizio missione) risultino funzionanti **non più** di due dei quattro componenti è la seguente:

$$\begin{aligned} R_{\leq 2/4}(t = 500) &= \sum_{j=0}^2 \binom{2}{j} \cdot \left[e^{-10^{-4} \cdot 500} \right]^j \cdot \left[1 - e^{-10^{-4} \cdot 500} \right]^{2-j} \\ &= 1 \cdot 1 \cdot \left[1 - e^{-10^{-4} \cdot 500} \right]^2 + 2 \cdot \left[e^{-10^{-4} \cdot 500} \right] \cdot \left[1 - e^{-10^{-4} \cdot 500} \right] + \\ &\quad + 1 \cdot \left[e^{-10^{-4} \cdot 500} \right] \cdot 1 = 0.013 \end{aligned}$$

in particolare è pari al complementare del quesito 3 perché $1 - 0.986 = 0.013$. Solo che in questo esercizio vale perché "per caso" il sistema funziona se 3 su 4 funzionano

Esercizio Soluzione del quesito 4

Il tasso di guasto del sistema $h(t)$ alle prime 500 ore di vita si calcola da:

$$h(t) = \frac{f_X(t)}{1 - F_X(t)}$$

ricorda le definizioni di $h(t)$
e di $F_X(t)$

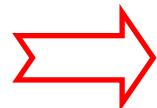
$F_X(t)$ nel caso specifico risulta:

$$F_X(t) = 1 - 4 \exp[-3\lambda t] + 3 \exp[-4\lambda t]$$

Da qui, $f_X(t)$ si ottiene derivando:

$$f_X(t) = 12 \lambda \exp[-3\lambda t] - 12 \lambda \exp[-4\lambda t]$$

Con $h(t)$ sarebbe da fare uno studio di funzione
per poi ottenere



$$h(t=500) = 0,000051$$

Esercizio

Soluzione del quesito 5

Le 500 ore sarebbero l'età e ricordando che la vita residua non dipende dall'età, quindi il rischio di guasto rimane uguale a prima

Nel caso in cui, a $t=500$ ore, si sia appena verificato il primo guasto, il sistema è diventato la serie di tre componenti. Cambia quindi la $F_X(t)$:

$$F_X(t) = 1 - \exp[-3\lambda t]$$

Il tasso di guasto, $h(t)$, si determina ricordando la proprietà di assenza di memoria della distribuzione esponenziale:

Ricordando che il rischio di guasto di UN componente è λ
Allora il rischio del sistema sarà la somma dei 3

$$h(t) = 3\lambda$$

Devi capire bene lambda

Non c'è bisogno della seguente:

$$h(t) = \frac{f_X(t)}{1 - F_X(t)}$$

Esercizio – Quesiti (2_di_2)

- 6) Effettuare lo studio di funzione su $h(t)$ al quesito 4
- 7) Calcolare un limite inferiore all'affidabilità del sistema all'istante $t = 500h$ vedi limiti al inizio
- 8) Definire la var. al. "tempo (residuo) al guasto del sistema" una volta che, per effetto del 1° guasto accaduto a $t=500h$, è stato deciso di riconfigurare il sistema come un "2-out-of-4". che accade?

Soluzione: lasciata allo studente ☺

Esercizio 1

(Problema e quesiti)

Il tasso di guasto dell'esponenziale è costante

Il “*rischio di guasto*” di un componente **sta crescendo** linearmente col tempo d'uso, secondo un coefficiente pari a $0,2 \text{ (u.t.}^{-2}\text{)}$.

da notare elevato a -2 e non -1 il tasso di guasto

1. Proporre un modello per la valutazione della affidabilità del componente. ovviamente esponenziale non va bene perché non è costante il rischio di guasto
2. Calcolare la probabilità che il componente funzioni ancora dopo 2 unità di tempo. (Affidabilità)
3. Calcolare la probabilità di un guasto nell'istante di tempo 0,25 e in un intorno $dt=0,01$ ovvero $[0,25; 0,26]$.

Esercizio 1

(Soluzione dei quesiti 1 e 2)

1. Si adotta un modello di guasto lineare

$$h(t) = 0,2 \cdot t$$

2. L'affidabilità si valuta nel modo seguente:

Applico la relazione tra tasso di guasto ed affidabilità

$$\begin{aligned} R(u) \Big|_{u=2} &= P(X \geq 2) = \exp \left[- \int_{t=0}^2 h(t) dt \right] = \\ &= \exp \left[- \int_{t=0}^2 0.2t dt \right] = \exp(-0.1 \cdot u^2) \Big|_{u=2} = 0.6703 \end{aligned}$$

Esercizio 1

(Soluzione del quesito 3)

Nella probabilità di guasto non si sa se arrivo a t. Mentre per il tasso di guasto è garantito che si arriva a t

3. La probabilità di un guasto in (t) si calcola come:

F è differenziabile per ipotesi

Per la probabilità uso $f(t)dt$
per il tasso uso $h(t)dt$

$$P(X = t) = dF(t) = f(t)dt \quad \text{con } dt = 0 \quad \rightarrow \quad P(X = t) = 0$$

La probabilità di un guasto in $[t; t+dt]$ si calcola come:

$$P(t \leq X \leq t + dt) = f(t) \cdot dt \quad \text{ricordando la relazione tra } f(t) \text{ e } R(t)$$

dove

$$f(t) = -R'(t) = h(t) \cdot R(t) = 0.2t \cdot e^{-0.1t^2}$$

R(t) è la condizionante
h(t)dt è la condizionata

$$\text{con } dt = 0.01 \Rightarrow$$

$f(t)dt = h(t)dt$ se $R(t)=1$ ovvero è certa

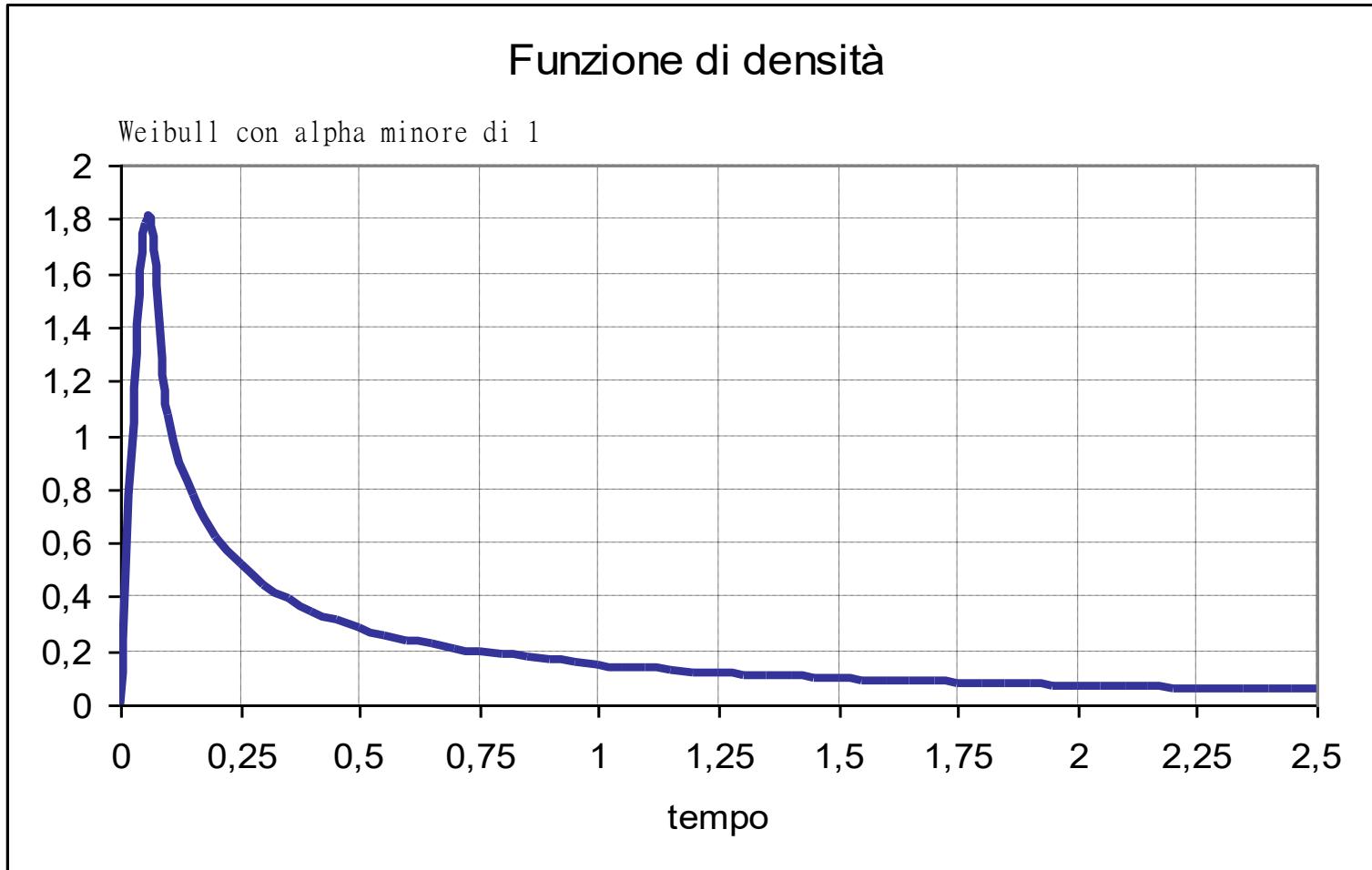
$$f(t)dt = f(0.25) \cdot 0.01 = 0.000497$$

Quantità
trascurabile

DIPENDE
DAL CONTESTO

Esercizio 1

(Considerazione sul quesito 3...)



Esercizio 2

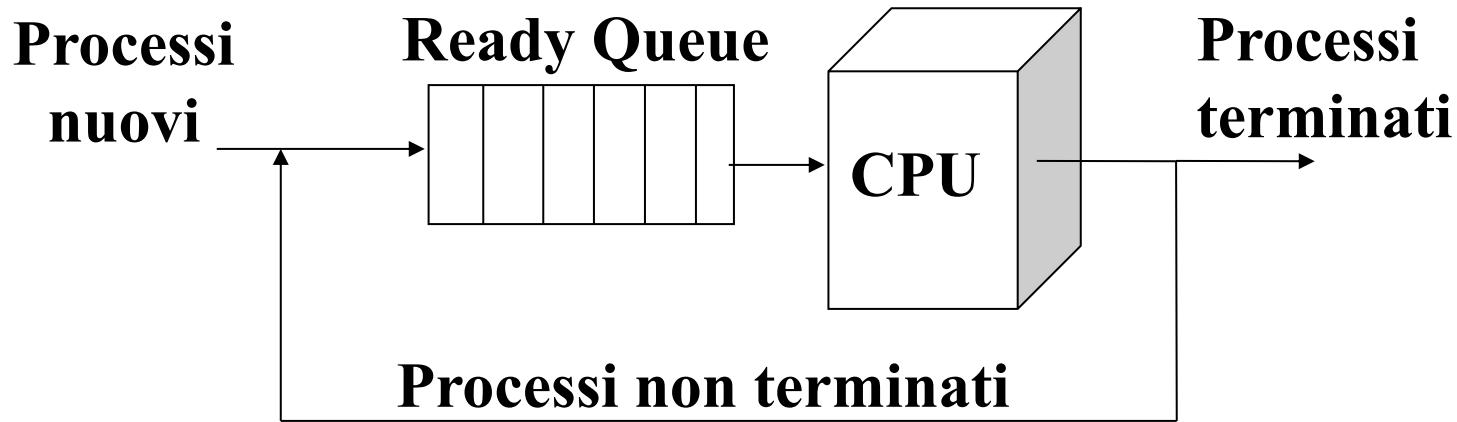
(Descrizione del problema)

Prova a riformularlo in un altro contesto

- Un sistemista deve stabilire la lunghezza del “*time slice*” per l’allocazione della CPU ai singoli processi in attesa nella “ready queue” di un sistema multiprogrammato.
- Al termine del *time slice* corrente il processo che non abbia ancora terminato l’esecuzione viene ricollocato nella “ready queue”, all’ultimo posto.
- Considerato che con un *monitor hardware* è stato determinato un tempo medio di esecuzione di 0,2sec per processo, il sistemista pensa di fissare a 0,02sec la durata del *time slice*.

Esercizio 2

(Descrizione del problema)



- Considerazioni?
- Ulteriori quesiti?

Esercizio 2 (Quesiti)

1. Proporre e giustificare un modello per il tempo di esecuzione di un processo. non si può usare Weibull perché si ha il tempo medio di esecuzione
2. Calcolare la probabilità che un processo abbia bisogno di più di un solo *time slice* per essere completamente eseguito.
3. Calcolare la probabilità che un processo abbia bisogno di un numero di *time slice* compresi tra 1 e 3, per essere completamente eseguito.
4. Posto che un processo sia stato rimesso in coda dopo aver esaurito il suo primo *time slice*, calcolare la probabilità che venga completato con il suo secondo *time slice*.

Esercizio 2 (Soluzione del quesito 1)

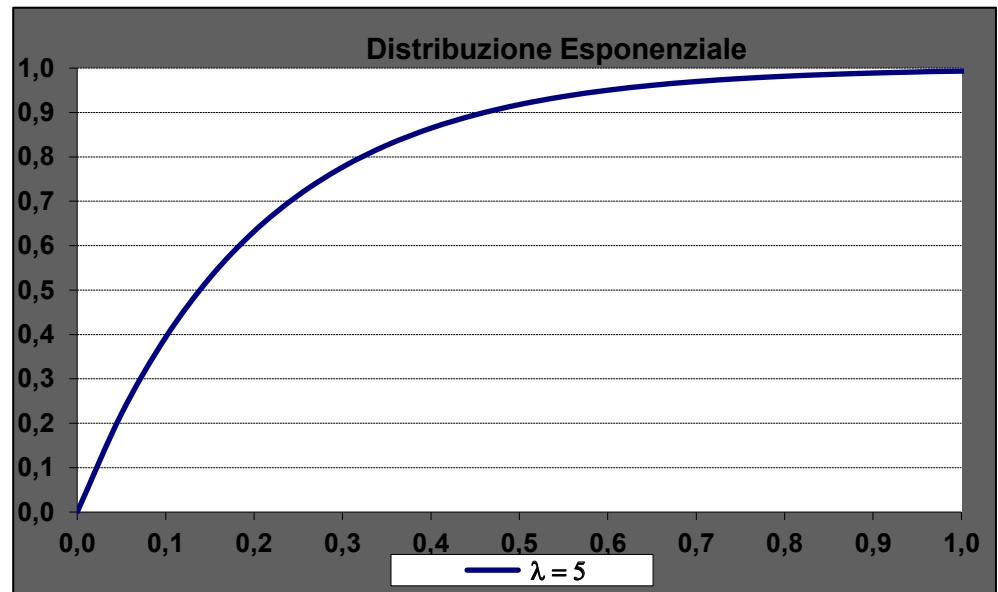
- Avendo solo la media (valore atteso) del tempo di esecuzione, il modello da adottare è quello esponenziale.

Sia X “la durata del tempo di esecuzione”

Allora: $E[X] = \frac{1}{\lambda} = \frac{2}{10} \Rightarrow \lambda = 5 \text{ sec}^{-1}$

e quindi

$$\begin{aligned} F_X(t) &= 1 - e^{-\lambda t} \\ &= 1 - e^{-5t} \end{aligned}$$



Esercizio 2 (Soluzione dei quesiti 2-4)

- La probabilità che sia sufficiente un unico *time slice* è:

$$P(X \leq 0.02) = \int_{0.0}^{0.2} f_X(u) du = F_X(0.02) = 0.095$$

- La probabilità che ne occorra più di uno solo è:

$$P(X > 0.02) = 1 - P(X \leq 0.02) = 0.905$$

- La probabilità che occorrono da 1 a 3 *time slice* è:

non uso il differenziale perché non è un tempo piccolo l'intervallo

$$P(0.02 \leq X < 0.06) = F_X(0.06) - F_X(0.02) = 0.164$$

- La probabilità che venga completato col suo secondo *time slice*, posto che non ce l'abbia fatta col suo primo.

$$P(X \leq 0.04 | X > 0.02) = P(X \leq 0.02) = 0.095$$

(Assenza di memoria)

Esercizio 2 (nuova soluzione del quesito 4)

SENZA RICORRERE ALLA PROPRIETA' DI ASSENZA DI MEMORIA

l'unica distribuzione che ha assenza di memoria è l'esponenziale

Sia: $Y \hat{=} X - \bar{x}$, dove: X è la “durata della vita del fenomeno”,
 Y la “vita residua” e x -segnato è l’età.

RISULTA:

$$P(Y \leq y \mid X \geq \bar{x}) = P(X - \bar{x} \leq y \mid X \geq \bar{x})$$

Vai a rivedere la dimostrazione alternativa

$$= \frac{P(\bar{x} \leq X \leq \bar{x} + y)}{P(X \geq \bar{x})} = \frac{F(\bar{x} + y) - F(\bar{x})}{1 - F(\bar{x})}$$

$$= \frac{F(0.02 + 0.02) - F(0.02)}{1 - F(0.02)} = 0.095$$

DOMANDA extra:

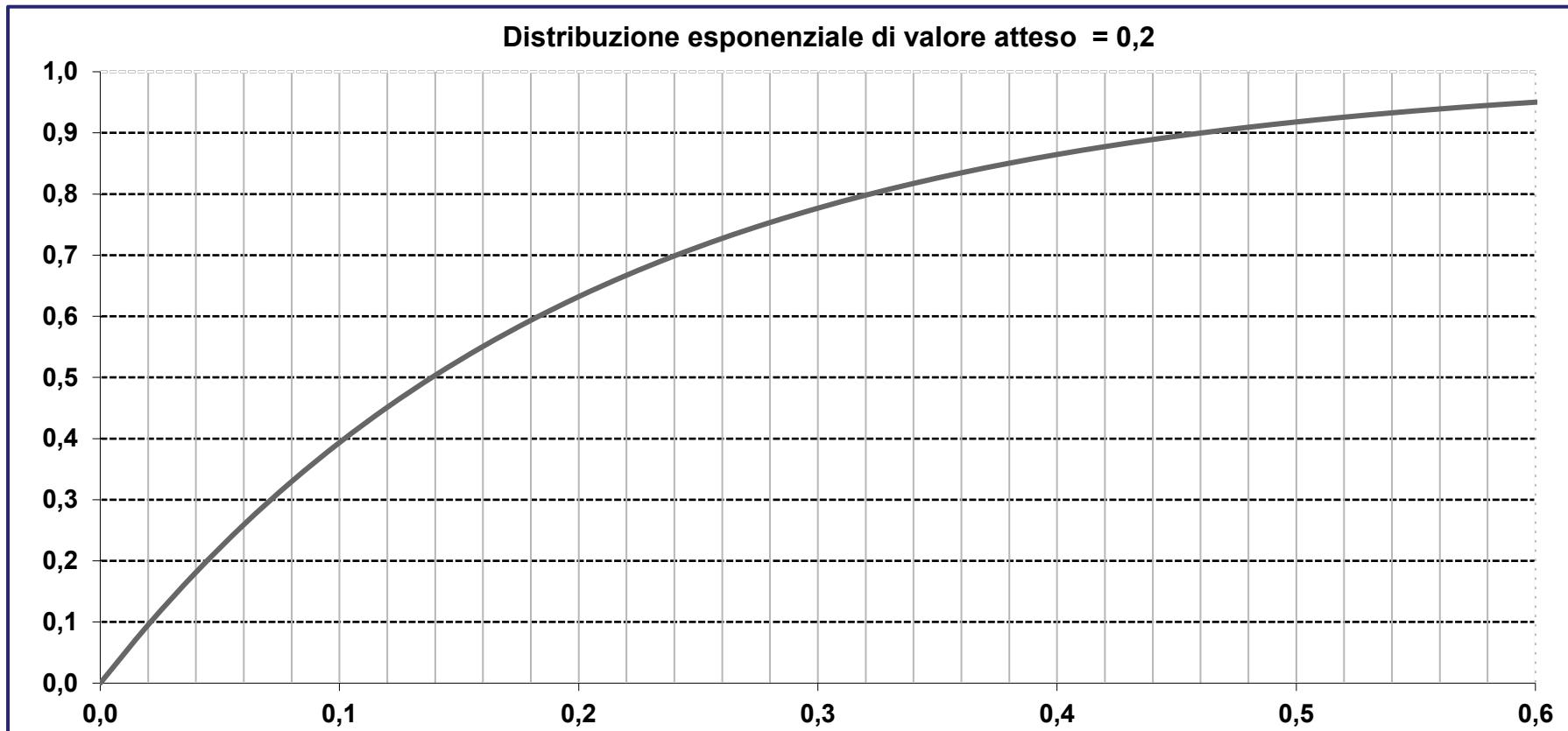
A cosa corrisponde la seguente?

$$\tilde{F}(x) \hat{=} \frac{F(x)}{F(\bar{x})}, \quad 0 \leq x \leq \bar{x}.$$

Se $F(x) = P(X \leq x)$ e $F(\bar{x}) = P(X \geq \bar{x})$
 $\tilde{F}(x) = P(X = x)?$

Oppure che la sia arrivato alla fine della durata. Cioè se pari a 1 ho terminato

Esercizio 2 (verifica calcoli)



Nota: il 39% delle realizzazioni non supera la metà del valore atteso, ma il 5% lo supera di almeno 3 volte! (si usa il complementare)

Domanda extra: tracciare il grafico troncato al valore 0,6, cioè corrispondente ad un modello che escluda durate superiori di almeno 3 volte il valore atteso.

Esercizio 3

(la vista degradante con la vecchiaia)

vedi file per soluzione

Un professore vuole acquistare un paio di occhiali da vista, a causa della presbiopia che sta per arrivare a lui a causa dell'età.

Al momento il professore ci vede benissimo a leggere da vicino, diciamo al 100%.

Secondo il suo oculista, “la capacità visiva” (“ $C(t)$ ”) si ridurrà nel tempo in maniera aleatoria, con una velocità media costante. Precisamente, alla velocità media di 0,25 gradi all’anno per i primi 4 anni dopo l’acquisto.

E’ possibile calcolare la capacità visiva del professore dopo 2 anni, assumendo che sia pari a 10 gradi al momento dell’acquisto?

RICAP.: distribuzione della v.a. “max” (1_di_2)

- La v.a. “max” è:

$$X \doteq \max \{X_1, \dots, X_n\}$$

– Allora:

$$F_X(t) = Pr(X_1 \leq t, \dots, X_n \leq t) = \prod_{i=1}^n F_{X_i}(t)$$

– L'affidabilità aumenta assai poco con n :

$$F_X(t) = (1 - e^{-\lambda t})^n$$

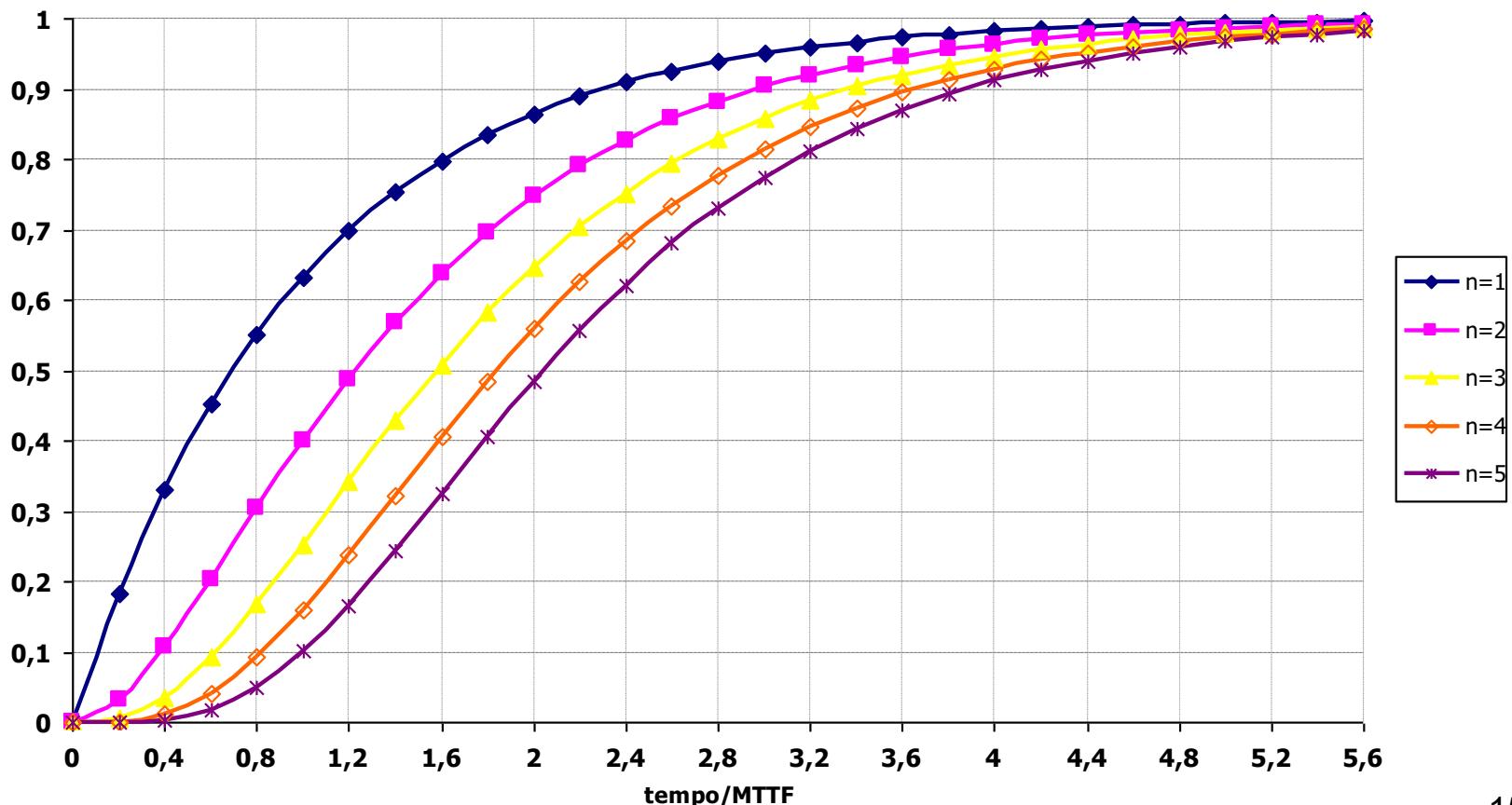
Caso
“esponenziale”

RICAP.: distribuzione della v.a. “max” (2_di_2)

Curve della $F_X(t)$ al variare di n

$$\lambda = 2$$

Funzione di Distribuzione v.a. "max"



RICAP.: distribuzione della v.a. “min” (1_di_2)

- La v.a. “min” è:

Allora:

$$X \doteq \min \{X_1, \dots, X_n\}$$

$$1 - F_X(t) = \Pr(X_1 > t, \dots, X_n > t) = \prod_{i=1}^n 1 - F_{X_i}(t)$$

L'affidabilità diminuisce assai velocemente con n :

$$F_X(t) = 1 - (e^{-\lambda t})^n$$

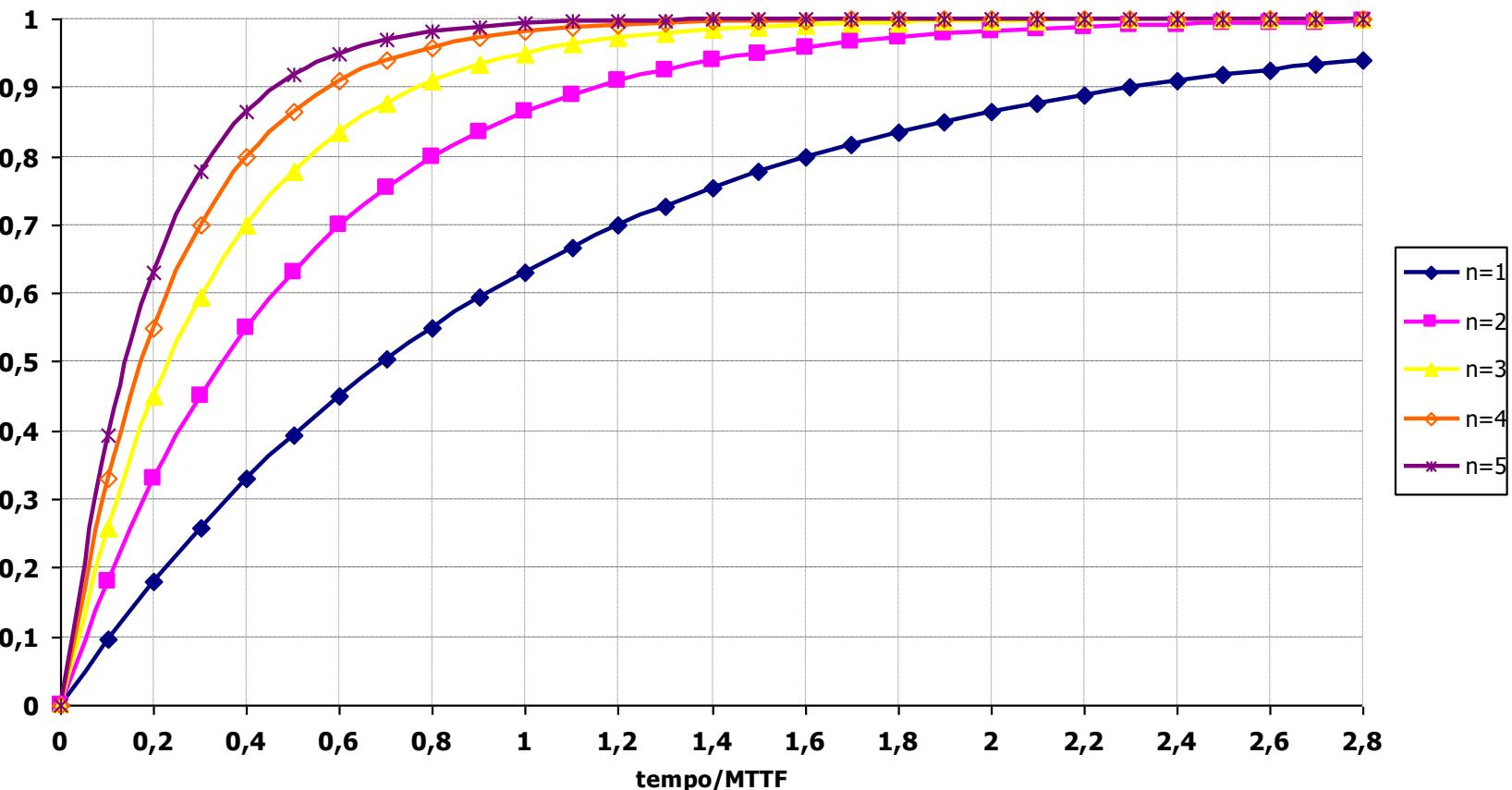
Caso
“esponenziale”

RICAP.: distribuzione della v.a. “min” (2_di_2)

Curve della $F_X(t)$ al variare di n

$$\lambda = 1$$

Funzione di Distribuzione v.a. "min"



RICAP.: generalizzazione per la v.a. “min di exp”

- Se

$$F_{X_i}(t) \hat{=} 1 - e^{-\lambda_i t}, \quad i = 1, \dots, n$$

e indipendenti

$$X \hat{=} \min \{X_1, \dots, X_n\}$$

allora la X segue ancora una legge esponenziale:

$$F_X(t) = 1 - \prod_{i=1}^n e^{-\lambda_i t} = 1 - \exp[-t \sum_{i=1}^n \lambda_i]$$

Esercizio 4

(Descrizione del problema)

Classica situazione per evitare la consistenza

- Un certo numero di processi-lettori può accedere alle tabelle di un database.
- Due processi-scrittori possono aggiornare il DB lavorando in **parallelo** (su tabelle diverse) e in modo **indipendente**.
- L'accesso dei processi-lettori al DB è bloccato per tutta la durata dell'aggiornamento.

Esercizio 4 (Quesiti)

- Adottando due modelli esponenziali, di media pari a 4 sec e 6 sec, per le rispettive durate del lavoro dei due processi-scrittori:
 1. Ricavare la distribuzione della v.a.
 X = “tempo di blocco degli accessi in lettura”.
(quanto tempo deve aspettare un processo per leggere?) il massimo fra i due tempi di esecuzione dei processi scrittori
 2. Calcolarne il valore atteso.

Esercizio 4 (Soluzione al quesito 1)

- Risulta:

$$X \doteq \max \{X_1, X_2\}$$

- dove X_i rappresenta il contributo alla durata del blocco dovuto al processo i (con $i=1,2$).
- Allora, usando due leggi exp di parametro μ_1 e μ_2 :

Utilizzo affidabilità nel caso esponenziale (vedi corollario)

$$\begin{aligned} F_X(t) &= (1 - e^{-\mu_1 t})(1 - e^{-\mu_2 t}) \\ &= 1 - (e^{-\mu_1 t} + e^{-\mu_2 t} - e^{-(\mu_1 + \mu_2)t}) \end{aligned}$$

Esercizio 4

(Soluzione al quesito 2)

CALCOLO

- Il valore atteso della durata del blocco:

$$\begin{aligned} E[X] &= \int_{t=0}^{\infty} (1 - F_X(t)) dt \quad \text{che tra l'altro è l'MTTF. Appunto integrale della sopravvivenza} \\ &= \int_{t=0}^{\infty} \left(e^{-\mu_1 t} + e^{-\mu_2 t} - e^{-(\mu_1 t + \mu_2 t)} \right) dt \\ &= \frac{1}{\mu_1} + \frac{1}{\mu_2} - \frac{1}{\mu_1 + \mu_2} \cong 7.5 \text{ sec} \end{aligned}$$

cos'è seconde te? è il valore atteso minimo che
debbo aspettare tra i 2

Esercizio 5

(Descrizione del problema)

- Due processi indipendenti elaborano dati letti dalle tabelle di un database e dopo aggiornano le tabelle.
- Quello che completa per **primo l'elaborazione** procede con l'operazione di **aggiornamento**.
- La **durata media** delle elaborazioni effettuate dai due processi è pari a 2 sec e 4 sec, rispettivamente.

Esercizio 5

Quesiti

- Adottando due modelli esponenziali, per le durate delle elaborazioni dei due processi:
 1. Ricavare la distribuzione della v.a.
 X = “tempo al prossimo aggiornamento”.
il minimo che serve per il prossimo aggiornamento
 2. Calcolarne il valore atteso.

Esercizio 5 (Soluzione)

- Risulta:

$$X \doteq \min \{X_1, X_2\}$$

- dove X_i rappresenta la durata dell'elaborazione effettuata dal processo i (con $i=1,2$).

utilizzando la definizione di \min vista pure nella ricapitolazione

- Allora, usando due leggi \exp di parametro μ_1 e μ_2 :

$$1 - F_X(t) = e^{-\mu_1 t} \cdot e^{-\mu_2 t}$$

$$F_X(t) = 1 - e^{-(\mu_1 + \mu_2)t}; \quad E[X] = \frac{1}{\mu_1 + \mu_2} = 1.33 \text{ sec}$$

valore atteso

Richiamo formule sistemi “m-out-of-n”

- Durate indipendenti e identiche, con stessa legge esponenziale e durata media = $1/\lambda$:

$$R_{1/n}^{\text{parallelo}}(t) = \sum_{j=1}^n \binom{n}{j} \left[e^{-\lambda t} \right]^j \left[1 - e^{-\lambda t} \right]^{n-j} = 1 - \left[1 - e^{-\lambda t} \right]^n$$

$$R_{m/n}(t) = \sum_{j=m}^n \binom{n}{j} \left[e^{-\lambda t} \right]^j \left[1 - e^{-\lambda t} \right]^{n-j}, \quad 1 < m < n$$

$$R_{n/n}^{\text{seriale}}(t) = \sum_{j=n}^n \binom{n}{j} \left[e^{-\lambda t} \right]^j \left[1 - e^{-\lambda t} \right]^{n-j} = \left[e^{-\lambda t} \right]^n$$

Esercizio 6

(il problema dell'attesa in pizzeria) (1_di_2)

È martedì sera; una piccola pizzeria apre alle ore 20 e 30 e tutti gli 8 tavoli disponibili sono prenotati esattamente per quell'ora. Una coppia di amici arriva proprio alle 20 e 30, però senza prenotazione.

Gli 8 tavoli prenotati sono tutti occupati da studenti universitari, che hanno comportamenti identici dal punto di vista probabilistico. Pertanto, si può assumere che il tempo medio di occupazione del generico tavolo sia di 60 min. Identici dal punto di vista probabilistico è diverso da dire uguale

Inoltre, non ci sono due o più tavoli occupati da uno stesso gruppo. Gli studenti seduti a tavoli diversi neppure di vista si conoscono!

Vuol dire che sono identicamente distribuiti
ma indipendenti!

Ipotizzando che un modello probabilistico con assenza di memoria per la variabile aleatoria “tempo di occupazione di uno qualunque dei tavoli” sia adeguato, si consideri pure trascurabile la probabilità che due o più tavoli si liberino nello stesso istante (per puro caso).

che modello si può usare? quale no?

La porzione di testo in blu è una ipotesi semplificativa
che varrà sempre ed è molto semplificativa

Esercizio 6

(il problema dell'attesa in pizzeria) (2_di_2)

Nelle ipotesi fatte,

- 1) calcolare la probabilità che la coppia di amici possa avere un tavolo entro 20 minuti. sarebbe il min

MODIFICA del problema

La stessa pizzeria non ha un orario di chiusura prestabilito e ha accettato solo il primo turno di prenotazioni, mandando via la coppia di amici senza prenotazione e chiunque altro arrivasse dopo. In tale nuova ipotesi,

- 2) calcolare la probabilità che il gestore possa chiudere entro 120 minuti dalle 20 e 30 sarebbe il calcolo del max

Lo stesso gestore ha due camerieri in servizio, ma ha deciso che ne manderà via uno alle 22 e 30 se per quell'ora saranno ancora occupati uno o due tavoli (sugli otto occupati dalle 20 e 30).

- 3) Calcolare la probabilità che ciò accada. m out of n caso M=1 e caso M=2

Data l'ipotesi di utilizzare un modello probabilistico con assenza di memoria, risulta naturale utilizzare il modello esponenziale: $F(t) = 1 - e^{-\lambda t}$

Il numero di tavoli totali è 8 e come suggerisce il testo, il tempo medio di occupazione di un tavolo è di 60 minuti. Dunque il valore atteso è 60 min e da esso si può ricavare lambda: $E[X] = \frac{1}{\lambda} = 60 \text{ min} \implies \lambda = \frac{1}{E[X]} = \frac{1}{60 \text{ min}} = 0.017 \text{ min}^{-1} = 0.00027778 \text{ s}^{-1}$

Risoluzione quesito 1): un tavolo qualunque si deve liberare entro 20 min

Consideriamo gli 8 tavoli come un sistema seriale. L'affidabilità è dunque pari a $R_{ser}(t)$ e prevede che nessun tavolo si liberi prima di un tempo t. Con $1-R_{ser}(t)$ calcolo la probabilità che almeno uno dei tavoli si liberi. Cioè quando un tavolo smette di essere occupato il sistema non è più "affidabile", se fossero pistone di una macchina sarebbe quando uno qualunque su 8 si rompe e va sostituito no?

Usando il risultato che esprime l'affidabilità per un sistema seriale si ottiene che:

$$R_{ser}(t) = 1 - F_{Y_1}(t) = 1 - (1 - \exp[-\sum_{i=1}^n \lambda_i]t) = \exp[-n\lambda t] = e^{-8 \cdot 0.00027778 \cdot 20 \cdot 60} = 0.069$$

In particolare se X="tempo che un tavolo si liberi"

$$\Pr(X \leq 20) = 1 - R_{ser}(t) = 1 - 0.069 = 0.931$$

Risoluzione quesito 2): ovvero tutti i tavoli si devono liberare entro 120 min

Stavolta considero gli 8 tavoli come un sistema parallelo. L'affidabilità in questo caso è pari a $R_{par}(t)$ e prevede che almeno 1 dei tavoli non si liberi prima di un tempo t. Con $1-R_{par}(t)$ calcolo la probabilità che tutti i tavoli si liberino entro un tempo t.

Usando il risultato che esprime l'affidabilità per un sistema parallelo si ottiene che:

$$R_{par}(t) = 1 - [1 - R(t)]^n = 1 - (F_X(t))^n = 1(1 - e^{-\lambda t})^n = 1 - (1 - e^{-0.00027778 \cdot 120 \cdot 60})^8 = 0.68$$

In particolare se X="tempo che tutti i tavoli siano liberi"

$$\Pr(X \leq 120) = 1 - R_{par}(t) = 1 - 0.68 = 0.32$$

Risoluzione quesito 3): max 2 tavoli occupati entro 2 ore perché il servizio parte alle 20:30

Considero un sistema 3 out of 8. L'affidabilità di questo sistema prevede che almeno 3 tavoli su 8 siano occupati al tempo t. Con $1-R_{3/8}(t)$ calcolo la probabilità che MENO DI 3 tavoli siano occupati al tempo t.

La distribuzione “normale” (di Gauss).

La distribuzione normale è piuttosto alla base dell’analisi statistica ed ha, invece, un ruolo più limitato nell’analisi probabilistica, non fosse altro che per il fatto che essa, non è esprimibile in forma chiusa (non si possono fare operazioni algebriche) ma solo come funzione integrale della sua densità. Dunque non è manipolabile più di tanto all’interno di formule fondamentali come la distribuzione totale.

Nel linguaggio dell’analisi probabilistica, si può dire che la funzione densità normale è la formalizzazione di un “modo di ripartizione” delle probabilità sullo spazio continuo delle realizzazioni di una variabile aleatoria fondata su due assunzioni:

1. Esiste una “realizzazione tipica”, nello spazio delle possibili realizzazioni, ed è quella che ricorre più frequentemente, valore atteso
2. Esiste un “livello di dispersione ben preciso”, più o meno grande, che dà ragione di scostamenti più o meno grandi dalla realizzazione tipica e, comunque, simmetrici. varianza

La ratio delle precedenti assunzioni si spiega con il fatto che, storicamente, con la densità normale si voleva descrivere un modo di ripartizione degli errori in esperimenti di misura di grandezze fisiche che poggiava sull’idea che l’errore tipico doveva essere nullo; ma ripetendo le misure era inevitabile la presenza di “disturbi di entità limitata”, ma assolutamente casuali e senza un verso preferenziale.

All’ipotesi 1. si può associare, in questa sede, l’interpretazione che la realizzazione tipica sia una sorta di durata vera e propria del fenomeno d’interesse, determinata da fattori endogeni. Mentre, con l’ipotesi 2., si aggiunge il fatto che, in assenza di ulteriori informazioni, si accetta l’idea che anticipi e ritardi (praticamente finiti) sulla durata del fenomeno siano equiprobabili.

Per comodità di trattazione, le ipotesi 1. e 2., saranno per il momento formalizzate imponendo che la forma di densità cercata sia riferita ad una variabile aleatoria, Z , standard, cioè di media nulla, $E[Z] = 0$, e varianza unitaria, $VAR[Z] = 1$. Poi si estenderà la funzione trovata al caso generico e, soprattutto, più naturale in questa sede, di una X con $E[X] = \mu$ e $VAR[X] = \sigma^2$ (si pone al quadrato perché, spesso, si lavora con la radice della varianza, detta deviazione standard e quindi pari a σ).

La media è il punto in cui si centra la Gaussiana. Quindi se media 0 è centrata in ascissa pari a 0. La normale standard/Gaussiana ha deviazione standard 1. La deviazione standard equivale ad un "fattore di distorsione". Nella standard dopo 3 deviazioni, la campana si appiattisce. Quindi se ho media 0 e varianza 16, la non standard con deviazione 4 "muore" a dopo 3 deviazioni, quindi ascissa pari a 12

Proprietà Gaussiana standard

Con le ulteriori condizioni:

per z che tende a infinito
la funzione tende a 0

funzione crescente per $z < 0$ $\frac{d f_Z(z)}{dz} = \begin{cases} > 0 & z < 0 \\ < 0 & z > 0 \end{cases}$ e $f_Z(z \rightarrow \pm \infty) \rightarrow 0$

Più:

condizione per punto di massimo o minimo $\frac{d f_Z(z)}{dz} \Big|_{z=0} = 0$ e $\frac{d^2 f_Z(z)}{dz^2} \Big|_{z=0} < 0$ concava

lasciate all'interpretazione dello studente, si può riconoscere che la più semplice equazione differenziale che le rappresenta tutte è la seguente:

Eq. diff. primo ordine $\rightarrow \frac{d f_Z(z)}{dz} = -f_Z(z) z$; con $f_Z(z) > 0$ per $-\infty < z < +\infty$

Integrando la precedente si ottiene:

$$f_Z(z) = k \exp\{-z^2/2\}, \text{ con } k > 0$$

La costante k si determina imponendo che risulti pari ad uno l'integrale della f esteso a tutto lo spazio delle possibili realizzazioni della Z . Si pone pari a 1 perché f è una densità ed il suo integrale tra $-\infty$ e $+\infty$ è pari a 1

Dunque: il due è 0 +infinito è perché è simmetrica

$$1 = \int_{-\infty}^{+\infty} k \exp\{-z^2/2\} dz = k \frac{2}{\sqrt{2}} \int_{t=0}^{+\infty} t^{1/2-1} \exp\{-t\} dt \quad (t \triangleq z^2/2)$$

Preso il secondo integrale di Eulero $\int_{t=0}^{+\infty} t^{\alpha-1} \exp\{-t\} dt = \Gamma(\alpha)$
si ottiene il risultato

e poiché riconosciamo la "funzione gamma" nell'ultimo integrale indefinito:

$$\int_{t=0}^{+\infty} t^{1/2-1} \exp\{-t\} dt = \Gamma(1/2) = \sqrt{\pi}$$

concludiamo che la **densità normale standard** è:

rispetto alla non standard al denominatore
sigma non c'è, perché pari a 1 $f_Z(z) = \frac{1}{\sqrt{2\pi}} \exp\{-z^2/2\}, \quad -\infty < z < +\infty$

e da qui la **distribuzione normale standard**, in forma di funzione integrale:

$$F_Z(z) = \int_{-\infty}^{u=z} \frac{1}{\sqrt{2\pi}} \exp\{-u^2/2\} du, \quad -\infty < z < +\infty \quad (\text{rimane funzione integrale})$$

I valori della "normale standard" sono ottenuti per integrazione numerica e sono tabulati. Alcune tabelle riportano la probabilità cumulativa della normale standard, ovvero la probabilità cade tra $-\infty$ and un dato valore $u > 0$, i.e., $P(-\infty \leq z \leq u)$

Osservazione: $F_Z(-z) = 1 - F_Z(z)$ $z \rightarrow 0$ c'è simmetria rispetto all'origine

We shall now prove that the $f_X(x)$ is a probability density function by showing that if

dimostrazione alternativa
dell'integrale precedente

$$I = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(1/2)x^2} dx$$

then $I = 1$. Direct evaluation of the integral is difficult. Hence we square I and write

$$\begin{aligned} I^2 &= \int_{-\infty}^{\infty} \frac{1}{2\pi} e^{-(1/2)x^2} dx \cdot \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(1/2)y^2} dy \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-(1/2)(x^2+y^2)} dx dy \end{aligned} \quad (6.4.6)$$

Substituting $x = r \cos(\theta)$ and $y = r \sin(\theta)$ in Eq. (6.4.5) we obtain,

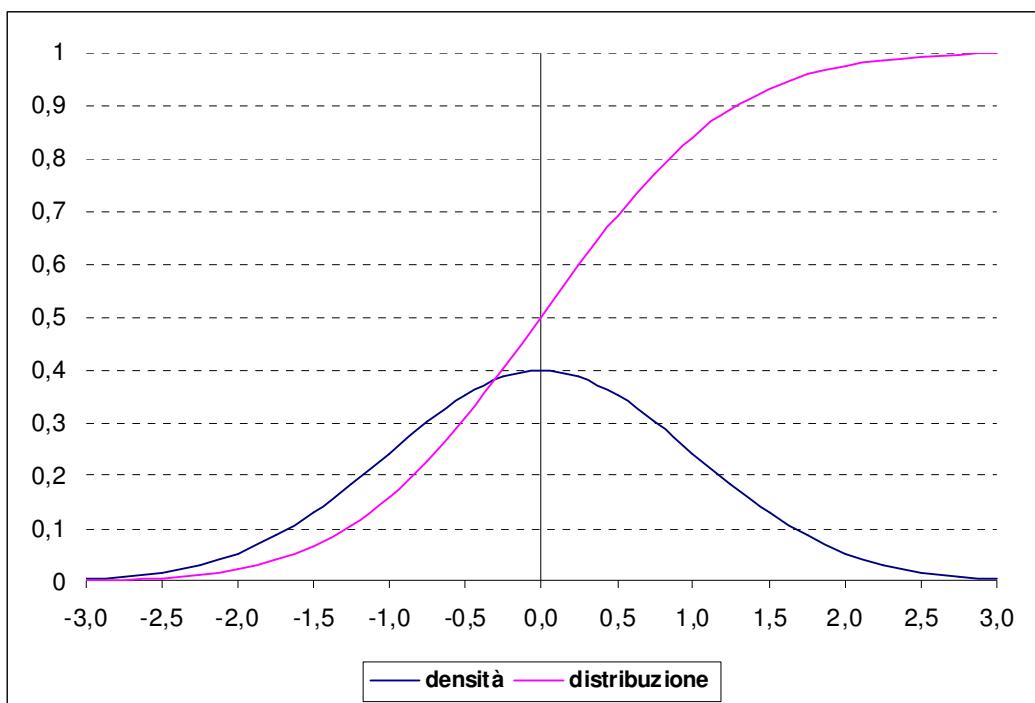
$$I^2 = \frac{1}{2\pi} \int_0^{2\pi} \int_0^{\infty} r e^{-(1/2)r^2} dr d\theta = 1 \quad (6.4.7)$$

and the result is proven.

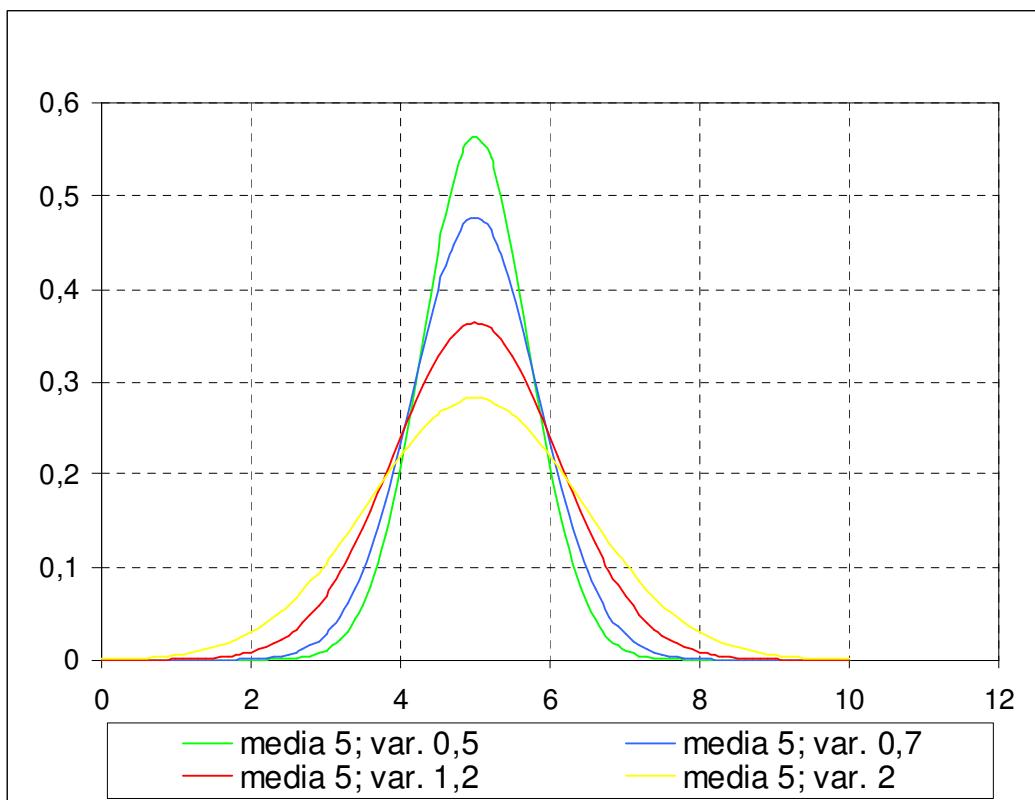
Example 6.4.1 The number of malfunctioning computers is Gaussian-distributed with $\mu = 4$ and $\sigma = 3$. We want to find the probability that the number of bad computers is

Rappresentazioni della normale

è "larga"
3 deviazioni
standard poi
tende a 0



La normale standard.



Esempi di normale NON standard.

DETALGO TECNICO_1

dimostrazione del perché esce sigma al quadrato nella varianza

PER DEFINIZIONE DI VARIANZA:

$$\text{VAR}[x] \triangleq E[(x - E[x])^2]$$

Allora:

$$\text{VAR}[\alpha x] = E[(\alpha x - E[\alpha x])^2]$$

$$= E[(\alpha x - \alpha E[x])^2]$$

$$= E[\alpha^2 \cdot (x - E[x])^2]$$

$$= \alpha^2 E[(x - E[x])^2]$$

$$= \alpha^2 \text{VAR}[x]$$

La distribuzione “normale” (9/10)

Sia $X \stackrel{\text{def}}{=} \sigma Z + \mu$, con $-\infty < \mu < +\infty$ e $\sigma > 0$
variabile aleatoria

X mantiene la forma “normale”, ma è traslata sul valore di μ
scommettiamo che

si dimostra che

$$E[X] = \sigma E[Z] + E[\mu] = \mu$$

Si dimostreranno
in seguito

La larghezza della “campana” è data dal valore di σ

$$\text{VAR}[X] = \sigma^2 \text{VAR}[Z] + \text{VAR}_{\mu} = \sigma^2$$

quelli del grafico precedente

perchè costante

I due punti di flesso corrispondono alle ascisse $\mu \pm \sigma$

creo una relazione tra la gaussiana standard ed una non standard

$$\Pr\{X \leq x\} = \Pr\left\{Z \leq \frac{x - \mu}{\sigma}\right\} \quad \rightarrow \quad z = \frac{x - \mu}{\sigma}$$

questo verrà usato per la non standard

densità non standard sarà:

$$f_X(x) = \frac{1}{\sigma \sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x - \mu}{\sigma}\right)^2\right\}, \quad -\infty < x < +\infty$$

DETTAGLIO TECNICO_2_e_3

utile per la relazione tra standard
e non standard

$$\Pr\{X \leq x\} = \Pr\left\{Z \leq \frac{x - \mu}{\sigma}\right\} \quad \rightarrow \quad z = \frac{x - \mu}{\sigma}$$

$$\begin{aligned} F_X(x) &\hat{=} \Pr\{X \leq x\} = \Pr\{\sigma Z + \mu \leq x\} = \\ &= \Pr\left\{Z \leq \frac{x - \mu}{\sigma}\right\} \hat{=} F_Z\left(\frac{x - \mu}{\sigma}\right) \end{aligned}$$

INOLTRE:

$$\begin{aligned} f_X(x) &\hat{=} \frac{d}{dx} F_X(x) = \frac{d}{dx} F_Z\left(\frac{x - \mu}{\sigma}\right) \\ &= \frac{1}{\sigma} f_Z\left(\frac{x - \mu}{\sigma}\right) \end{aligned}$$

si dimostra
nella lezione
di funzioni di
variabili aleatorie

da qui nasce il
sigma della slide precedente

Risultati fondamentali sulla normale

L'importanza della distribuzione normale è in larga parte dovuta all'esistenza dei due risultati che saranno presentati adesso. Il primo è noto come **teorema di riproducibilità** della distribuzione normale perché stabilisce che la "forma" normale si mantiene rispetto all'operazione di somma (di variabili aleatorie), mentre il secondo è noto come **teorema centrale del limite** perché stabilisce che la legge normale è la forma a cui tende la somma di variabili aleatorie di media e varianza note ed ha, per questo, un ruolo centrale nell'analisi statistica.

Teorema di riproducibilità della normale

Sia X_1, X_2, \dots, X_n una collezione di variabili aleatorie indipendenti e ciascuna distribuita secondo una legge normale di media μ_i e varianza σ_i^2 ($i = 1, \dots, n$). Allora, introdotte le costanti reali a_1, a_2, \dots, a_n , la variabile aleatoria

$$S_n \doteq \sum_{i=1}^n a_i X_i \text{ combinazione lineare}$$

è distribuita con una legge normale di media

$$= \sum_{i=1}^n a_i \mu_i \text{ e varianza } \sigma^2 = \sum_{i=1}^n a_i^2 \sigma_i^2. \text{ combinazioni lineari}$$

Teorema centrale del limite

Sia X_1, X_2, \dots, X_n una collezione di variabili aleatorie indipendenti e ciascuna distribuita secondo una stessa legge, arbitraria e con media μ_i e varianza σ_i^2 anche diverse. Allora la variabile aleatoria:

$$Z_n \text{ tende alla Gaussiana standard} \quad \tilde{Z}_n \doteq \frac{\sum_{i=1}^n X_i - \sum_{i=1}^n \mu_i}{\sqrt{\sum_{i=1}^n \sigma_i^2}} \quad \begin{array}{l} \text{Si sta standardizzando} \\ \text{ad una Gaussiana} \end{array}$$

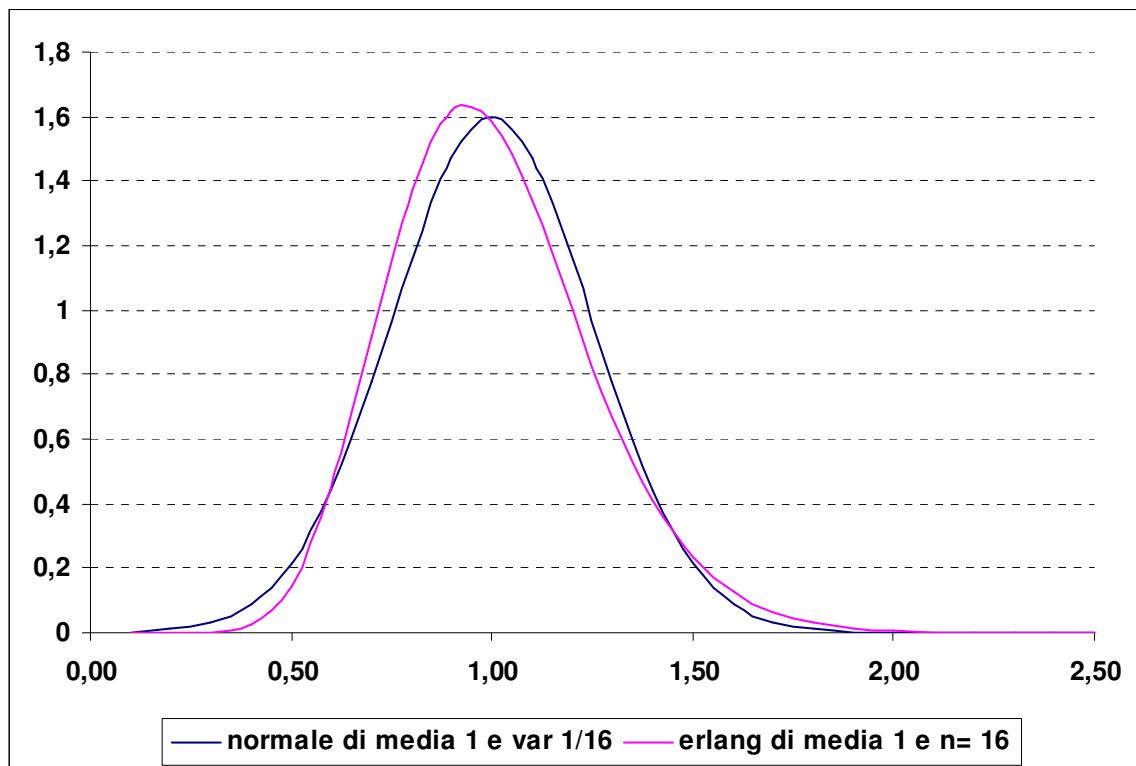
ma la sommatoria delle variabili X_i no

che ha, per costruzione, media nulla e varianza unitaria converge in forma alla normale standard al crescere di n , cioè:

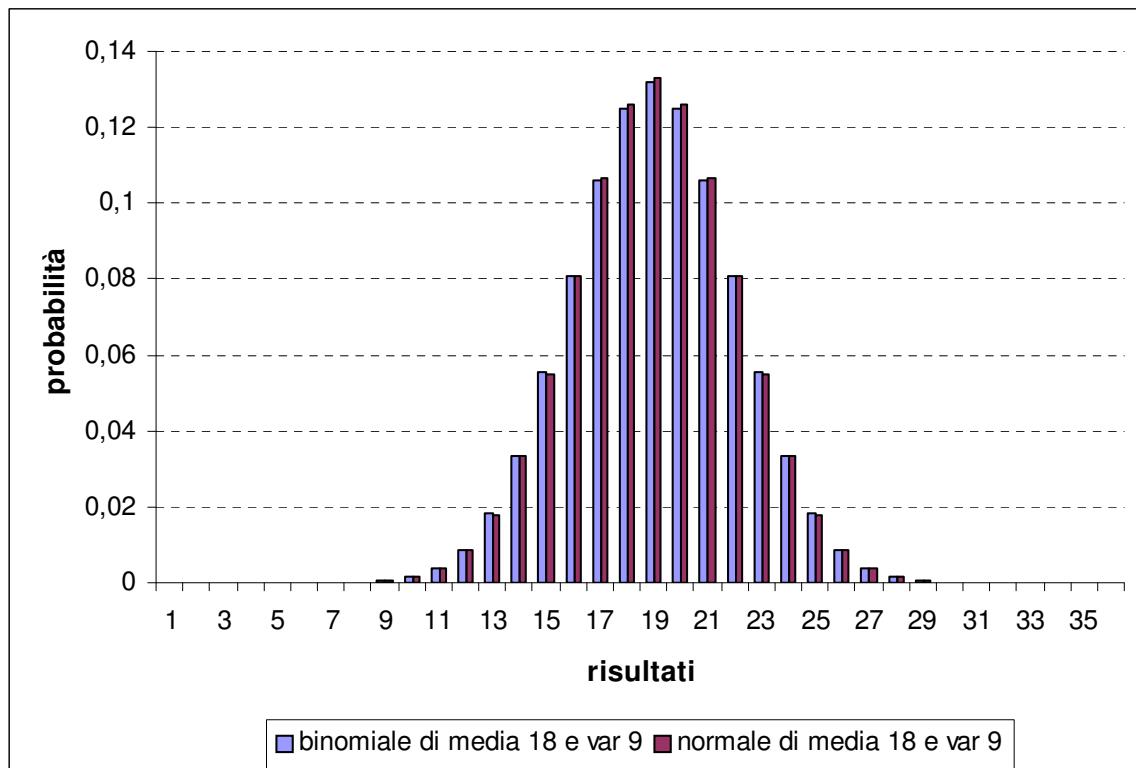
$$\begin{aligned} n \rightarrow \infty \Rightarrow F_{Z_n}(z) &\xrightarrow{\text{distr}} \int_{-\infty}^{u=z} \frac{1}{\sqrt{2\pi}} \exp\{-u^2/2\} du, \quad -\infty < z < +\infty \\ \text{già con } n=30 \\ \text{si ha una normale} \\ \text{La dimostrazione dei due teoremi è tralasciata.} \end{aligned}$$

In pratica prese cose asimmetriche dopo n passaggi ottengono una Gaussiana che è simmetrica

Illustrazione del teorema limite centrale



Erlang è somma di exp come vedremo dopo!



La binomiale è una somma di prove di Bernoulli

Distribuzione congiunta

La funzione di distribuzione congiunta, di due var. al. (continue):

Caso d'uso:
tempi al guasto di due
componenti soggetti sia a cause
individuali di guasto sia
a cause comuni di guasto

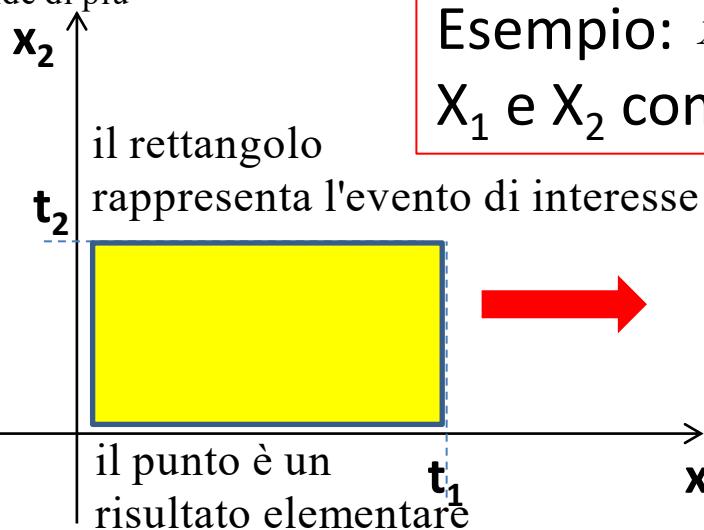
$$F_{X_1 X_2}(t_1, t_2) \hat{=} \Pr\{X_1 \leq t_1 \cap X_2 \leq t_2\}$$

$$0 \leq t_1 < \infty, 0 \leq t_2 < \infty$$

se ad esempio ho 2 processi
in parallelo che devono
concludere un job, la
semplificazione
sta nel rendere $t_1=t_2$.

misura la probabilità congiunta che le durate (X_1 e X_2) delle due
attività siano non superiori a t_1 e t_2 , rispettivamente, quando non è
lecito trattare X_1 e X_2 separatamente, l'una a prescindere dall'altra.
 X_1, X_2 sono correlate (quale che sia)

Esempio: l'attesa in modello client-server è collegata
alla durata dei servizi. Se un servizio dura di più si
attende di più



Esempio: X_1, X_2 processi che comunicano nell'esecuzione
 X_1 e X_2 come durate del viaggio di due amici, in \mathbb{R}^2_+

VIAGGIO come ESPERIMENTO ALEATORIO.
DURATE DEL VIAGGIO come RISULTATI ELEM.,
cioè punti di coordinate (x_1, x_2) .
AREA come EVENTO (CUMUL.) di INTERESSE
l'evento di interesse passa da un intervallo/punto (1 dimensione)
ad un'area (2 dimensioni)

Ipotizzando di voler analizzare il comportamento di due processi che ogni lunedì compiono un job partendo assieme terminando però in tempi diversi. L'esperimento aleatorio è il compimento del job ogni lunedì. Il tempo di conclusione è la durata dell'attività che è misurata nel continuo. Ogni lunedì si hanno realizzazioni diverse delle variabili aleatorie. Come visto, l'area rappresentata sull'asse cartesiano è l'insieme di eventi elementari, in particolare ogni punto delle ascisse indica il tempo di conclusione del primo processo mentre ogni punto delle ordinate indica il tempo di conclusione del secondo processo.

$f(x_1, x_2)$ è la densità della congiunta che il primo processo si concluda tra x_1 e $x_1 + dx_1$ e che il secondo processo termini tra x_2 e $x_2 + dx_2$

$f(x_1, x_2)dx_1dx_2$ è invece la probabilità congiunta che due eventi accadano

La probabilità dell'evento di interesse come sappiamo, si calcola andando a sommare le singole probabilità degli eventi elementari che lo compongono. Essendo nel continuo ed essendo un'area la somma sarà espresse da un integrale doppio.

Cosa devo fare per escludere
che uno o entrambi gli amici
non rientrino prima di un certo tempo sicuramente?

Distribuzione marginale

La distribuzione congiunta può essere ottenuta a partire
dalla definizione preliminare di una funzione densità
congiunta, f ,

$$F_{X_1 X_2}(t_1, t_2) = \int_{u_1=0}^{t_1} \int_{u_2=0}^{t_2} f_{X_1 X_2}(u_1, u_2) du_2 du_1$$

In due dimensioni il volume (risultato dell'integrale) sarà unitario, differentemente dal caso monodimensionale dove è l'area ad essere unitaria.

Funzione di distribuzione marginale:

$$F_{X_i}^{(m)}(t_i) \hat{=} P(X_i \leq t_i) \quad i = 1, 2$$

probabilità che X_i
finisca prima di t_i
quale che sia lo stato
di X_j

A PRESCINDERE DA X_j !
Quale che sia

la m sta per
marginale che
si ricava e non è un
modello primitivo

Supponendo di conoscere la distribuzione congiunta delle
durate dei rispettivi viaggi dei due amici, a cosa può
servire la distribuzione marginale? a calcolare la probabilità che un amico
finisca il viaggio entro t_i , quale che sia
la durata del viaggio del altro

Distribuzione congiunta e distribuzione marginale

Qualora sia data solo la densità congiunta, la distribuzione marginale della X_1 si ottiene così:

Per ottenere la marginale di X_1 , occorre portare all'infinito X_2 in questo modo integro su tutto lo spazio dei valori di X_2 e resta tutto lo spazio di X_1 che è indipendente da X_2

$$\lim_{t_2 \rightarrow \infty} F_{X_1 X_2}(t_1, t_2) = \int_{u_1=0}^{t_1} \int_{u_2=0}^{\infty} f_{X_1 X_2}(u_1, u_2) du_2 du_1 = F_{X_1}(t_1)$$

Si conviene di definire **densità marginale** della X_1 la seguente:

che integrata tra $u_1 = 0, t_1$ ritorna la distribuzione marginale F_{X_1}

$$F_{X_1(t_1)}^{(m)} \triangleq \int_{u_1=0}^{t_1} f_{X_1}(u_1) du_1$$

$$f_{X_1}^{(m)}(u_1) \triangleq \int_{u_2=0}^{\infty} f_{X_1 X_2}(u_1, u_2) du_2 < \infty$$

quindi la marginale è un caso limite dove una densità è definita come integrale indefinito!

Se converge ad un valore finito, si è associato un valore a u_1 . Se ottengo più valori potrei generare una funzione.

È detta marginale in quanto l'attributo/variabile aleatoria dipende da altri fattori

Ripartizione congiunta e marginale nel discreto

RIFERIMENTO:

N e K indicano, rispettivamente, il numero di oggetti presenti in due buffer comunicanti secondo uno schema produttore-consumatore

$$P_{N,K}(n,k) \stackrel{\text{probabilità congiunta}}{=} \Pr\{N = n \cap K = k\} \quad \text{Ripartizione congiunta}$$
$$n = 0, 1, 2, \dots ; \quad k = 0, 1, 2, \dots ;$$

la marginale è pari alla somma dei possibili valori di k

$$P_N^{(m)}(n) \stackrel{\infty}{=} \sum_{k=0} P_{N,K}(n,k) \quad \text{Ripartizione marginale}$$

Ripartizione marginale
CUMULATIVA

$$C_N^{(m)}(l) = \sum_{n=0}^l P_N^{(m)}(n)$$

INDIPENDENZA STOCASTICA

Variabili aleatorie indipendenti se e solo se:

distribuzione congiunta

resta una probabilità

congiunta

$$F_{X_1, X_2}(t_1, t_2) = F_{X_1}^{(m)}(t_1) F_{X_2}^{(m)}(t_2)$$

prodotto tra le marginali
che sono probabilità

saremmo tentati di definire la densità anche come prodotto

$$f_{X_1, X_2}(t_1, t_2) \doteq f_{X_1}^{(m)}(t_1) f_{X_2}^{(m)}(t_2)$$

$$0 \leq t_1 < \infty, \quad 0 \leq t_2 < \infty$$

GIUSTIFICAZIONE:

partendo dalla definizione
di densità, ovvero che
la realizzazione sia compresa
tra t e t più un delta, si va a
giustificare l'indipendenza
stocastica

$$\begin{aligned} & P(t_1 < X_1 \leq t_1 + dt_1 \cap t_2 < X_2 \leq t_2 + dt_2) \\ & \approx f(t_1, t_2) dt_1 dt_2 \doteq f(t_1) dt_1 \cdot f(t_2) dt_2 \\ & = P(t_1 < X_1 \leq t_1 + dt_1) \cdot P(t_2 < X_2 \leq t_2 + dt_2) \end{aligned}$$

Distribuzione congiunta e distribuzione marginale

"Spiegazione" non rigorosa di Trivedi della congiunta

Le funzioni di distribuzione marginale si ricavano a partire dalla distribuzione congiunta:

$$(X_1 \leq t_1) \equiv (X_1 \leq t_1 \cap X_2 < \infty) = \lim_{\substack{\text{rilasso il vincolo} \\ \substack{\text{non finisce per forza dopo } t_1. \\ X_2 < t_1 \text{ è ammesso}}} (X_1 \leq t_1 \cap X_2 \leq t_2)$$

$$P(X_1 \leq t_1) = P\left(\lim_{t_2 \rightarrow \infty} (X_1 \leq t_1, X_2 \leq t_2)\right) = \lim_{t_2 \rightarrow \infty} P(X_1 \leq t_1, X_2 \leq t_2)$$

sfruttando la continuità, ovvero una funzione è continua se il valore del limite nel punto coincide con il valore della funzione.

$$F_{X_1}^{(m)}(t_1) = \lim_{t_2 \rightarrow \infty} F_{X_1, X_2}(t_1, t_2)$$

Se le variabili aleatorie non sono indipendenti lavoro con la probabilità congiunta

$$F_{X_2}^{(m)}(t_2) = \lim_{t_1 \rightarrow \infty} F_{X_1, X_2}(t_1, t_2)$$

Esempio classico: Il problema dell'incontro (1/4)

problema di sincronia

I signori Tizio e Caio si trattengono tutte le sere in un noto bar di Rende, rispettivamente Δt_1 minuti e Δt_2 minuti, nell'intervallo $[0, T]$, sotto le ipotesi seguenti: $0 \leq \Delta t_1 \leq T$, e $0 \leq \Delta t_2 \leq T$.

sono diversi i tempi di attesa che rientrano nell'intervallo

Escludiamo che i due abbiano un appuntamento al bar e siano pure puntuali, ma assumiamo che gli istanti di arrivo, x e y , di entrambi siano il risultato di un comportamento “congiuntamente aleatorio”.

Più precisamente, col termine “congiuntamente aleatorio” intendiamo sintetizzare un comportamento non necessariamente indipendente fra i due e neppure completamente casuale per quanto riguarda l’istante di arrivo al bar di ciascuno dei due.

Io e Giulio tutte le sere alle 19 finiamo e ci dirigiamo al bar dell'università. Senza darci un appuntamento se ci incontriamo al bar ci fermiamo per un aperol. Io aspetto t1 minuti quando arrivo mentre Giulio aspetta t2 minuti. Quindi può capitare che magari una sera io termino alle 18:50 e quindi mi trattengo di meno al bar oppure finisco più tardi e passo quando Giulio se nè già andato via.

Esempio classico: Il problema dell'incontro (2/4)

Usando lo spazio reale euclideo a due dimensioni, vogliamo individuare i sottospazi corrispondenti agli eventi di interesse seguenti:

C= “Tizio arriva non dopo Caio” e D= “Tizio e Caio si incontrano”

Indicando con x e y i rispettivi istanti di arrivo,
risulta:

$$C \hat{=} \{(x, y) \in [0, T] \times [0, T] : x \leq y\}$$

Tizio arriva prima di Caio

$$D \hat{=} D_1 \cup D_2$$

$$\text{con } D_1 \hat{=} \{x, y \in [0, T] : x \leq y \cap y \leq x + \Delta t_1\}$$

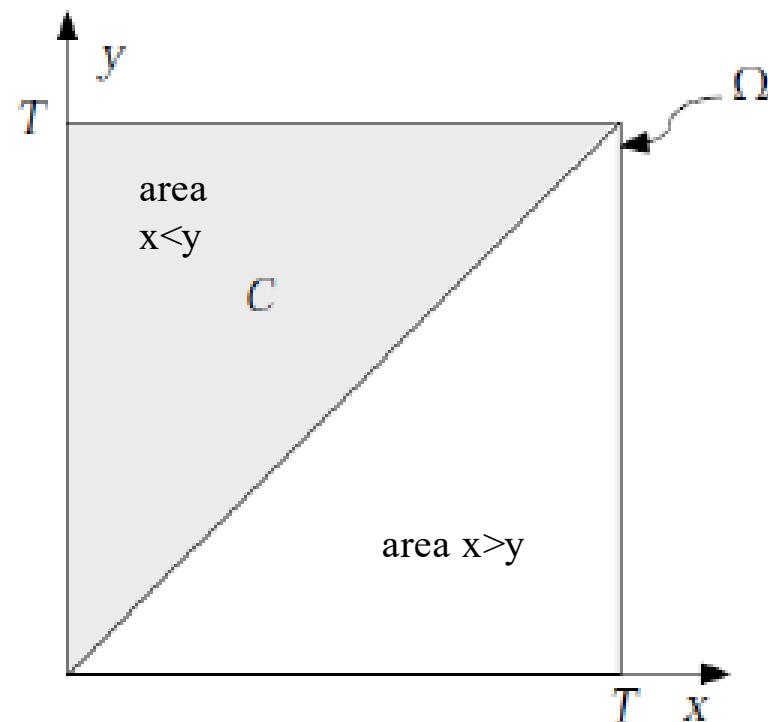
$$\text{e } D_2 \hat{=} \{x, y \in [0, T] : y \leq x \cap x \leq y + \Delta t_2\}$$

Può avvenire anche il contrario. Da notare che
non aspettano la stessa quantità di tempo Δt_i

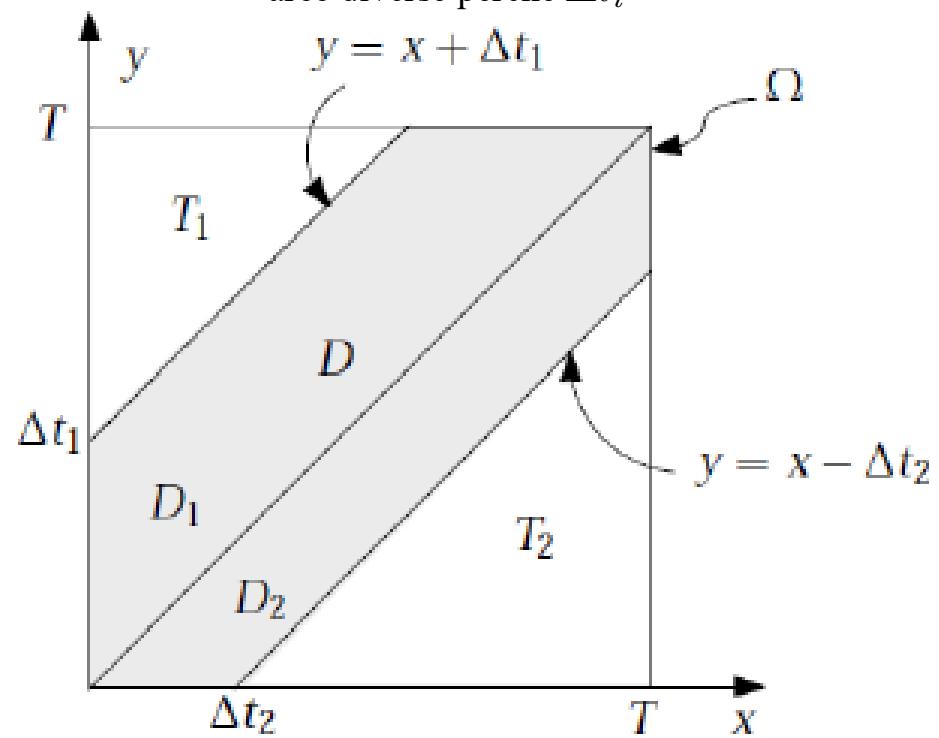
Esempio classico: Il problema dell'incontro (3/4)

RAPPRESENTAZIONE DEGLI EVENTI DI INTERESSE:

sulla diagonale arrivano allo stesso istante



aree diverse perché Δt_i diversi



Avendo una funzione densità che attribuisca probabilità ai risultati elementari, potremmo calcolare la probabilità degli eventi di interesse per integrazione prima su C e poi su D .

Esempio classico: Il problema dell'incontro (4/4)

CALCOLO DELLE PROBABILITA' DEGLI EVENTI DI INTERESSE:

Rinunciando alla densità congiunta, ma con le seguenti ipotesi :

- 1) Tizio e Caio sono indipendenti;
- 2) L'istante di arrivo di ciascuno di loro è la realizzazione di una variabile aleatoria con distribuzione uniforme in $[0, T]$.

RISULTATO DEL CALCOLO NELLE IPOTESI 1) E 2):

$x < y$

Prob("Tizio arriva prima di Caio") =

$$= \text{AREA}_C / \text{AREA}_\Omega = (1/2)T^2 / T^2 = 1/2 \text{ appunto area divisa a metà dalla diagonale nella slide precedente}$$

Prob("Tizio e Caio si incontrano") = $\text{AREA}_D / \text{AREA}_\Omega =$

$$= (\text{AREA}_\Omega - \text{AREA}_{T_1} - \text{AREA}_{T_2}) / \text{AREA}_\Omega$$

$$= T^2 - ((T - \Delta t_1)^2/2) - ((T - \Delta t_2)^2/2) / T^2$$

(Con $T = 60\text{min}$ e $\Delta t_1 = \Delta t_2 = 5\text{min}$ allora Prob("incontro") = 0.139)

Introduction to Probability

CDF=densità

50 PDF=distribuzione
partono assieme

Dimitri P. Bertsekas

General Random Variables

Chap. 3

Si cerca il ritardo di sincronia

Example 3.29. Romeo and Juliet have a date at a given time, and each, independently, will be late by an amount of time that is exponentially distributed with parameter λ . What is the PDF of the difference between their times of arrival?

Let us denote by X and Y the amounts by which Romeo and Juliet are late, respectively. We want to find the PDF of $Z = X - Y$, assuming that X and Y are independent and exponentially distributed with parameter λ . We will first calculate the CDF $F_Z(z)$ by considering separately the cases $z \geq 0$ and $z < 0$ (see Fig. 3.26).

nel esempio non dice
che partono allo stesso
istante temporale, perché
si usano modelli esponenziali
quindi c'è l'assenza di memoria

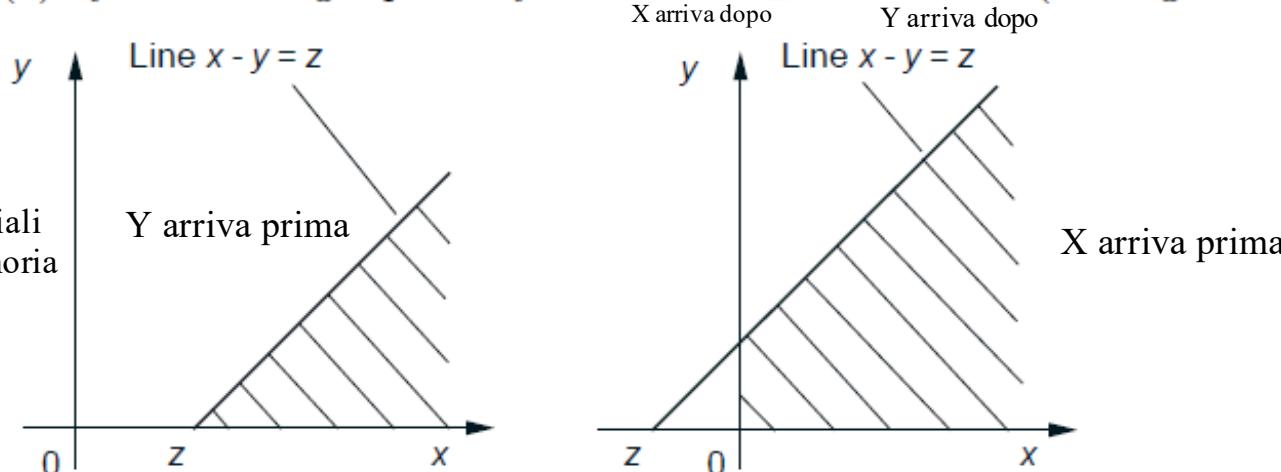


Figure 3.26: The calculation of the CDF of $Z = X - Y$ in Example 3.29. To obtain the value $P(X - Y > z)$ we must integrate the joint PDF $f_{X,Y}(x,y)$ over the shaded area in the above figures, which correspond to $z \geq 0$ (left side) and $z < 0$ (right side).

Combining the two cases $z \geq 0$ and $z < 0$, we obtain

$$F_Z(z) = \begin{cases} 1 - \frac{1}{2}e^{-\lambda z} & \text{if } z \geq 0, \\ \frac{1}{2}e^{\lambda z} & \text{if } z < 0, \end{cases}$$

We now calculate the PDF of Z by differentiating its CDF. We obtain

$$f_Z(z) = \begin{cases} \frac{\lambda}{2}e^{-\lambda z} & \text{if } z \geq 0, \\ \frac{\lambda}{2}e^{\lambda z} & \text{if } z < 0, \end{cases}$$

or

$$f_Z(z) = \frac{\lambda}{2}e^{-\lambda|z|}.$$

This is known as a **two-sided exponential PDF**, also known as the **Laplace PDF**.

For $z \geq 0$, we have (see the left side of Fig. 3.26)

$$\begin{aligned}
F_Z(z) &= \mathbf{P}(X - Y \leq z) \\
&= 1 - \mathbf{P}(X - Y > z) = 1 - \int_0^\infty \left(\int_{z+y}^\infty f_{X,Y}(x,y) dx \right) dy \\
&= 1 - \int_0^\infty \lambda e^{-\lambda y} \left(\int_{z+y}^\infty \lambda e^{-\lambda x} dx \right) dy = 1 - \int_0^\infty \lambda e^{-\lambda y} e^{-\lambda(z+y)} dy \\
&= 1 - e^{-\lambda z} \int_0^\infty \lambda e^{-2\lambda y} dy = 1 - \frac{1}{2} e^{-\lambda z}.
\end{aligned}$$

For the case $z < 0$, we can use a similar calculation, but we can also argue using symmetry. Indeed, the symmetry of the situation implies that the random variables $Z = X - Y$ and $-Z = Y - X$ have the same distribution. We have

$$F_Z(z) = \mathbf{P}(Z \leq z) = \mathbf{P}(-Z \geq -z) = \mathbf{P}(Z \geq -z) = 1 - F_Z(-z).$$

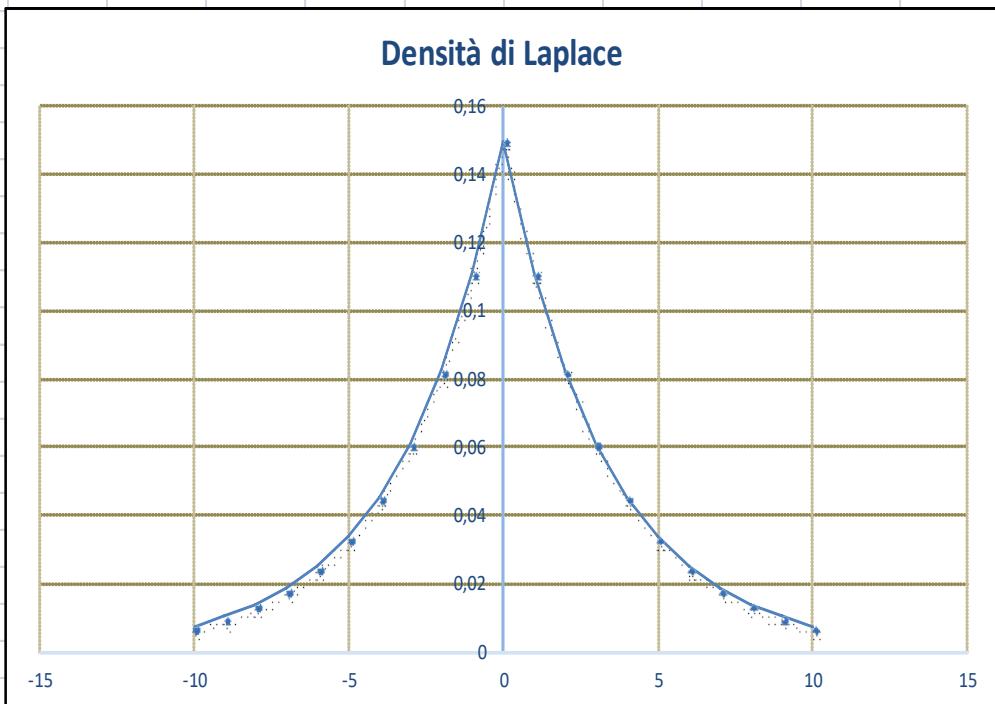
With $z < 0$, we have $-z \geq 0$ and using the formula derived earlier,

$$F_Z(z) = 1 - F_Z(-z) = 1 - \left(1 - \frac{1}{2} e^{-\lambda(-z)} \right) = \frac{1}{2} e^{\lambda z}.$$

Densità di Laplace con EXCEL

Costituita da due esponenziali come visto ed impiegata per problemi di ritardo di sincronia.

z	f(z)
-10	0,007468
-9	0,010081
-8	0,013608
-7	0,018368
-6	0,024795
-5	0,03347
-4	0,045179
-3	0,060985
-2	0,082322
-1	0,111123
0	0,15
1	0,111123
2	0,082322
3	0,060985
4	0,045179
5	0,03347
6	0,024795
7	0,018368
8	0,013608
9	0,010081
10	0,007468



La distribuzione congiunta in Affidabilità

(nei modelli di sopravvivenza)

$$F_{X_1, X_2}(t, t) \hat{=} P(X_1 \leq t \cap X_2 \leq t) \quad 0 \leq t < \infty$$

Può essere usata per esprimere la probabilità che entrambi i componenti non sopravvivano oltre l'istante t , nell'ipotesi che agiscano sia cause di guasto individuale sia cause di guasto comune

NON INTERESSA LA DISPOSIZIONE SE IN SERIE O PARALLELO PERCHÈ HO CAUSE DI GUASTO INDIVIDUALI!

AFFIDABILITÀ CONGIUNTA DEI DUE COMPONENTI:
(sopravvivenza)

$$R(t, t) \hat{=} P(X_1 > t \cap X_2 > t)$$

t è il tempo del PRIMO guasto, ma non si sa quale dei due componenti si guasti per primo!

Modello primitivo quando non è possibile caratterizzare separatamente il tempo al guasto di uno qualunque dei due componenti, indipendentemente dall'altro.

Nel caso ad una dimensione $R(t)=1-F(t)$. Vale anche in questo caso? NO

La distribuzione congiunta in Affidabilità

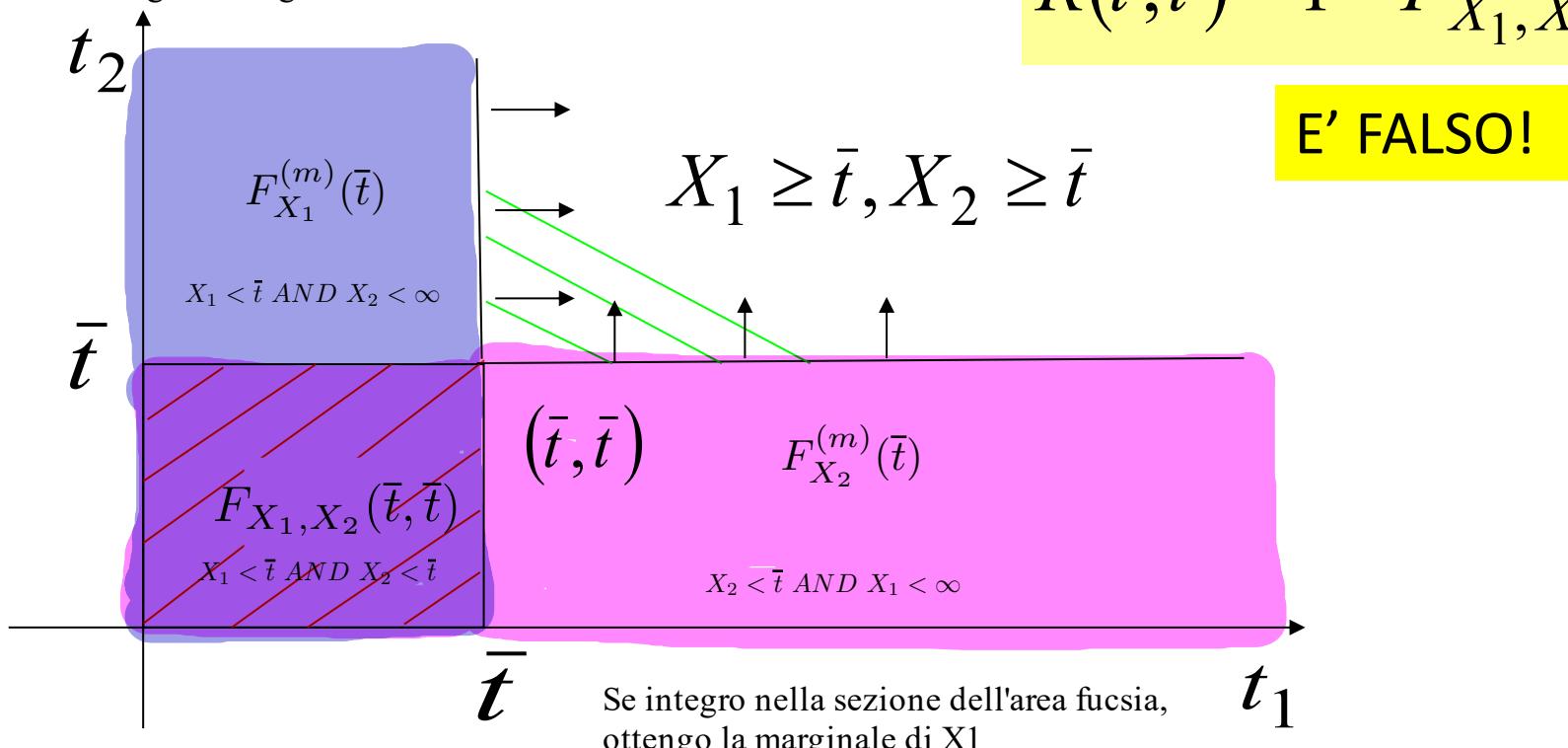
$$R(\bar{t}, \bar{t}) \hat{=} P((X_1 \cap X_2) > \bar{t}) \hat{=} \int_{u_1=\bar{t}}^{\infty} \int_{u_2=\bar{t}}^{\infty} f_{X_1, X_2}(u_1, u_2) du_1 du_2$$

Il complemento a entrambi che tornano prima di t, non è tutti e due non tornano dopo di t, ma che uno dei due torni prima di t

Importante

Se integro nella sezione dell'area blu, ottengo la marginale di X_2

$$R(\bar{t}, \bar{t}) = 1 - F_{X_1, X_2}(\bar{t}, \bar{t})$$



Distribuzione congiunta e distribuzione marginale

Viaggi, incontri e ritardi di sincronia a parte,

QUANDO ANCORA SI USA UNA DISTRIBUZIONE CONGIUNTA?

- 1) quando si voglia definire una successione di variabili aleatorie su di uno stesso spazio dei risultati, al fine di rappresentare la successione delle durate di uno stesso fenomeno che si ripete nel tempo (degenze ripetute in ospedale, fidanzamento e matrimonio);
- 2) quando si voglia considerare uno spazio dei risultati che sia lo spazio "prodotto" di coppie (o n-ple) di realizzazioni di variabili aleatorie che rappresentano "cose" diverse ma logicamente dipendenti fra di loro (attesa dal parrucchiere e numero di persone trovate al proprio arrivo);

Distribuzione della somma di variabili aleatorie

Si vuole calcolare la funzione di distribuzione di una somma di variabili aleatorie, ovvero di una funzione di variabili aleatorie

(indipendenti)

L'ipotesi di indipendenza nascerà più avanti

$$\{(x_1, x_2); x_1 \geq 0, x_2 \geq 0\}$$

v.a. tempo che passa dall'inizio della prima attività alla fine della seconda attività

$$X \hat{=} X_1 + X_2 \Rightarrow X \leq t$$

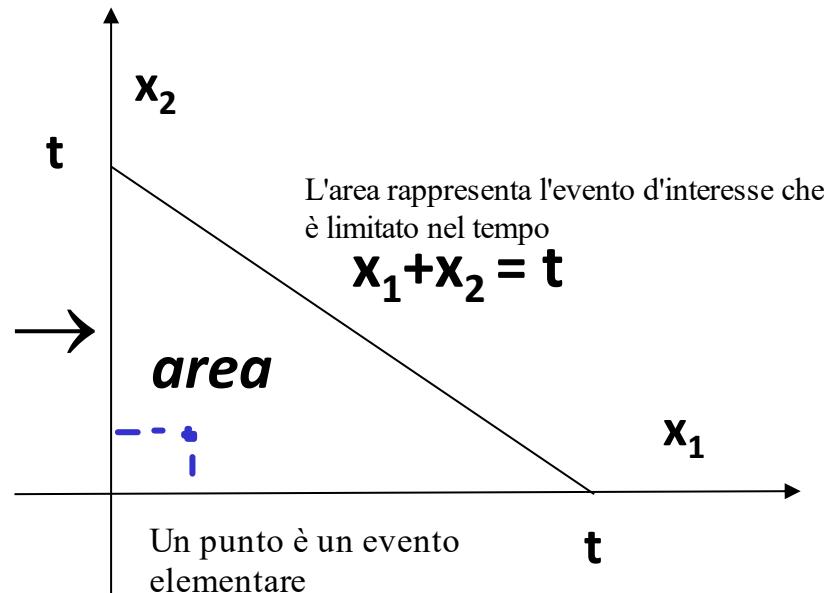
Durata complessiva
evento d'interesse

$$F_X(t) \hat{=} \Pr\{X_1 + X_2 \leq t\}$$

Per ipotesi le variabili sono ancora dipendenti tra loro quindi utilizzo la congiunta

$$F_X(t) = \int_{x_2=0}^t \int_{x_1=0}^{t-x_2} f_{X_1, X_2}(x_1, x_2) dx_1 dx_2$$

$$\equiv \int_{x_1=0}^t \int_{x_2=0}^{t-x_1} f_{X_1, X_2}(x_1, x_2) dx_1 dx_2$$



$$F_X(t) = \iint_{\text{area}} f_{X_1, X_2}$$

**SENZA INDIPENDENZA
E' LA FORMULA FINALE!**

Distribuzione della somma di variabili aleatorie (indipendenti)

Mi gioco l'ipotesi di indipendenza,
quindi passo a prodotto di densità individuali come visto nella lezione precedente

$$\int_{x_1=0}^t \int_{x_2=0}^{t-x_1} f_{X_1, X_2}(x_1, x_2) dx_2 dx_1 = \int_{x_1=0}^t \left(\int_{x_2=0}^{t-x_1} f_{X_2}(x_2) dx_2 \right) f_{X_1}(x_1) dx_1$$

probabilità che il componente attivo si guasti entro t e che il

**X1 e X2 variabili
aleatorie
indipendenti !**

In verde probabilità che
X2 termini prima di
t-x1. In arancione che
il primo componente si
guasti entro x1
Non è escluso
che x1 duri t-epsilon
o che duri 0+epsilon

Se conoscessi l'istante in cui X1 si guasta
l'integrale non sarebbe necessario, basterebbe $x_2=0$
 $F_{X_2}(t - \bar{x}_1) f_{X_1}(\bar{x}_1) dx_1$

$$\int_{x_1=0}^t F_{X_2}(t - x_1) f_{X_1}(x_1) dx_1 =$$

integrale di convoluzione
La convoluzione nasce dall'indipendenza
 t oppure se il primo integrale è tra $x_1 = 0, t - x_2$

$$= \int_{x_2=0}^t F_{X_1}(t - x_2) f_{X_2}(x_2) dx_2$$

$$\int_{x_1=0}^t F_{X_2}(t - x_1 | x_1) \cdot f_{X_1}(x_1) dx_1$$

sarebbe una probabilità totale
La sostituzione $F_{X_2}(t - x_1) = F_{X_2}(t - x_1 | x_1)$
vale perché si ha indipendenza

**convoluzione delle funzioni densità
dell'una con la distribuzione dell'altra**

Convoluzione di esponenziali identiche

e indipendenti

Sommando n esponenziali

$$F_{X_i}(t) = 1 - e^{-\lambda t} \quad i = 1, 2, \dots, n \quad \rightarrow \quad X \doteq X_1 + X_2 + \dots + X_n$$

TESI:

$$f_X(t) = \frac{(\lambda t)^{n-1}}{(n-1)!} \lambda e^{-\lambda t}$$

DENSITA'

DI ERLANG DI ORDINE "n"

PROVA:

per induzione

$$\text{Posto: } Y \doteq X_1 + X_2 + \dots + X_{n-1} \quad \Rightarrow \quad X = Y + X_n$$

$$\Rightarrow f_X(t) = \frac{d}{dt} F_X(t) = \frac{d}{dt} \int_{u=0}^t F_{X_n}(t-u) \cdot f_Y(u) \cdot du$$

$$\text{derivando} \quad = \int_{u=0}^t f_{X_n}(t-u) f_Y(u) du = \int_{u=0}^t f_{X_n}(t-u) \cdot \frac{(\lambda u)^{n-2}}{(n-2)!} e^{-\lambda u} \lambda \cdot du$$

$$= \int_{u=0}^t \lambda e^{-\lambda(t-u)} \cdot e^{-\lambda u} \frac{(\lambda u)^{n-2}}{(n-2)!} d(\lambda \downarrow u) = \lambda e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!}$$

Convoluzione di esponenziali identiche

non è domanda

$$F_X(t) = \int_0^t e^{-\lambda u} \frac{(\lambda u)^{n-1}}{(n-1)!} d(\lambda u) \quad \text{Si risolve per parti}$$

$$= e^{-\lambda t} \frac{(\lambda t)^n}{n!} + \int_0^t e^{-\lambda u} \frac{(\lambda u)^n}{n!} d(\lambda u) = e^{-\lambda t} \frac{(\lambda t)^n}{n!} + \int_0^t e^{-\lambda u} d\left(\frac{(\lambda u)^{n+1}}{(n+1)!} \right)$$

$$= e^{-\lambda t} \frac{(\lambda t)^n}{n!} + e^{-\lambda t} \frac{(\lambda t)^{n+1}}{(n+1)!} + \int_0^t e^{-\lambda u} d\left(\frac{(\lambda u)^{n+2}}{(n+2)!} \right) = \dots$$

$$\dots = e^{-\lambda t} \sum_{i=n}^{\infty} \frac{(\lambda t)^i}{i!}$$

**DISTRIBUZIONE
DI ERLANG DI ORDINE "n"**

Dopo tutto quanto ottengo

Il modello di Erlang -n

Densità:

$$f_X(t) = \frac{(\lambda t)^{n-1}}{(n-1)!} \lambda e^{-\lambda t}$$

Distribuzione:

$$F_X(t) = \int_0^t e^{-\lambda u} \frac{(\lambda u)^{n-1}}{(n-1)!} d(\lambda u) \quad \dots = e^{-\lambda t} \sum_{i=n}^{\infty} \frac{(\lambda t)^i}{i!}$$

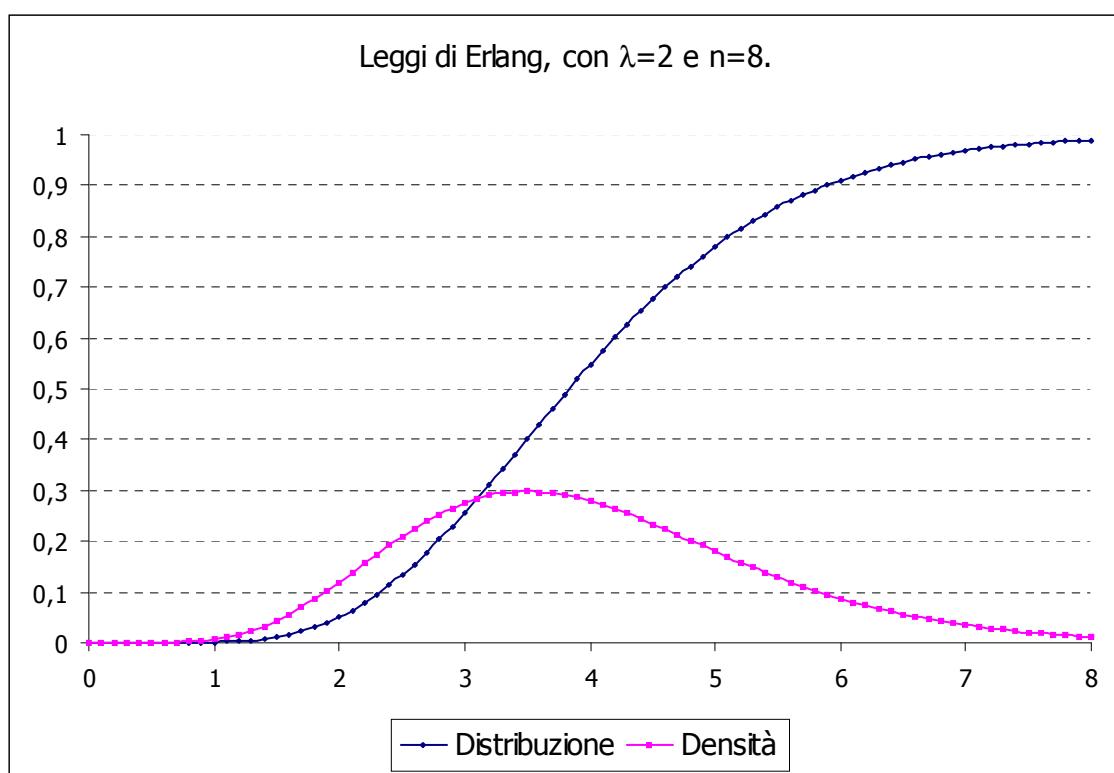
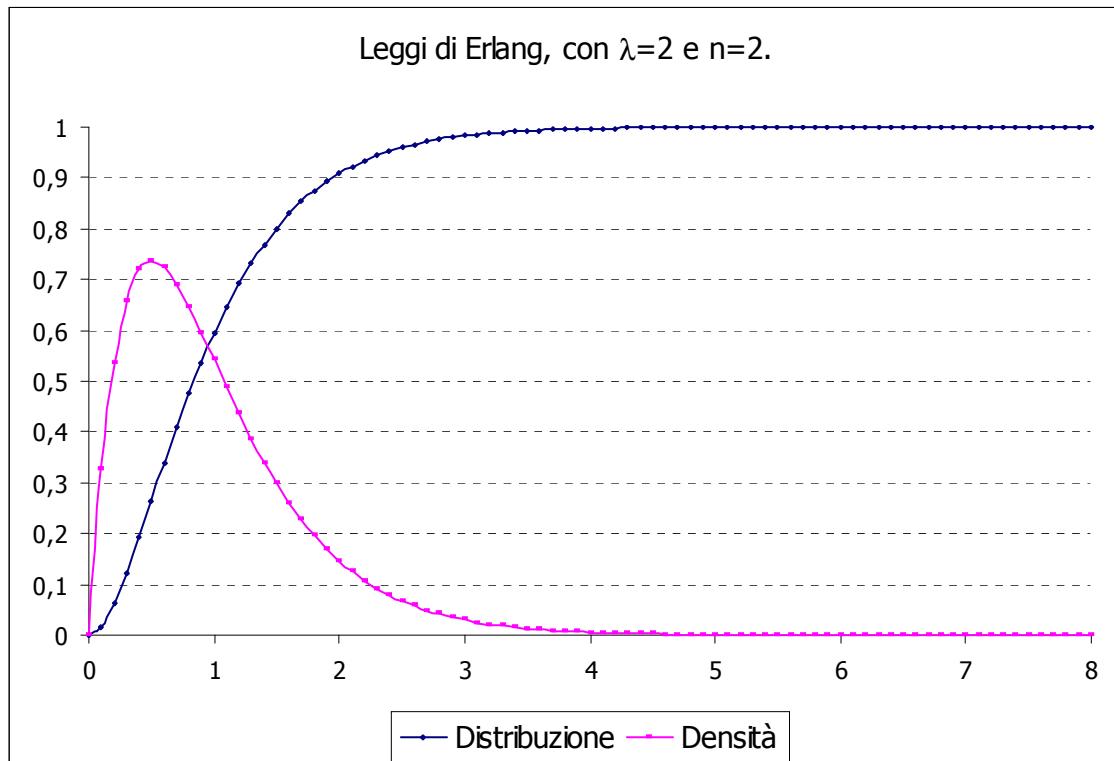
$$= e^{-\lambda t} \left(\sum_{i=0}^{\infty} \frac{(\lambda t)^i}{i!} - \sum_{i=0}^{n-1} \frac{(\lambda t)^i}{i!} \right) = e^{-\lambda t} \left(e^{\lambda t} - \sum_{i=0}^{n-1} \frac{(\lambda t)^i}{i!} \right)$$

ci interessa fermarci a n-1

$$= 1 - e^{-\lambda t} \sum_{i=0}^{n-1} \frac{(\lambda t)^i}{i!}$$

quindi questa è la distribuzione che si usa per fare calcoli
Erlang non ha assenza di memoria perché mi dice la
posizione in cui mi trovo, grazie all'indice!

Illustrazione del Modello di Erlang (1)



Modulare la varianza attorno alla stessa media

Sia X , la v.a. che rappresenta la durata di un fenomeno e ipotizziamo di conoscere solo la durata media: $E[X]$, e di non avere informazioni sulla varianza. Di solito sceglioamo di rappresentare la durata di quel fenomeno in base ad una legge esponenziale di parametro pari al reciproco di $E[X]$.

Così facendo, implicitamente, adottiamo una varianza che è: $VAR[X] = E[X]^2$

Adesso si farà vedere che possiamo fare di meglio!

Pensiamo alla X come:

$$X \doteq X_1 + X_2 + \dots + X_n \quad \text{con} \quad X_i \approx \exp(\lambda) \quad i = 1, \dots, n \quad (\text{anche indipendenti})$$

e, sapendo che: $E[X] = n E[X_i] = n \frac{1}{\lambda}$, $VAR[X] = n VAR[X_i] = n \frac{1}{\lambda^2}$

osserviamo che, ponendo:

$$\lambda \doteq n \frac{1}{E[X]}$$

riusciamo a mantenere lo stesso valore medio $E[X] = n \frac{1}{\lambda} = n \frac{E[X]}{n}$, ma a

modificare la varianza: $VAR[X] = n \frac{E[X]^2}{n^2} = \frac{E[X]^2}{n}$ rispetto al valore che si otterrebbe

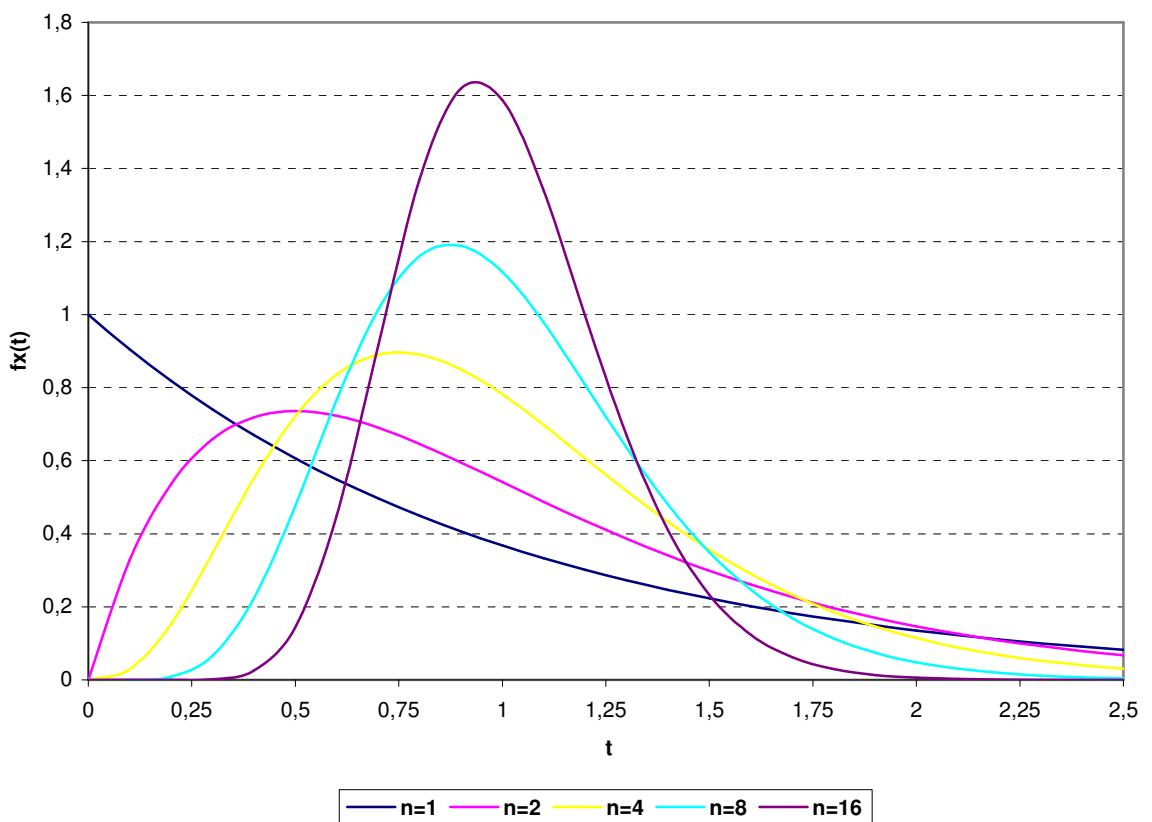
rappresentando la X con un'unica legge esponenziale.

Usando "n" come parametro di nostra scelta, possiamo far diminuire quanto vogliamo la varianza (rispetto a quella originariamente prevista dalla legge esponenziale).

Più precisamente, passiamo ad una rappresentazione del fenomeno basata su una "opportuna" distribuzione di Erlang di ordine "n", che cambia forma al crescere di "n", però mantiene sempre la stessa media.

Dunque la chiameremo Legge di Erlang MODULATA.

Densità di Erlang "modulata"



Esempio

Si supponga che un componente, una volta guasto, possa essere riparato secondo un processo di riparazione che si compone di quattro fasi in sequenza. La durata media dell'intero processo è stimata in 2 unità di tempo e non si hanno informazioni sulle durate parziali medie delle singole fasi: potrebbero corrispondere ciascuna a 0.5 unità di tempo.

Proporre una funzione densità per la variabile aleatoria "durata del processo di riparazione" e calcolare la probabilità che una riparazione duri non più di 3 unità di tempo.

Rappresentare con un grafico in Excel la suddetta densità e discutere della differenza di andamento in confronto ad una densità esponenziale caratterizzata dalla stessa media. E cosa si può dire a proposito della varianza?

Affidabilità dei sistemi con riserve pronte

Componenti NON identici quindi affidabilità diversa. Ad esempio la riserva è più scadente per risparmiare soldi



X = “Vita del sistema”;

X_1 = “Vita comp. att.”;

X_2 = “Vita riserva att.”;

La commutazione è il passaggio da componente attivo a componente di riserva

- la commutazione sia istantanea e che riesca con certezza = PERFETTA
- la riserva non possa subire guasti quando permane pronta.

$$\Rightarrow X \doteq X_1 + X_2$$

\downarrow tempo all'ultimo guasto

$$F_X(t) = \int_{x_1=0}^t [1 - e^{-\lambda_2(t-x_1)}] \lambda_1 e^{-\lambda_1 x_1} dx_1$$

\uparrow

$$X_i \approx \exp(\lambda_i) \quad i = 1, 2.$$

$$R(t) = 1 - F_X(t) = 1 - \left(\frac{\lambda_2}{\lambda_2 - \lambda_1} e^{-\lambda_1 t} + \frac{\lambda_1}{\lambda_1 - \lambda_2} e^{-\lambda_2 t} \right)$$

Affidabilità dei sistemi con riserve pronte

GENERALIZZAZIONE:

“n-1” componenti di riserva al componente attivo. Tutti identici.

$$X \doteq X_1 + X_2 + \dots + X_n \quad X_i \approx \exp(\lambda) \quad i = 1, \dots, n$$



$$R(t) = e^{-\lambda t} \sum_{i=0}^{n-1} \frac{(\lambda t)^i}{i!}$$

ottenuto da $R(t) = 1 - F_X(t)$

Affidabilità nel
MODELLO DI ERLANG

Ipotesi:

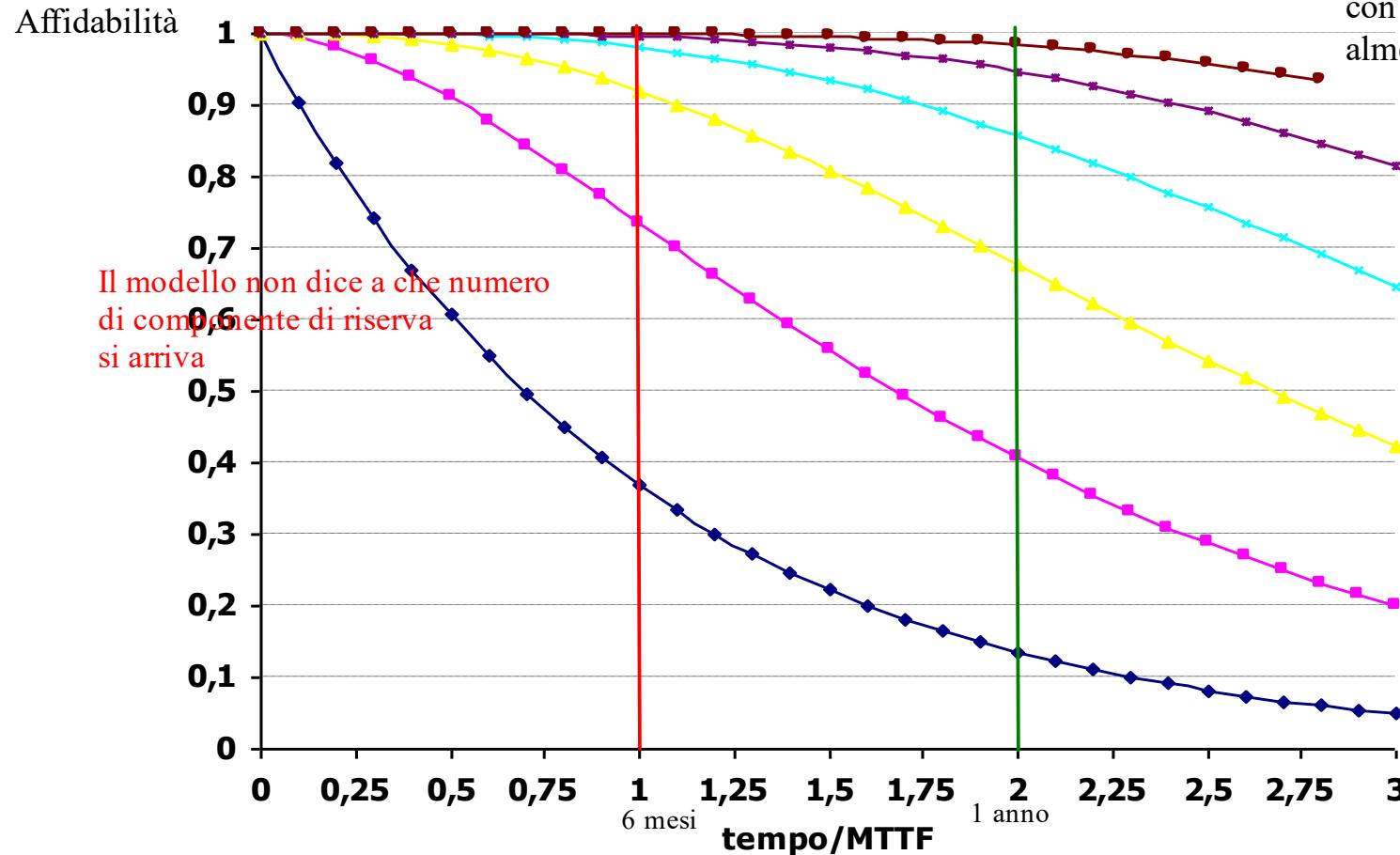
- *la commutazione riesce sicuramente;*
- *la commutazione si realizza in un tempo trascurabile ;*
- *i componenti sono identici e non riparabili;*
- *la legge comune di guasto è esponenziale di parametro λ .*

Affidabilità dei sistemi con riserve pronte

Fissato un istante di tempo calcolo la probabilità che:

1. ho rotto tutti i componenti
2. arrivo all'istante fissato con probabilità > 0.9 con almeno una riserva

**Affidabilità di sistema con un componente attivo e $n-1$ di riserva
(modello di Erlang)**



Fisso l'istante di tempo a 1 che corrisponde a 6 mesi ad esempio. Allora per ogni curva posso calcolare la probabilità di che tutti i componenti siano rotti come 1-affidabilità%, ad esempio per la blu ho $1-0.38=0.62$. Mentre se volessi superare con 0.9 di probabilità i 6 mesi dovrei adoperare un sistema con almeno 3-2 di riserva, se invece volessi superare un anno dovrei adoperare un sistema 5-4 di riserva.

\bullet $n=1$ - 0 di riserva	\blacksquare $n=2$ - 1 di riserva	\blacktriangle $n=3$ - 2 di riserva
\times $n=4$ - 3 di riserva	\blacksquare $n=5$ - 4 di riserva	\bullet $n=6$ - 5 di riserva

Il modello non dice a quale di componente di riserva si arriva, anche perchè non comunica quando si rompono i componenti!

La distribuzione condizionata

Date due variabili aleatorie X e Y , con Y continua, la funzione di distribuzione condizionata $F_{Y/X}(y/x)$ viene introdotta per misurare la probabilità dell'evento " $Y \leq y / X = x$ ". Per proporre una definizione formale che si riferisca al caso in cui la X è una variabile aleatoria continua con funzione di distribuzione ottenuta integrando la densità, si può pensare all' evento condizionante come ad un evento limite :

$$\begin{array}{ll} "Y \leq y / X = x" \hat{=} \lim_{\Delta x \rightarrow 0} "Y \leq y / x \leq X \leq x + \Delta x" & Y \text{ è la condizionata} \\ & X \text{ è la condizionante} \end{array}$$

e assumere che valga la seguente:

$$P(\lim_{\Delta x \rightarrow 0} "Y \leq y / x \leq X \leq x + \Delta x") = \lim_{\Delta x \rightarrow 0} P(Y \leq y / x \leq X \leq x + \Delta x)$$

Cosicché :

$$F_{Y/X}(y/x) \hat{=} \lim_{\Delta x \rightarrow 0} P(Y \leq y / x \leq X \leq x + \Delta x)$$

Con la definizione proposta, la distribuzione condizionata è posta in relazione ad un evento condizionante la cui probabilità è misurabile con gli strumenti a nostra disposizione e si può ricavare la seguente formula di calcolo:

$$F_{Y|X}(y|x) = \int_{v=0}^y \frac{f_{Y,X}(v,x)}{f_X^{(m)}(x)} dv$$

Da cui è immediato riconoscere la funzione densità condizionata :

$$f_{Y/X}(v/x) \hat{=} \frac{f_{Y,X}(x,v)}{f_X^{(m)}(x)}$$

Dalla formula di calcolo è possibile ricavare la densità marginale per la variabile aleatoria Y

$$F_Y^{(m)}(y) = \int_{x=0}^{\infty} F_{Y|X}(y|x) \cdot f_X^{(m)}(x) dx \quad \text{dove ad esempio se ha l'attesa di un client Y}$$

e la durata del servizio X , l'integrale viene calcolato per tutti i possibili valori di durata del servizio. Inoltre la margina di X rappresenta la probabilità che il servizio arrivi all'istante x e finisca subito dopo entro un dx . Infine sorge una domanda: il tempo di attesa può essere calcolato come un modello primitivo? È linearizzabile? La risposta in entrambi i casi è no

Derivazione della formula di calcolo

Prima di ricavarla, è bene far vedere che la probabilità dell'evento condizionante è misurabile con la densità marginale:

$$\begin{aligned} P(x \leq X \leq x + \Delta x) &= \lim_{y \rightarrow \infty} P(x \leq X \leq x + \Delta x \cap Y \leq y) \\ &= \int_{u=x}^{x+\Delta x} \int_{v=0}^{\infty} f_{Y,X}(v,u) dv du = f_X^{(m)}(x) \Delta x \end{aligned}$$

A questo punto:

Formula concettuale della probabilità condizionata per Trivedi

$$F_{Y/X}(y/x) \triangleq \lim_{\Delta x \rightarrow 0} P(Y \leq y / x \leq X \leq x + \Delta x)$$

$$\begin{aligned} &= \lim_{\Delta x \rightarrow 0} \frac{P(Y \leq y \cap x \leq X \leq x + \Delta x)}{\lim_{y \rightarrow \infty} P(x \leq X \leq x + \Delta x \cap Y \leq \infty)} \\ &= \lim_{\Delta x \rightarrow 0} \frac{\int_{v=0}^y \int_{u=x}^{x+\Delta x} f_{Y,X}(v,u) du dv}{\Delta x f_X^{(m)}(x)} \\ &= \lim_{\Delta x \rightarrow 0} \frac{\int_{v=0}^y f_{Y,X}(v, \bar{u}) \Delta x \, dv}{\Delta x f_X^{(m)}(x)} \quad \text{ove } x \leq \bar{u} \leq x + \Delta x \end{aligned}$$

$$= \int_{v=0}^y \frac{f_{Y,X}(v,x)}{f_X^{(m)}(x)} \, dv$$

Formula della distribuzione totale

La formula della distribuzione totale mette in relazione la funzione di distribuzione marginale $F_Y^{(m)}(y)$ con la funzione di distribuzione condizionata $F_{Y|X}(y/x)$, purché si conosca pure la densità marginale della X . È una formula alla quale si ricorre spesso, perché è quasi sempre più facile conoscere direttamente la distribuzione condizionata che non la marginale. Da un punto di vista teorico, non è altro che l'estensione della formula della probabilità totale alle variabili aleatorie.

Per ricavarla, si può partire dalla relazione:
$$F_Y^{(m)}(y) = \lim_{x \rightarrow \infty} F_{Y,X}(y,x)$$

Ricordando che:

$$F_{Y,X}(y,x) \triangleq \int_{v=0}^y \int_{u=0}^x f_{Y,X}(v,u) du dv$$

si ha:

da notare come

l'integrale rispetto

x cambi estremo superiore

$$F_Y^{(m)}(y) = \int_{v=0}^y \int_{x=0}^{\infty} f_{Y,X}(v,x) dx dv = \int_{v=0}^y \left[\int_{x=0}^{\infty} f_{Y|X}(v|x) f_X^{(m)}(x) dx \right] dv$$

cambio l'ordine
densità congiunta

$$= \int_{x=0}^{\infty} \left[\int_{v=0}^y f_{Y|X}(v|x) dv \right] \cdot f_X^{(m)}(x) dx$$

e da qui la formula finale:

$$F_Y^{(m)}(y) = \int_{x=0}^{\infty} F_{Y|X}(y|x) f_X^{(m)}(x) dx$$

A parole: la probabilità del risultato $Y \leq y$ è pari alla somma - su tutti i risultati condizionanti $0 \leq x < \infty$ - dei prodotti tra la probabilità di ciascun risultato condizionato $Y \leq y | X \in (x, x+dx)$ e la probabilità del rispettivo condizionante $X \in (x, x+dx)$.

Infine, per derivazione della formula finale, si ottiene la formula della densità totale:

$$f_Y^{(m)}(y) = \int_{x=0}^{\infty} f_{Y|X}(y|x) f_X^{(m)}(x) dx$$

Quando la variabile condizionante X ha realizzazioni non negative, ma nel discreto, allora le formule di distribuzione totale e di densità totale diventano le seguenti:

$$F_Y^{(m)}(y) = \sum_{x=0}^{\infty} F_Y(y|x) \cdot P_X^{(m)}(x) \quad f_Y^{(m)}(y) = \sum_{x=0}^{\infty} f_Y(y|x) \cdot P_X^{(m)}(x).$$

Commutazione imperfetta ed affidabilità

Utilizziamo la condizionante

Un componente attivo ed uno di riserva (identici)

v.a. discreta
condizionante

$$X = \begin{cases} 1 & \hat{\text{commutazione riesce}} \\ 0 & \hat{\text{commutazione fallisce}} \end{cases}$$

Ho solo una chance per commutare i componenti per semplicità

Si cerca la distribuzione del tempo di vita del sistema dove

Y = "tempo di vita del sistema"

Passo 1, uso la distribuzione totale

$$① F_{Y|X=0}(t) = 1 - e^{-\lambda t};$$

esprime il vincolo che la commutazione non riesce, ovvero il sistema dura fino a quando il componente attivo dura

Distribuzione totale:

non si usa l'integrale!

$$F_Y^{(m)}(t) = F_{Y|X=1}(t) \cdot P_X(1) + F_{Y|X=0}(t) \cdot P_X(0)$$

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

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

in base al valore di c mi riconduco al caso ① oppure ②

Ipotesi numero 1 si ha un

Meccanismo bernoulliano
per la commutazione

$$P_X(1) = c, \quad P_X(0) = 1 - c$$

c è detto fattore di coverage

Ipotesi numero 2:

Leggi di guasto esponenziali
perchè si usa la convoluzione quindi userò Erlang

$$① F_{Y|X=1}(t) = 1 - e^{-\lambda t} \sum_{i=0}^{2-1} \frac{(\lambda t)^i}{i!} = 1 - e^{-\lambda t} (1 + \lambda t)$$

uso Erlang

La 0) è la mortalità, mentre la 1) è l'affidabilità

UN SOLO TENTATIVO!

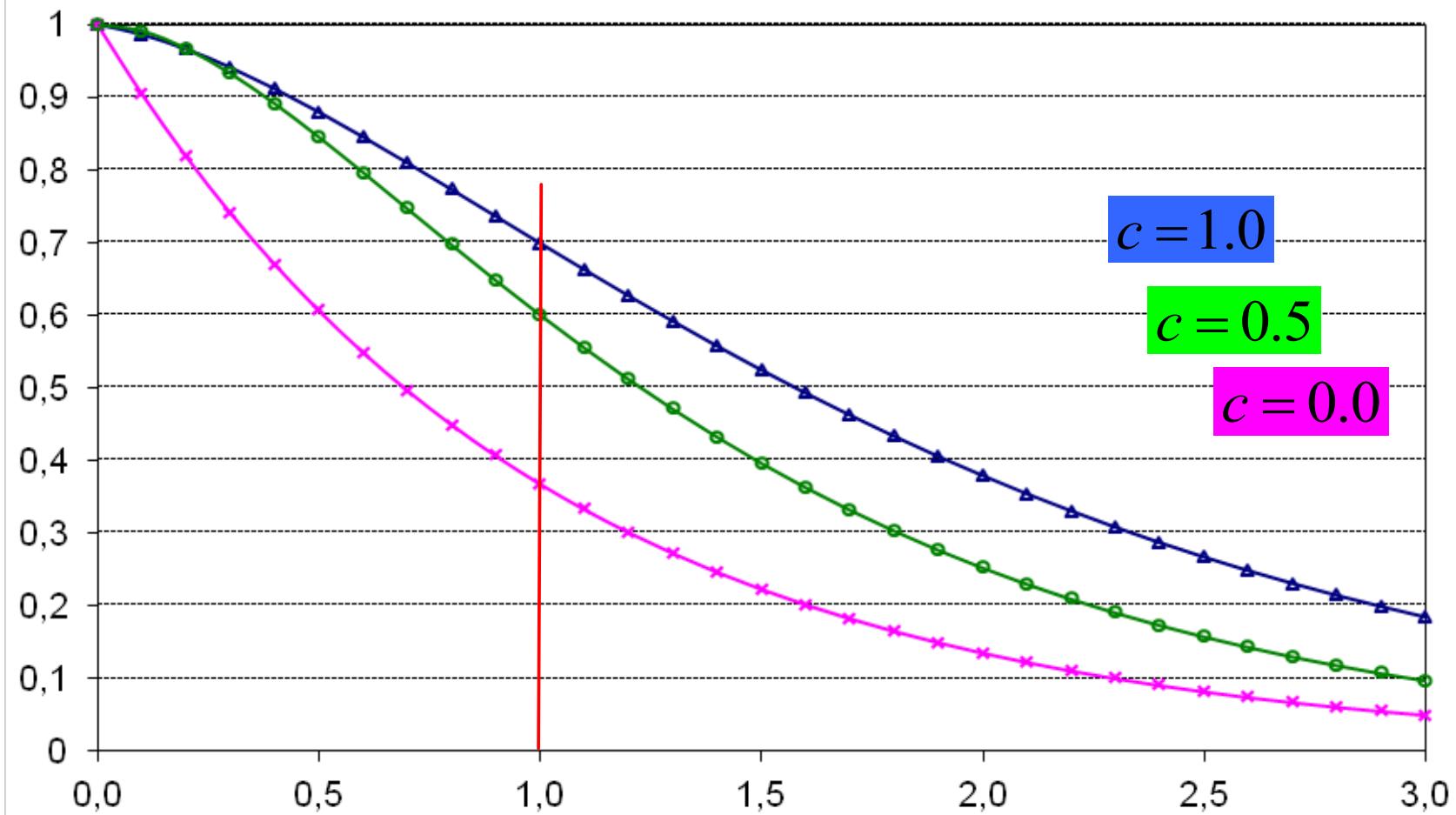
$$\Rightarrow R(t) = (1 + c \cdot \lambda t) \cdot e^{-\lambda t}$$

se $c=0$ ho che $R(t) = e^{-\lambda t}$

Effetto del fattore “c” sulla curva di affidabilità

Si nota dal grafico che c influenza l'affidabilità, se si vuole un'affidabilità del 50%, con $c=0$ si ha fino a circa 750 ore mentre con $c=1$, si ha fino a 1500 ore. Praticamente il doppio della vita!

$$R(t) = (1 + c\lambda t)e^{-\lambda t}, \quad t \geq 0$$



Da notare come a 1000 ore l'affidabilità non ha **Tempo (ore x mille)** divario molto grande tra $c=0.5$ e $c=1$

Illustrazione del “peso” della commutazione (1/2)

$$R_{PAR}(t) = \left(e^{-\lambda \cdot t} + e^{-\lambda \cdot t} - e^{-(\lambda + \lambda) \cdot t} \right)$$

$$R_{RIS}(t) = (1 + c\lambda t)e^{-\lambda t}, \quad t \geq 0$$

I parallelo avrei
un sistema m-out-of-n.
Con buone prestazioni
e cosiddetta ridondanza
attiva.

Secondo componente in parallelo o di riserva ?

(risultato con $c = 0,9$ e $1/\lambda = 2000$ ore)

MEGLIO IN RISERVA!

fai più strada con riserva ma ottieni meno prestazioni sul lungo termine, infatti all'inizio le prestazioni sono quasi identiche

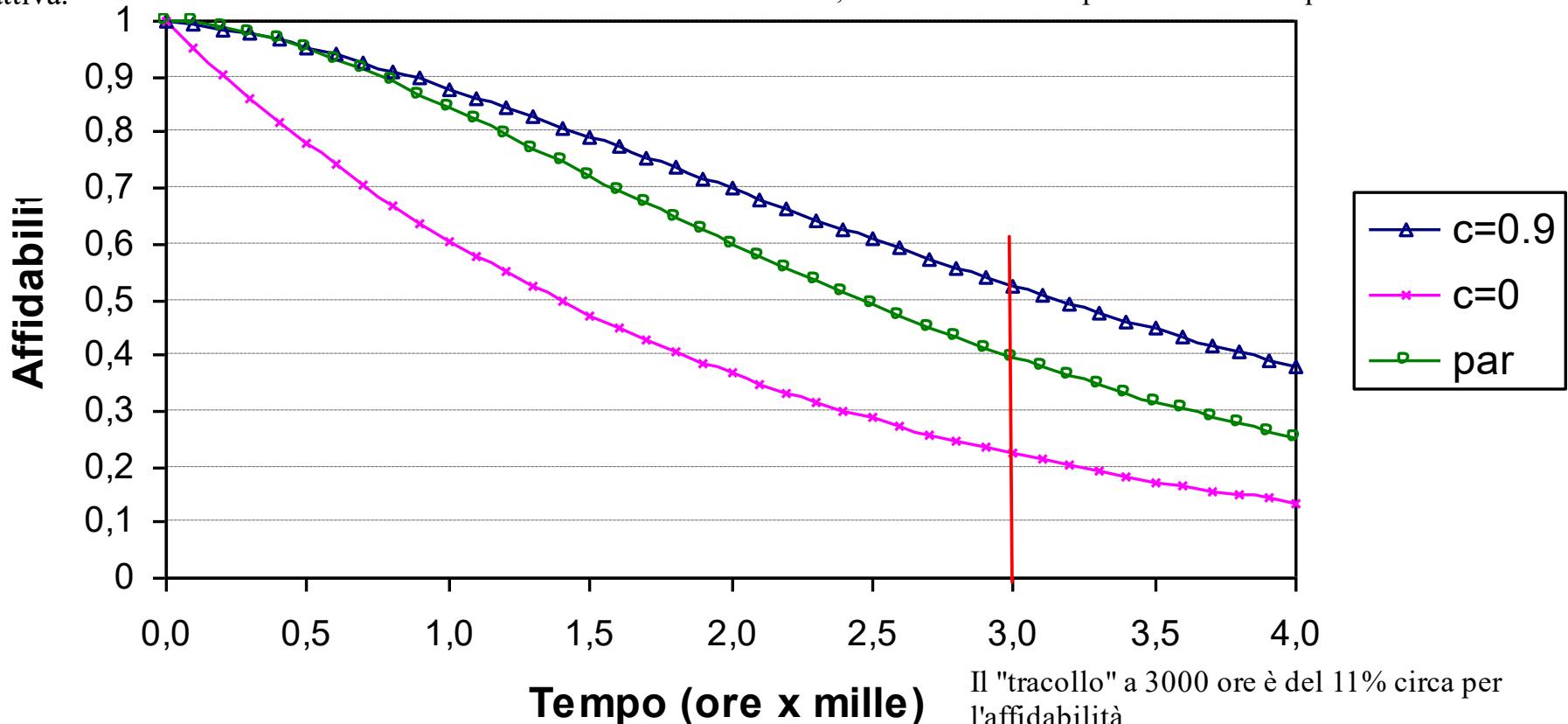


Illustrazione del “peso” della commutazione (2/2)

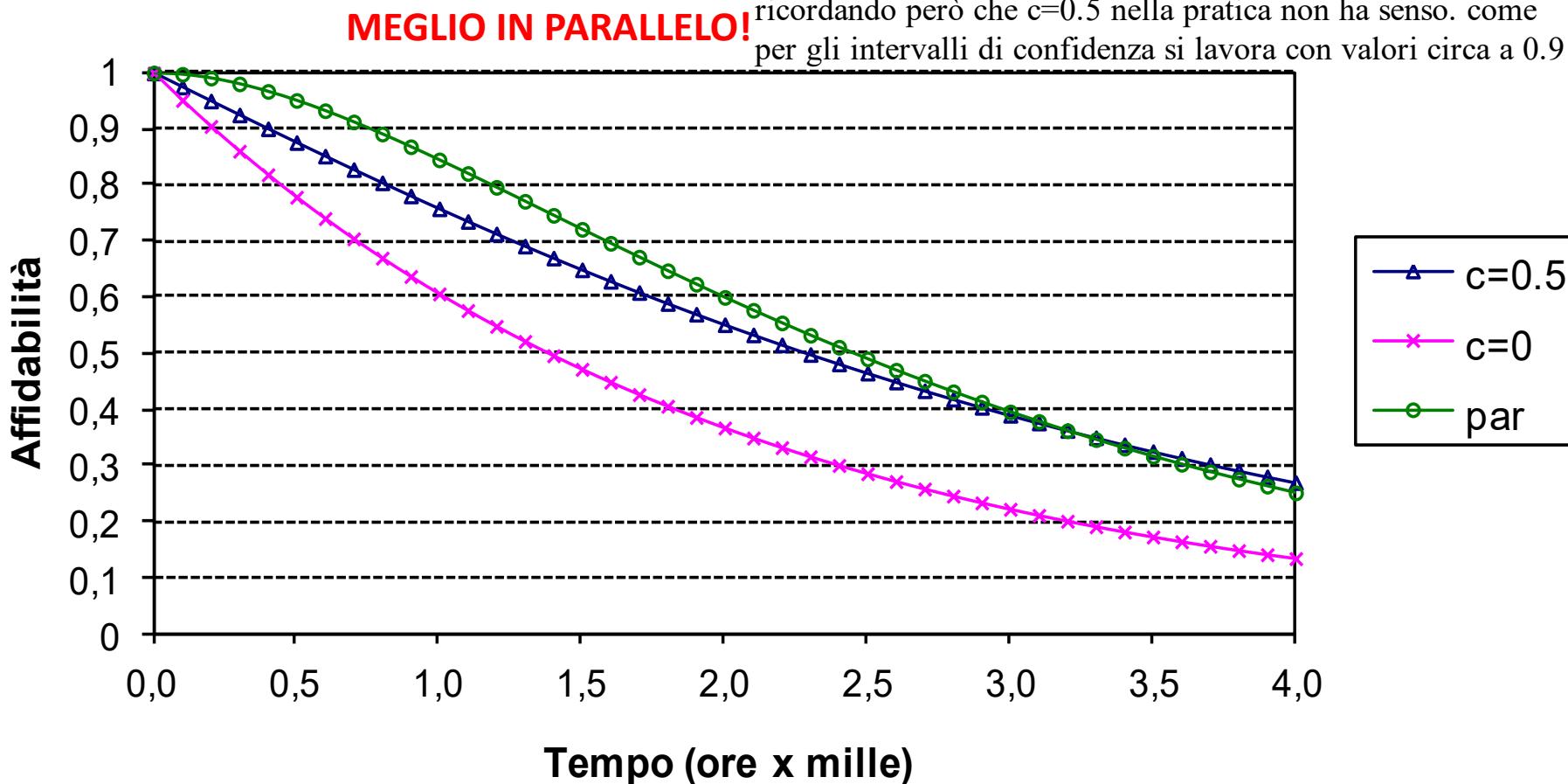
$$R_{PAR}(t) = \left(e^{-\lambda \cdot t} + e^{-\lambda \cdot t} - e^{-(\lambda + \lambda) \cdot t} \right)$$

$$R_{RIS}(t) = (1 + c\lambda t)e^{-\lambda t}, \quad t \geq 0$$

Secondo componente in parallelo o di riserva ?

(risultato con $c = 0,5$ e $1/\lambda = 2000$ ore)

nel caso $c=0.5$ si ha che il parallelo è meglio sfruttarlo
ricordando però che $c=0.5$ nella pratica non ha senso. come
per gli intervalli di confidenza si lavora con valori circa a 0.9



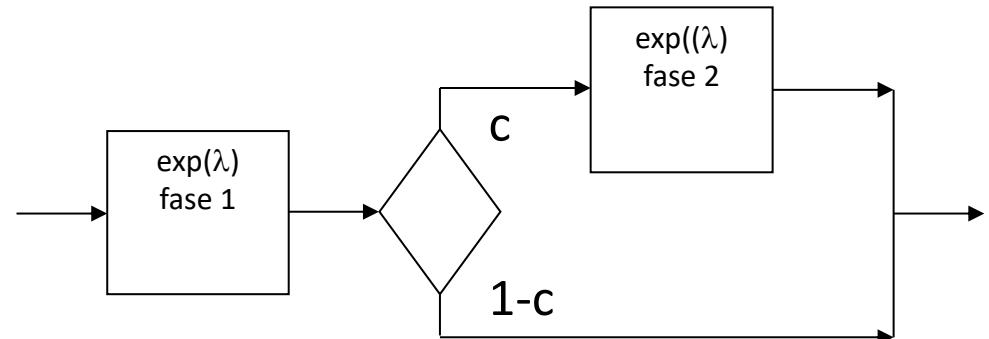
Esercizio: MTTF del sistema “1 + 1 di riserva”

$$MTTF = \int_{t=0}^{\infty} R(t)dt = \int_{t=0}^{\infty} (1 + c\lambda t)e^{-\lambda t} \cdot dt = \\ = \dots = \frac{1}{\lambda} (1 - c) + \left(\frac{1}{\lambda} + \frac{1}{\lambda} \right) \cdot c$$

tempo al guasto componente attivo
o tempo di vita del sistema se la
commutazione non riesce

tempo al guasto componente attivo
più tempo al guasto della riserva
con commutazione perfetta

Il risultato si può ottenere anche
applicando la formula del valore
atteso totale al seguente
diagramma a blocchi che illustra la
durata della vita del sistema:



Infatti, $(1/\lambda)$ è il valore atteso condizionato dall'evento “la commutazione non riesce” e $(1-c)$ è la probabilità dello evento condizionante; mentre $(2/\lambda)$ è il val atteso condizionato dall'evento “la commutazione riesce” e (c) è la probabilità che la commutazione riesca.

Esercizio: MTTF del sistema “2 in parallelo”

ricordando dalle statistiche dell'ordinamento che la distribuzione del massimo di due durate è pari al prodotto delle distribuzioni

$$X_{PAR} \doteq \max \{X_1, X_2\} \Rightarrow F_{X_{PAR}}(t) = F_{X_1}(t) \cdot F_{X_2}(t)$$

$$\begin{aligned} F_{X_{PAR}}(t) &= (1 - e^{-\lambda \cdot t})(1 - e^{-\lambda \cdot t}) \\ &= 1 - (e^{-\lambda \cdot t} + e^{-\lambda \cdot t} - e^{-(\lambda + \lambda) \cdot t}) \end{aligned}$$

ottengo la mortalità

$$\begin{aligned} E[X_{PAR}] &= \int_{t=0}^{\infty} (1 - F_{X_{PAR}}(t)) dt \\ &= \int_{t=0}^{\infty} (e^{-\lambda t} + e^{-\lambda t} - e^{-(\lambda t + \lambda t)}) dt \end{aligned}$$

affidabilità

valore atteso
del 1° componente

valore atteso
del 2° componente

valore atteso
del tempo al
primo guasto,
ma non so di chi

$$= \frac{1}{\lambda} + \frac{1}{\lambda} - \frac{1}{\lambda + \lambda}$$

Esercizio: Confronto fra i due MTTF

In definitiva, fissando $1/\lambda=2000\text{h}$ e $c=0.9$:

$$MTTF_{PAR} = \frac{1}{\lambda} + \frac{1}{\lambda} - \frac{1}{2\lambda} = (1+0.5) \cdot \frac{1}{\lambda} = 3000 \text{ h}$$

$$MTTF_{RIS} = \frac{1}{\lambda}(1-c) + \frac{2}{\lambda} c = (1+c) \cdot \frac{1}{\lambda} = 3800 \text{ h}$$

800 ore in più
circa 27%

Dunque:

finché risulta $c > 0.5 \Rightarrow MTTF_{RIS} > MTTF_{PAR}$

Esempio: faccio un giro con il motoscafo con 2 serbatoi. Basandomi sul consumo ad ore se vado piano passo più tempo in acqua altrimenti se accelero e vado a forti velocità finisco prima il carburante e passo meno tempo sul motoscafo.

Esempio informatico: ho da concludere un job, ed ho 2 processori a disposizione la cui velocità si può modulare, allora potrei o fare meno calcoli al secondo ma per più tempo oppure potrei spremere tutta la loro potenza di calcolo per calcolare più in fretta ma in un tempo inferiore. Ad esempio ho 2 GPU e voglio minare un bitcoin in 30 minuti, invece che in 4 ore perché poi arriva la bolletta della luce cara

IL PROBLEMA DEL RITARDO A LEZIONE

problema di sincronia

Il professore di “Processi Aleatori e Ingegneria del Teletraffico” arriva a lezione il venerdì mattina alla prima ora **con un ritardo aleatorio che va da 0 a 20 minuti massimo, ma arriva!**

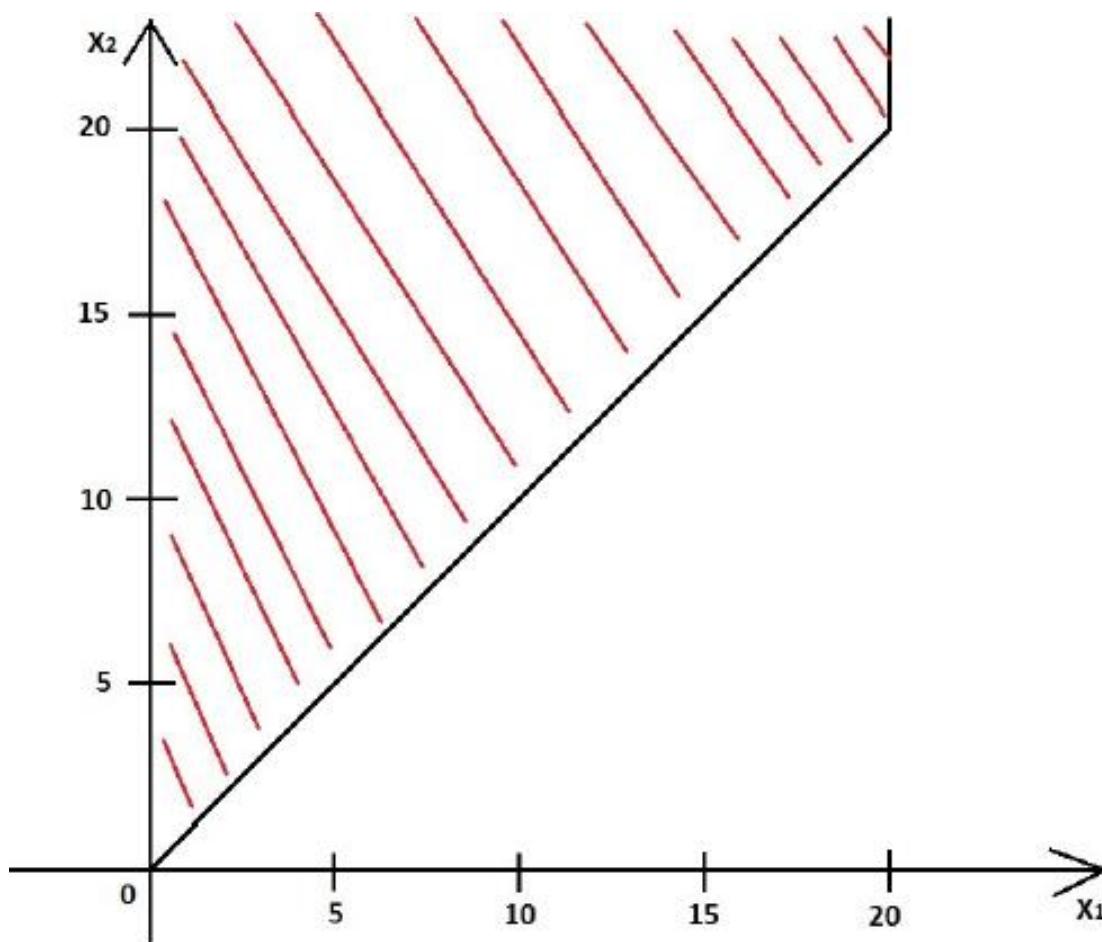
Lo studente di Arcavacata arriva invece con un **ritardo medio di 5 minuti** e distribuito secondo il modello esponenziale.
dunque non posso usare il modello di Weibull
uso l'esponenziale con lambda pari 5

Si chiede di:

- 1) Individuare l'area (A) che corrisponde all'evento di interesse così definito: “lo studente arriva dopo il professore”; X_1 sarà il docente X_2 sarà lo studente. Ci interessa $X_2 < X_1$
- 2) Adottando una funzione di densità lineare, ma compresa fra 0 e 20 minuti per il ritardo del professore, calcolare la probabilità dell'evento di interesse.

QUESITO 1:

Considerando che il docente (X_1) arriva con un ritardo espresso con una densità lineare e che lo studente (X_2) arriva con un ritardo **distribuito esponenzialmente** con, in media, 5 minuti di ritardo (ma potrebbe anche non arrivare!), si vede dalla seguente figura qual è l'area che corrisponde all'evento desiderato:



QUESITO 2:

I MODELLI DEL RITARDO PER DOCENTE E STUDENTE

Densità della variabile aleatoria X_1 "ritardo del docente":

$f_{X_1}(t) = k \cdot t$, con la costante k dalla seguente
"docente arriva sicuramente entro 20 minuti"

$$\int_{t=0}^{t=20} k \cdot t \cdot dt = 1 \quad \xrightarrow{\hspace{1cm}} \quad k \cdot \left[\frac{t^2}{2} \right]_{t=0}^{t=20} = 1 \quad \Rightarrow k = \frac{1}{200}$$

Densità della variabile aleatoria X_2 "ritardo dello studente":

$$E[X_2] = 5 \quad \Rightarrow f_{X_2}(t) = \frac{1}{5} \cdot e^{-\frac{t}{5}} \quad \text{"inversa" esponenziale}$$

QUESITO 2:

PROBABILITA' CHE LO STUDENTE ARRIVI DOPO IL DOCENTE

risolvo applicando la distribuzione totale

$$P(X_2 > X_1) =$$

$$= \int_{t=0}^{\infty} \lim_{dt \rightarrow 0} P\{X_2 > t \mid t < X_1 \leq t + dt\} \cdot P\{t < X_1 \leq t + dt\}$$

$$= \int_{t=0}^{\infty} [1 - F_{X_2|X_1}(t)] \cdot f_{X_1}^{(m)}(t) dt$$

ma professore e studente sono correlati o indipendenti?

$$\text{ipotesi di indipendenza} \Rightarrow = \int_{t=0}^{\infty} [1 - F_{X_2}(t)] \cdot f_{X_1}(t) dt$$

$$= \int_{t=0}^{20} e^{-\frac{t}{5}} \cdot \frac{t}{200} dt = \frac{1}{40} \cdot \int_{t=0}^{20} \frac{t}{5} e^{-\frac{t}{5}} dt$$

CALCOLO:

$$P(X_2 > X_1) = \frac{1}{40} \cdot \int_{t=0}^{t=20} \frac{t}{5} e^{-\frac{t}{5}} \cdot dt$$

Ponendo $x \triangleq t / 5 \Rightarrow dt = 5dx \Rightarrow \frac{5}{40} \int_{x=0}^{x=4} xe^{-x} dx$

e integrando per parti: $\int xe^{-x} dx = -\int d(xe^{-x}) + \int e^{-x} dx$

$$\begin{aligned} P(X_2 > X_1) &= \frac{1}{8} \cdot \left\{ \left[-xe^{-x} \right]_0^4 + \left[-e^{-x} \right]_0^4 \right\} = \\ &= \frac{1}{8} \cdot \left\{ -4e^{-4} - e^{-4} + 1 \right\} \\ &= \frac{1}{8} \cdot \left\{ -0,0732 - 0,0183 + 0,9084 \right\} = 0,113 \end{aligned}$$

E se il docente
ritarda al max
10minuti?

t=10 e k=1/100
insomma la metà

QUESITO 2: APPROCCIO ALTERNATIVO, integrando la congiunta

più immediato VALE PERCHÈ IPOTESI INDIPENDENZA

$$P(X_2 > X_1) = \int_A f_{X_1 X_2}(t_1, t_2) \cdot dt_1 dt_2$$

$$= \int_{t_1=0}^{20} \int_{t_2=t_1}^{\infty} f_{X_1 X_2}(t_1, t_2) \cdot dt_1 dt_2$$

$$A \triangleq \{0 \leq t_1 \leq 20 \cap t_1 \leq t_2 < \infty\}$$

$$f_{X_1, X_2}(t_1, t_2) = f_{X_1}(t_1) \cdot f_{X_2}(t_2)$$

RICORDANDO:

GRAZIE ALL'INDIPENDENZA
(lezione congiunta)

con: $f_{X_1}(t_1) = \frac{t_1}{200}$, $f_{X_2}(t_2) = \frac{1}{5} e^{-\frac{t_2}{5}}$

SI OTTIENE:

$$P(X_2 > X_1) = \int_{t_1=0}^{20} \int_{t_2=t_1}^{\infty} \frac{t_1}{200} \cdot \frac{e^{-\frac{t_2}{5}}}{5} \cdot dt_1 \cdot dt_2$$

CALCOLIAMO L'INTEGRALE INTERNO:

$$\int_{t_2=t_1}^{\infty} \frac{t_1}{200} \cdot \frac{e^{-\frac{t_2}{5}}}{5} \cdot dt_2 = \lim_{t \rightarrow \infty} \int_{t_2=t_1}^t \frac{t_1}{200} \cdot \frac{e^{-\frac{t_2}{5}}}{5} \cdot dt_2$$

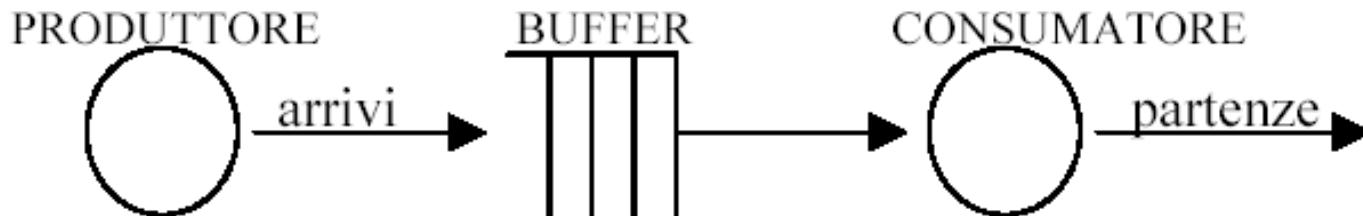
$$= \frac{t_1}{200} \cdot \lim_{t \rightarrow \infty} \left[\left(1 - e^{-\frac{t}{5}} \right) - \left(1 - e^{-\frac{t_1}{5}} \right) \right]$$

$$= \frac{t_1}{200} \cdot e^{-\frac{t_1}{5}} = \frac{1}{40} \cdot t_1 \cdot \frac{1}{5} e^{-\frac{t_1}{5}}$$

E RISCRIVIAMO: $P(X_2 > X_1) = \frac{1}{40} \int_{t_1=0}^{20} \frac{t_1}{5} e^{-\frac{t_1}{5}} \cdot dt_1$

Il problema della distribuzione del tempo di attesa

Ipotesi: il server non viene interrotto da nessun altro lavoro e non rifiuta richieste



Ipotesi: clienti infiniti e che sono caratterizzati solo dal ritmo di arrivo λ

$$F_W^{(m)}(t) \triangleq \Pr[W \leq t]$$

$$= \sum_{n=0}^{\infty} \Pr[W \leq t \cap N = n]$$

è una congiunta. Attendo meno di t e ci siano n utenti presenti

$$= \sum_{n=0}^{\infty} \Pr[W \leq t \mid N = n] \cdot \Pr[N = n]$$

l'attesa W è una marginale condizionata dal numero di utenti presenti N

$$= \sum_{n=0}^{\infty} F_{W|N}(t \mid n) \cdot \Pr[N = n]$$

distribuzione totale

Trovo n clienti in coda, qual è la probabilità che vengo servito entro un tempo?

Se arrivi e partenze sono esponenziali $F_{W|N}(t|n)$ è modello di Erlang

Non vado a studiare il comportamento di $W(m)$, ad esempio $W(10)$ ovvero del decimo consumatore. Ma eseguo uno studio a lungo termine

Var. al. sul lungo termine:
quindi ne studio una sola
 $W \triangleq \text{"attesa nel buffer"}$

$N \triangleq \text{"utenti presenti"}$
sul lungo termine,
in corrispondenza
di istanti
completamente
casuali

C'è differenza tra utenti trovati ed utenti presenti. Il numero di utenti trovati dipendono dal profilo probabilistico che si immagina di avere, estraendo il mio istante di arrivo. Quindi a t_1 trovo n utenti mentre a t_2 trovo m utenti. Gli utenti presenti non dipendono dall'arrivo ma da un osservatore casuale. Se considero uno stimatore casuale gli istanti di arrivo sarebbero distribuiti in maniera equiprobabile. Nel problema della distribuzione del tempo di attesa considero gli arrivi in maniera completamente casuale, (Poissoniani: distribuiti uniformemente nel orizzonte temporale) perciò il punto di vista dell'osservatore casuale coincide con il punto di vista del singolo utente, ovvero statisticamente indistinguibili. Se non fosse dovrei impiegare una simulazione monte carlo

Il teorema (sulla distribuzione del tempo di attesa)

μ velocità di servizio ed anche produttività limite, quando il server è pieno

IPOTESI: λ ritmo degli arrivi

$\lambda = \rho\mu$ equazione di continuità, ciò che entra esce

o meglio disciplina first in first served (F.I.F.S.)

Il modello P-C ha un solo server gestito con disciplina “F.I.F.O.”,

“ $1/\lambda > 0$ ” è il valore atteso del tempo tra due arrivi consecutivi, A,

“ $1/\mu > 0$ ” è la durata attesa del servizio reso al generico cliente, S,

probabilità che arrivi un nuovo cliente entro t

$$\{A : F_A(t) = 1 - e^{\lambda t}, t \geq 0\}$$

probabilità che il servizio abbia durata meno di t

$$\{S : F_S(t) = 1 - e^{\mu t}, t \geq 0\}$$

Altra IPOTESI

TESI: è valida ses: $1 > (\lambda / \mu) \triangleq \rho$

$$1. P_N(n) = (1 - \rho) \cdot \rho^n, \quad n \geq 0$$

$$2. F_W(t) = F_{W|N=0}(t)P_{N=0} + \sum_{n=1}^{\infty} F_{W|N=n}(t)P_N(n)$$

distribuzione tempo di attesa

VALIDO SOLO PER F.I.F.S.

$$= 1 \cdot (1 - \rho) + (1 - e^{-\mu(1-\rho)t}) \cdot \rho = 1 - \rho e^{-\mu(1-\rho)t}$$

$$\left(\begin{array}{l} \rho = \Pr \left\{ \text{"utenti almeno uno} \geq 1 \right\} \\ = \Pr \left\{ \text{"attesa} > 0 \right\} \end{array} \right)$$

ρ è prob che ci sia almeno 1 è uguale a quella che il server stia lavorando

Si dimostra che si ottiene il modello geometrico

$$P_N(n) = (1 - \rho) \cdot \rho^n \quad n \geq 0 \quad \text{se } n=0 \text{ non c'è nessuno}$$

probabilità
server è libero/buffer vuoto

probabilità di occupazione del buffer con n clienti

E che il modello probabilistico è il seguente

$$F_W(t) = F_{W|N=0}(t)P_{N=0} + \sum_{n=1}^{\infty} F_{W|N=n}(t)P_N(n)$$

Probabilità evento condizionante
che non ci sia nessuno

Distribuzione
dell'attesa (totale)

Distribuzione
condizionata
che non ci sia nessuno

$$F_W(t) = 1 \cdot (1 - \rho) + (1 - e^{-\mu(1-\rho)t}) \cdot \rho = 1 - \rho e^{-\mu(1-\rho)t}$$

attesa condizionata
di chi non aspetta

Erlang

se non c'è nessuno
l'unica realizzazione è pari
a 0. dunque tutta la
prob. è concentrata

probabilità
che non ci
sia nessuno

attesa condizionante
di chi aspetta

attesa in generale
che NON è una
funzione
esponenziale

Il modello P_C e l'attesa dal medico di base

(vale SEMPRE l'ipotesi di arrivi casuali)

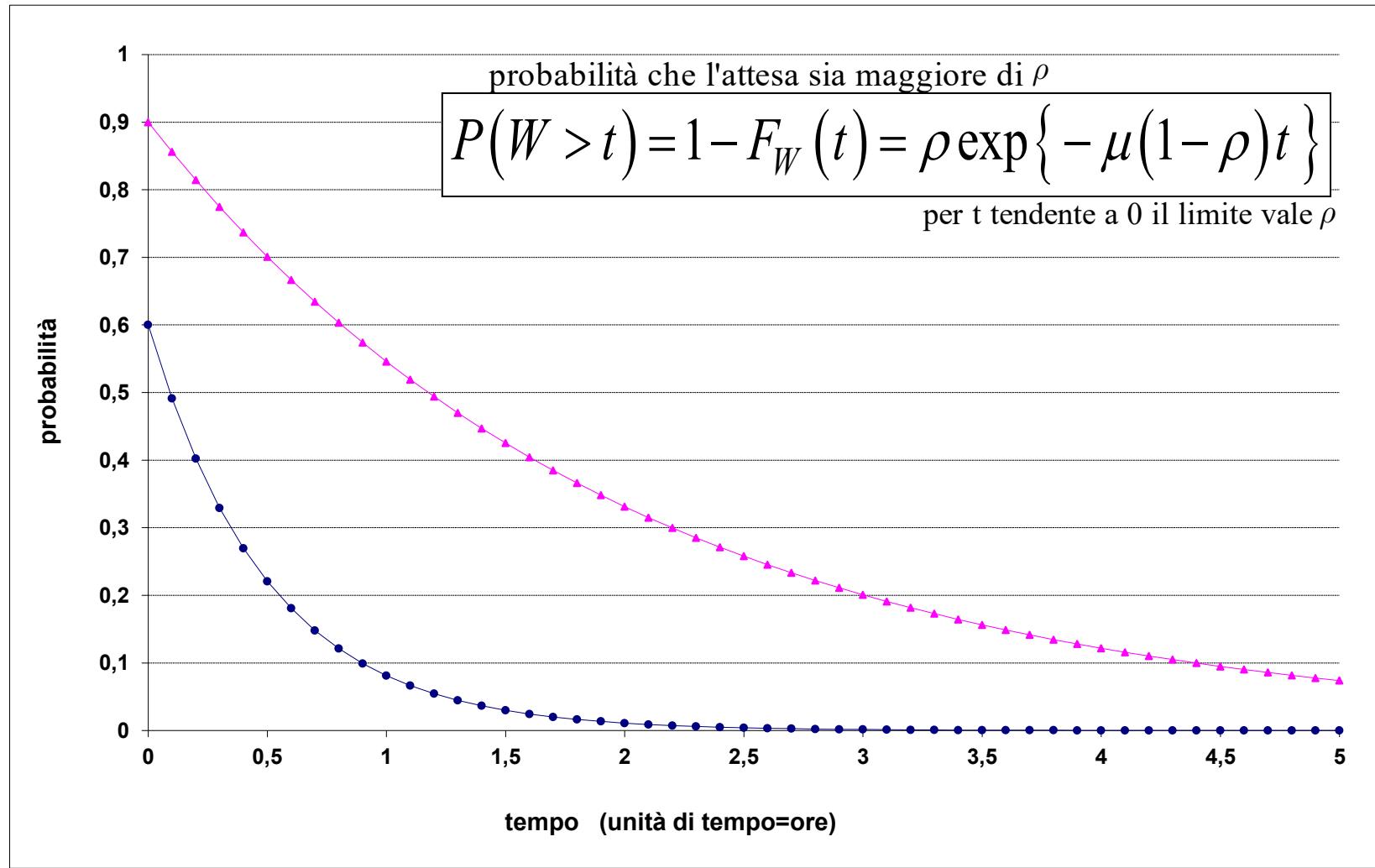
Il dott. Tizio, medico di base, valuta che il flusso degli arrivi di pazienti al suo studio possa essere quantificato in **3 pazienti all'ora**, in media su tutto l'intervallo giornaliero di apertura dell'ambulatorio. L'ambulatorio si compone di una sala d'attesa e dello studio per le visite.

$$\mu = 12 * 5 = 60$$

I pazienti vengono visitati in ordine di arrivo e ogni visita dura mediamente **12 minuti**. Verificare che è praticamente nulla la probabilità che un paziente debba aspettare più di 2 ore prima di essere visitato e che, d'altra parte, il 78% dei pazienti in arrivo aspetta non più di mezz'ora.

Considerato che il dott. Tizio non andrà in ferie d'estate, valutare se può accettare nel proprio ambulatorio anche parte dei pazienti del dott. Caio che, invece, andrà in ferie. Valutando che si avrebbe un incremento del ritmo orario medio degli arrivi da $3h^{-1}$ a ^{incremento 50%} $4,5h^{-1}$ nell'ambulatorio del dott. Tizio, verificare che la probabilità di aspettare più di 2 ore prima di essere visitati aumenterebbe da quasi zero al valore 0,33.

Probabilità di attesa dal medico di base



La gara fra due durate esponenziali indipendenti

IPOTESI

X_1 e X_2 sono le durate di due attività in parallelo svolgimento: (hanno cominciato assieme)

$$P(X_1 < X_2) = \Pr(X_2 > t | t < X_1 \leq t + dt)$$

distribuzione totale = $\int_{t=0}^{\infty} \left[1 - F_{X_2|X_1}(t) \right] \cdot f_{X_1}^{(m)}(t) dt$

probabilità condizionata
sopravvivenza

probabilità variabile condizionante che duri tra t e $t+dt$

se sono indipendenti

$$F_{X_2|X_1}(t) = F_{X_2}(t)$$

$$f_{X_1}^{(m)} = \frac{f_{X_1, X_2}}{f_{X_2/X_1}} = f_{X_1}$$

(INDIPENDENZA STOCASTICA)

caso più semplice

Per due modelli ESPONENZIALI con parametri μ_1 e μ_2 :

$$P(X_1 < X_2) = \int_{t=0}^{\infty} e^{-\mu_2 t} \cdot \mu_1 e^{-\mu_1 t} dt$$

$$= \frac{\mu_1}{\mu_1 + \mu_2} \int_{t=0}^{\infty} (\mu_1 + \mu_2) e^{-(\mu_2 + \mu_1)t} dt = \frac{\mu_1}{\mu_1 + \mu_2}$$

come visto in precedenza

velocità del primo modello
fratto la somma delle due velocità

La durata della gara e il vincitore

Siano
 X_1, X_2 esponenziali e indipendenti

$$\Rightarrow P(X_1 < X_2) = \frac{\mu_1}{\mu_1 + \mu_2} \quad \text{estensione possibile a 4 partecipanti: } X_1 < X_2 \cap X_3 & X_4$$

Se aggiungiamo che: $Z \triangleq \min(X_1, X_2)$ statistica dell'ordinamento

probabilità che X_1 finisca entro t

$$\Rightarrow F_Z(t) = 1 - (1 - F_{X_1}(t)) \cdot (1 - F_{X_2}(t)) = 1 - e^{-(\mu_1 + \mu_2)t}$$

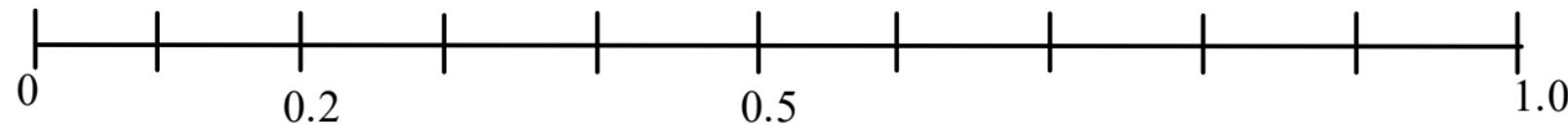
probabilità che X_1 finisca dopo t probabilità che X_2 finisca dopo t

$$\begin{aligned} P(Z \leq t \cap X_1 < X_2) &= \\ &= P(Z \leq t \mid X_1 < X_2) \cdot P(X_1 < X_2) \\ &\quad \text{condizionata} \qquad \text{condizionante} \\ ? \text{ qui si gioca l'indipendenza} \\ &= P(Z \leq t) \cdot P(X_1 < X_2) \\ &= (1 - e^{-(\mu_1 + \mu_2)t}) \cdot \frac{\mu_1}{\mu_1 + \mu_2} \end{aligned}$$

Ecco la
probabilità che il
primo fenomeno
finisca entro t e
che sia proprio il
fenomeno 1

Esperimento di Bernoulli generalizzato per il caso generico di n esp in gara

Genero un numero casuale ed in base al suo valore vedo in quale intervallo cade. Esempio:



Se casca in 0.2 allora ho che la prima attività a finire è $\frac{\mu_1 = 0.2}{\sum_{j=1}^n \mu_j}$

Se casca in 0.3 il primo a finire è $\frac{\mu_2 = 0.3}{\sum_{j=1}^n \mu_j}$

La distribuzione iperesponenziale

non è una convoluzione

$$Y \doteq \sum_{i=1}^n \alpha_i \cdot Y_i ;$$

ma combinazione convessa di exp

$$\alpha_i > 0, \quad \sum_{i=1}^n \alpha_i = 1;$$

$$F_{Y_i}(y) \doteq 1 - \exp\{-\lambda_i y\}$$

se exp tutte uguali non si
cambia niente

**COMBINAZIONE
CONVESSA DI
EXPONENZIALI**

Interpretazione:

$$P_X(i) \doteq \alpha_i, \quad i = 1, \dots, n$$

prob. eventi condizionanti

Distribuzione totale:

$$F_Y(y) = \sum_{i=1}^n F_{Y|X=i}(y | i) \cdot P_X(i)$$

condizionata della durata

condizionante

Risultato:

$$F_Y(y) = \sum_{i=1}^n \alpha_i (1 - \exp\{-\lambda_i y\})$$

Dunque l'iperesponenziale nasce dalla distribuzione totale

$$f_Y(y) = \sum_{i=1}^n \alpha_i \lambda_i \exp\{-\lambda_i y\}$$

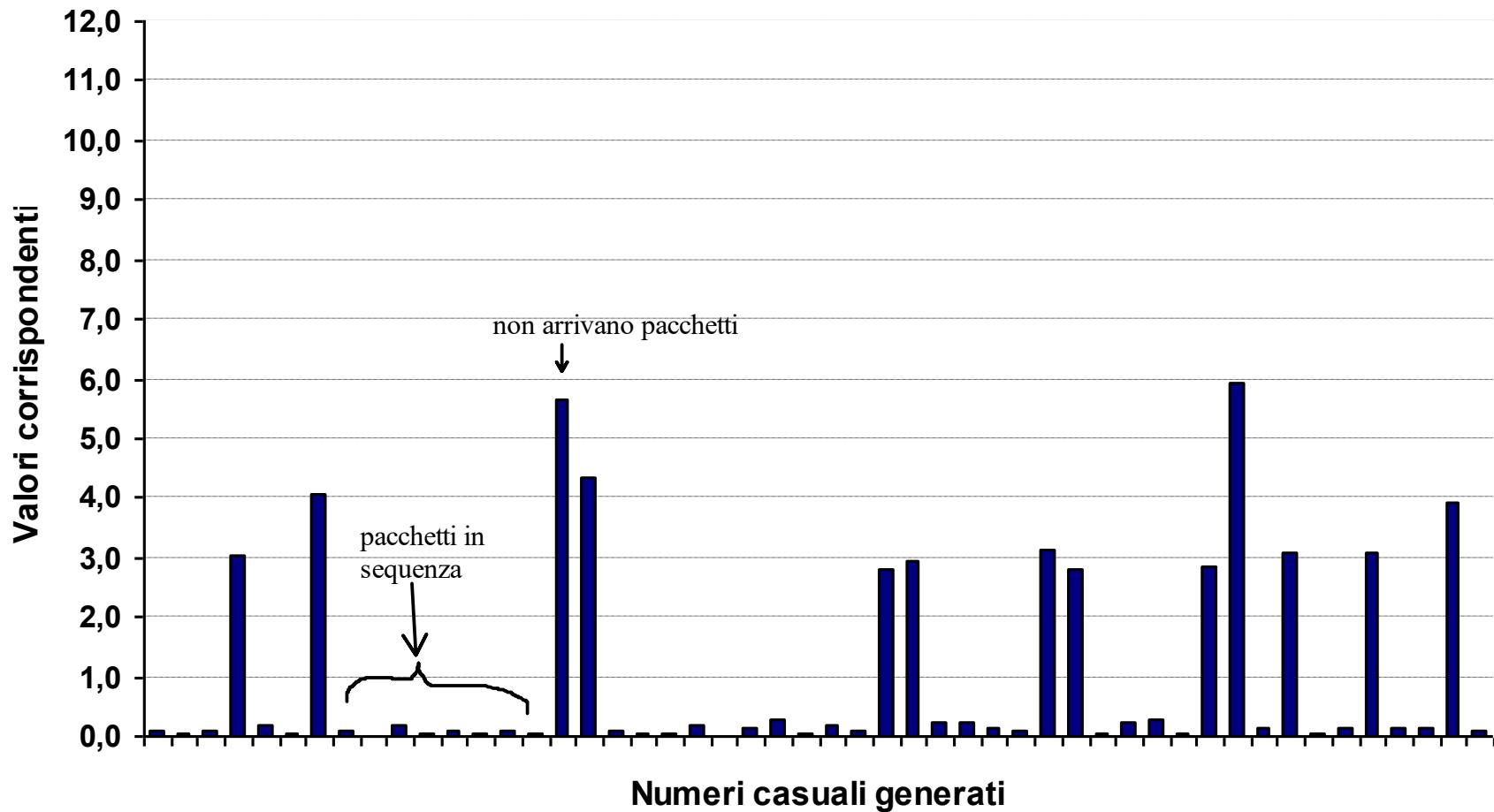
Quindi la densità:

Ricordando che l'iperexp mostra l'effetto di burst nelle reti oppure di esecuzione di processi in un OS

Realizzazioni casuali dalla legge IPEResponenziale

(alfa = 0,75 - lambda_1 = 5 - lambda_2 = 0,5) tra i lambda ho un fattore 10

in questo caso il burst non è molto chiaro



... **attesa media:**

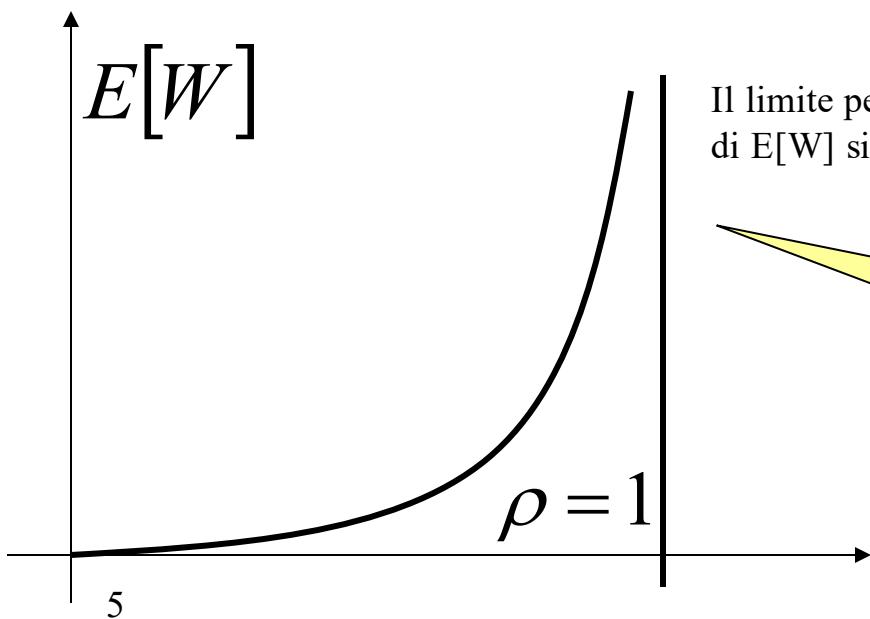
$$E[W] = \int_{t=0}^{\infty} (1 - F_W(t)) dt$$

$$= \int_{t=0}^{\infty} \rho e^{-\mu(1-\rho)t} dt = \frac{\rho}{\mu(1-\rho)}$$

$$E[D] = E[W] + \mu^{-1} = \frac{1}{\mu(1-\rho)}$$

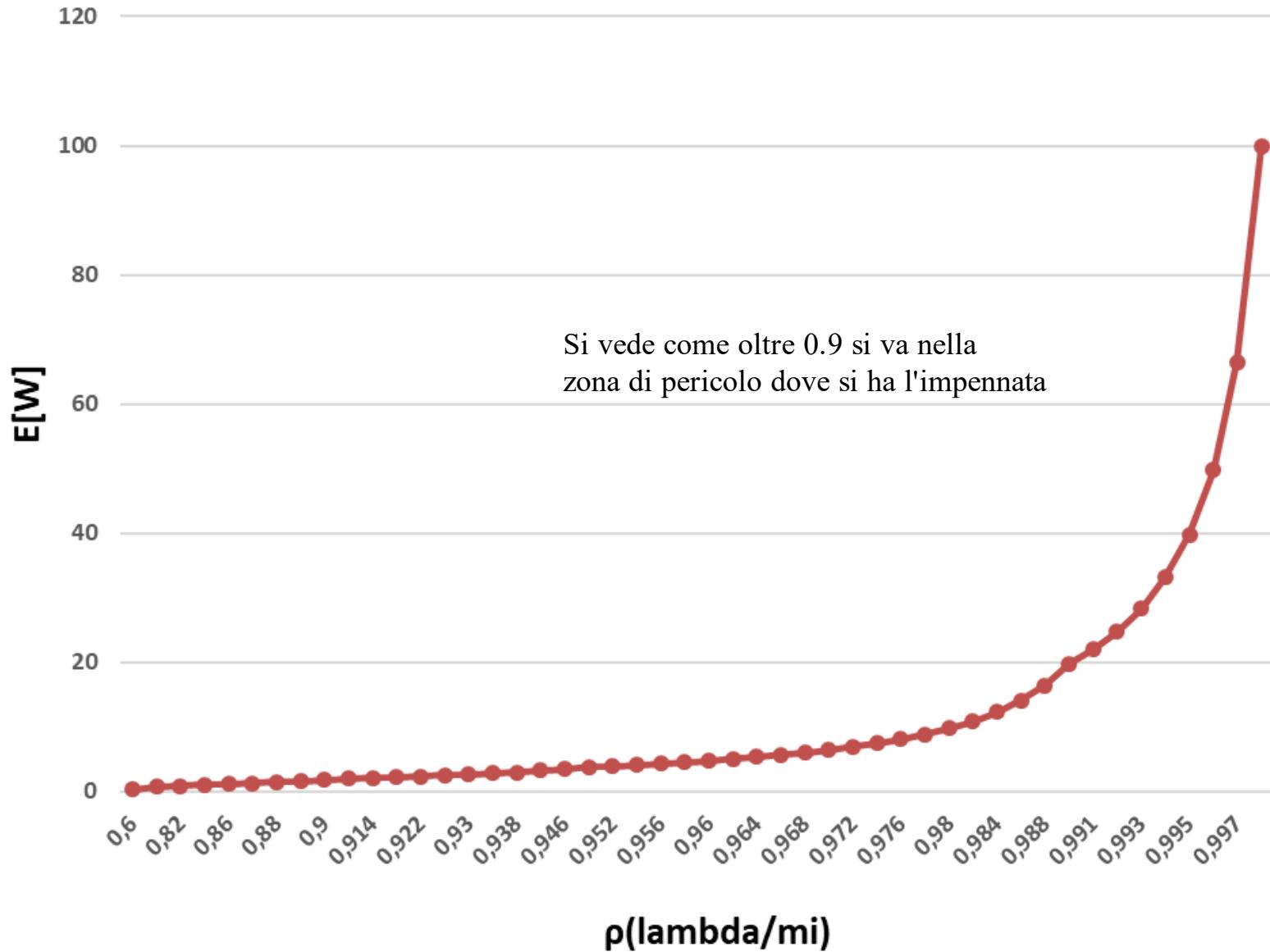
Tempo (Delay)
speso nel
modello P-C

Il limite per rho tendente a 1
di E[W] si ha un asintoto quindi si attende infinitamente



L'utente trova una coda
infinitamente lunga

ATTESA MEDIA ALL'AUMENTARE DI RHO



Il modello P_C e l'attesa dal medico di base

Il risultato appena verificato non è accettabile per la qualità del servizio che il dott. Tizio vuole garantire a tutti i pazienti, alcuni dei quali sarebbero pure tentati di abbandonare la coda di attesa in ambulatorio. Allora, egli decide di adottare una politica di "rigetto senza ritorno" nei confronti di pazienti in arrivo.

La politica consiste nel chiudere la porta dell'ambulatorio quando si raggiunga il numero di 3 pazienti in attesa oltre quello sotto visita e dirottare presso la guardia medica o un qualsivoglia altro collega chiunque trovi la porta chiusa quando arriva in ambulatorio.

Il modello P_C e l'attesa dal medico di base

Adottando il modello geometrico troncato per la variabile aleatoria “numero di pazienti presenti in ambulatorio” ricalcolare la probabilità che in presenza della condivisione di pazienti (e dunque ritmo medio orario degli arrivi pari a 4,5) un paziente in arrivo non aspetti più di mezz’ora, posto che venga accettato.

Domanda finale riferita al modello a buffer illimitato:

E se volessimo calcolare la probabilità che il tempo totale di permanenza dal dottore (attesa + visita) risulti non superiore alle 2 ore, quando il ritmo degli arrivi sale a 4,5h-1?

Risposta: avremmo bisogno della funzione di distribuzione riferita al tempo totale (attesa + visita), che si dimostra essere la seguente:

$$P(W+V \leq t) \hat{=} F_{W+V}(t) = 1 - \exp \left\{ -\mu(1-\rho)t \right\}$$

La troncata a partire dalla geometrica NON troncata

Dopo un tot di clienti si ha un rigetto da parte del serve. È un richiamo sulla geometrica

$$P_n = (1 - \rho)\rho^n, \quad n = 0, 1, \dots, k, k+1, \dots$$

Evento di interesse:

"esattamente n utenti presenti | al più k possono essere accettati" $\hat{=}$ $n | n \leq k$

Probabilità
evento
condizionante:

$$P_{n \leq k} = \sum_{n=0}^k (1 - \rho)\rho^n$$

Probabilità
evento
d'interesse:

$$P_{n|n \leq k} = \frac{P_n}{P_{n \leq k}} = \frac{(1 - \rho)\rho^n}{\sum_{n=0}^k (1 - \rho)\rho^n} \quad n = 0, 1, \dots, k$$

Derivazione formula finale per

$$P_{n|n \leq k}$$

$$n = 0, 1, \dots, k$$

Calcoli per capire non per esame

Si sviluppa:

$$\sum_{n=0}^k \rho^n = \sum_{i=0}^{\infty} \rho^n - \sum_{n=k+1}^{\infty} \rho^n$$

$$\text{con } \rho < 1 \text{ e } j \hat{=} n - (k+1) \Rightarrow \quad = \frac{1}{1-\rho} - \sum_{j=0}^{\infty} \rho^{k+1} \rho^j$$

$$= \frac{1}{1-\rho} - \frac{\rho^{k+1}}{1-\rho} = \frac{1-\rho^{k+1}}{1-\rho}$$

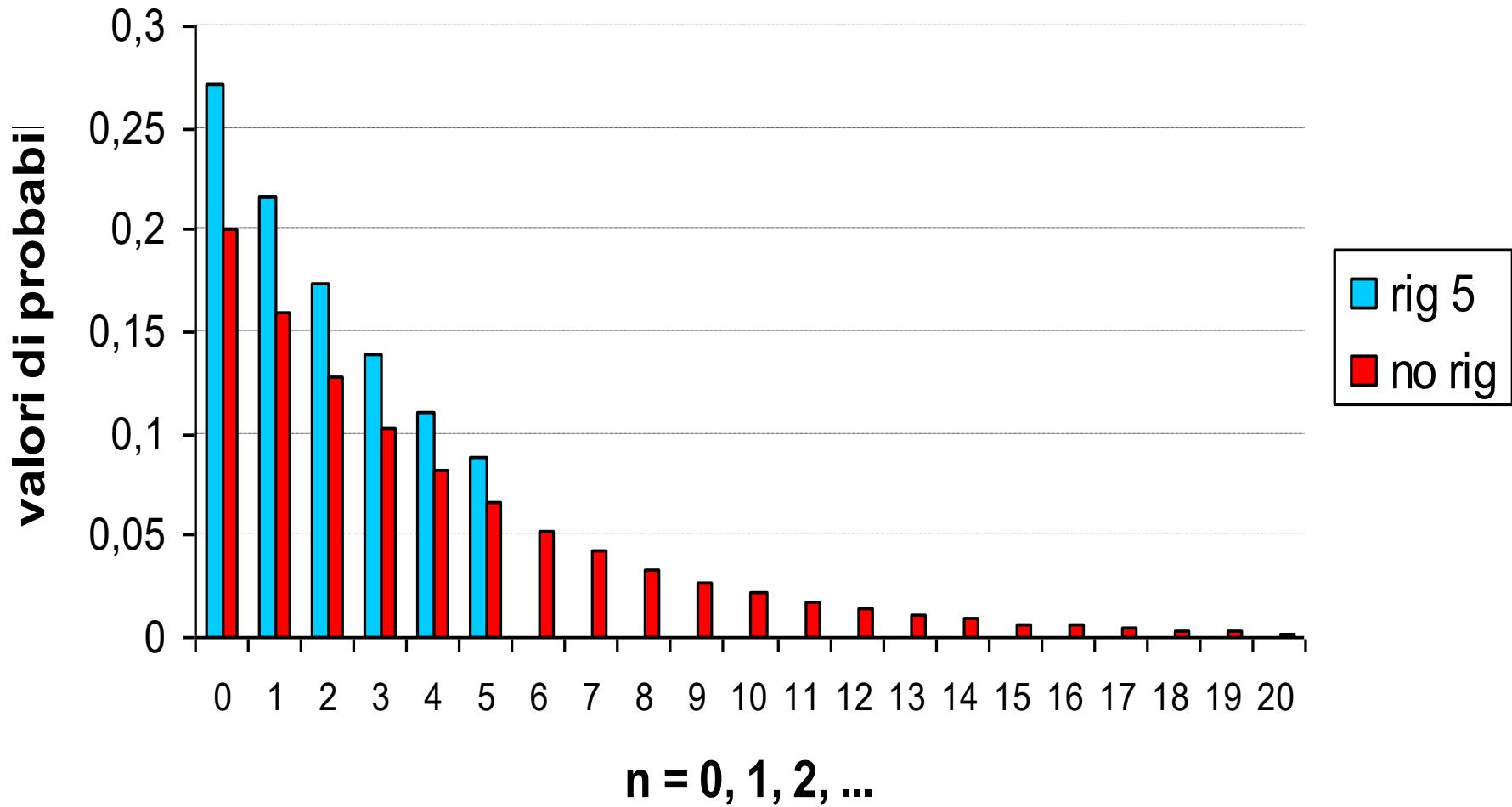
Formula finale:

$$P_{n|n \leq k} = \frac{\rho^n}{\sum_{n=0}^k \rho^n} = \frac{(1-\rho)}{1-\rho^{k+1}} \rho^n$$

$$n = 0, 1, \dots, k$$

Prob stazionaria (utenti presenti = n), con $\rho = 0.8$

Di più sul foglio EXCEL



Formula del valore atteso totale (per var. al. continue)

Si parte dalla definizione di valore atteso della Y:

$$E[Y] \triangleq \int_{y=0}^{\infty} y \cdot f_Y^{(m)}(y) \cdot dy = \text{passaggio con la congiunta}$$

Estensione del concetto per variabile condizionata

e si esprime la densità marginale della Y come prodotto della densità condizionata di $Y|X$ per la densità marginale della condizionante, X:

$$= \int_{y=0}^{\infty} y \cdot \left[\int_{x=0}^{\infty} f(y|x) f_X^{(m)}(x) dx \right] dy = \int_{x=0}^{\infty} \left[\int_{y=0}^{\infty} y \cdot f(y|x) dy \right] f_X^{(m)}(x) dx$$

sommo tutti i casi possibili della variabile condizionante

A questo punto, si definisce il valore atteso condizionato:

e si ottiene la seguente:

$$E[Y|X=x] = \int_{y=0}^{\infty} y \cdot f(y|x) dy < +\infty$$

fisso il valore condizionante a x ad esempio impiego 10 secondi per un processo al variare di x (valore condizionante) si ha una funzione, una curva detta **curva di regressione**

$$E[Y] = \int_{x=0}^{\infty} E[Y|X=x] \cdot f_X^{(m)}(x) \cdot dx$$

valore atteso condizionato per la condizionante sommato su tutti i casi possibili quindi ottengo il valore atteso totale

Nota come formula del valore atteso totale

Formula del valore atteso totale nel modello produttore consumatore

$S \hat{=} "tempo di consumo"$ (non si sa la durata)
o meglio servizio

$N_S \hat{=} "\# di produzioni durante un consumo"$
arrivi servizio

Qual è il numero atteso di arrivi durante un servizio?

Val. att. condizionato:
della durata del servizio

$$E[N_S | t] = \lambda \cdot t$$

pongo un vincolo
con un valore fisso

significa $S=t$

→ nel protocollo Aloha si ha che
 $G = \lambda T$

Val. att. totale

$$E[N_S] \hat{=} \int_{t=0}^{\infty} E[N_S | S = t] \cdot \Pr\{S \in [t, t + dt]\} dt$$

$$= \int_{t=0}^{\infty} \lambda t \cdot \mu e^{-\mu t} dt = \frac{\lambda}{\mu} \hat{=} \rho$$

Num. medio di
arrivi durante
servizio
un consumo

Ricapitolando quanto detto nella lezione precedente abbiamo:

λ : ritmo degli arrivi

ρ : numero medio di arrivi durante un servizio e probabilità che ci sia almeno un cliente da servire

μ : capacità di servizio del server e produttività massima del server, nel caso in cui, ci sia sempre almeno un cliente che si sta servendo e che arrivino clienti con un ritmo di arrivo imposto, anche se questo porterebbe ad avere code infinitamente lunghe sul lungo termine

$$\rho = \frac{\lambda}{\mu} \quad \text{per definizione} \quad \rho \cdot \mu \quad \text{il prodotto è il ritmo delle uscite, il cosiddetto throughput del sistema}$$

Un sistema in cui $\frac{\lambda}{\mu} > 1$ non è ammissibile!

VALORE ATTESO E VARIANZA DI FUNZIONI DI V.A.

Siano X e Y due var. al. continue e non negative $(\Omega \triangleq \{(x, y) : 0 \leq x < \infty \cap 0 \leq y < \infty\})$
(perchè ci occupiamo di durate) lo spazio di lavoro è quello delle congiunte!

Sia $g(X, Y)$ una delle 4 operazioni aritmetiche che configurano i punti (x, y) di un esperimento aleatorio congiunto di X and Y in un sottoinsieme di R^2 sul quale è possibile integrare la densità congiunta di X e Y.

Allora:

$$E[g(X, Y)] \triangleq \int_{x=0}^{\infty} \int_{y=0}^{\infty} g(x, y) \cdot f_{X,Y}(x, y) dx dy < +\infty \quad \text{ovviamente se non converge non esiste}$$

e' il valore atteso della coppia di var al X e Y sotto l'operazione "g"

$$Var[g(X, Y)] \triangleq \int_{x=0}^{\infty} \int_{y=0}^{\infty} (g(x, y) - E[g(X, Y)])^2 \cdot f_{X,Y}(x, y) dx dy$$

è la varianza della coppia di var al X e Y sotto l'operazione "g"

FORMULE UTILI PER VALORE ATTESO E VARIANZA

$$E[X + Y] = E[X] + E[Y]$$

senza indipendenza!

(additività) che finalmente dimostriamo

$$E[cX + d] = cE[X] + d$$

(c e d arbitrarie)

con l'ipotesi di indipendenza: come anticipato ad inizio corso

$$E[XY] = E[X]E[Y]$$

$E[E[X]]$ =valore atteso di
un numero! $E[X]$ è un numero!

(forma prodotto)

$$VAR[X] \doteq E[(X - E[X])^2]$$

$$= E[X^2 - 2XE[X] + E[X]^2]$$

non dimostrabile senza
aver dimostrato prima
l'additività del val.att.

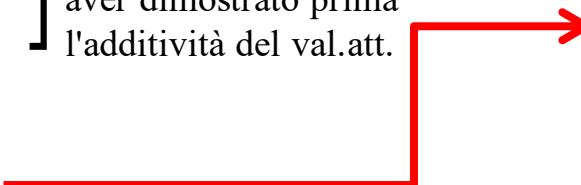
$$= E[X^2] - E[X]^2$$

vista con la Gaussiana

$$VAR[a \cdot X]$$

$$= a^2 VAR[X]$$

$$a \in \mathbb{R}, a \neq 0$$



Additività del valore atteso - prova

Per la dimostrazione serviva la congiunta e marginale! Ecco perchè non l'abbiamo dimostrata prima

$$E[X + Y] \hat{=} \int_{y=0}^{\infty} \int_{x=0}^{\infty} (x + y) f(x, y) dx \cdot dy$$

$$= \int_{x=0}^{\infty} x \int_{y=0}^{\infty} f(x, y) dy \cdot dx + \int_{y=0}^{\infty} y \int_{x=0}^{\infty} f(x, y) dx \cdot dy$$

la trasformiamo nella
marginale di X la trasformiamo nella
marginale di Y

$$= \int_{x=0}^{\infty} x \cdot f_X^{(m)}(x) \cdot dx + \int_{y=0}^{\infty} y \cdot f_Y^{(m)}(y) \cdot dy$$

come per definizione

$$= E[X] + E[Y]$$

La differenza tra valori attesi si è vista con Laplace (problema di sincronia)

**Senza
ipotesi di
indipendenza!**

altrimenti "uccido" il ragionamento

$$\int_{x=0}^{\infty} f(x, y) dx dy = f_Y^{(m)}(y) dy$$

da congiunta a marginale

Forma prodotto del valore atteso - prova

$$E[X \cdot Y] \hat{=} \int_{y=0}^{\infty} \int_{x=0}^{\infty} x \cdot y \cdot f(x, y) dx \cdot dy$$

Con ipotesi di
indipendenza!

$$\begin{aligned} & \text{probabilità congiunta} & \text{probabilità} & \text{probabilità} \\ f(x, y) dx dy & \hat{=} f_X(x) dx \cdot f_Y(y) dy \end{aligned}$$

valido perché la probabilità congiunta è pari al prodotto delle singole probabilità solo in presenza di ipotesi di indipendenza! Come visto nella lezione 20

$$= \int_{x=0}^{\infty} x \cdot f(x) dx \int_{y=0}^{\infty} y \cdot f(y) \cdot dy$$

$$= E[X] \cdot E[Y]$$

Il quoziente è difficile da dimostrare e viene utilizzato nel modello di Cauchy

Riprendiamo la definizione di varianza per singola variabile aleatoria

PER DEFINIZIONE DI VARIANZA:

$$\text{VAR}[X] \triangleq E[(X - E[X])^2]$$

Allora:

$$\text{VAR}[\alpha X] = E[(\alpha X - E[\alpha X])^2]$$

$$= E[(\alpha X - \alpha E[X])^2]$$

$$= E[\alpha^2 \cdot (X - E[X])^2]$$

$$= \alpha^2 E[(X - E[X])^2]$$

$$= \alpha^2 \text{VAR}[X]$$

VARIANZA DELLA SOMMA DI DUE VARIABILI ALEATORIE

$$VAR[X + Y] \doteq E[(X + Y) - E[X + Y]]^2 \quad \text{grazie all'additività spezzo } E[X+Y]$$

$$= E[((X + Y) - E[X] - E[Y])^2] = E[((X - E[X]) + (Y - E[Y]))^2] \quad \begin{matrix} \text{semplici passaggi algebrici} \\ \text{mi avvicino alla varianza} \end{matrix}$$

sviluppo il quadrato e mi rendo conto di avere 2 varianze

$$= E[(X - E[X])^2 + (Y - E[Y])^2 + 2(X - E[X])(Y - E[Y])]$$

$$= E[(X - E[X])^2] + E[(Y - E[Y])^2] + 2E[(X - E[X])(Y - E[Y])] \quad \begin{matrix} \text{raggruppo e vedo le 2 varianze} \\ \text{mi avvicino alla varianza} \end{matrix}$$

$$= VAR[X] + VAR[Y] + 2E[(X - E[X])(Y - E[Y])]$$

covarianza che permetterà di studiare in qualche misura la dipendenza tra variabili aleatorie intesa come correlazione

$$\Rightarrow VAR[X + Y] = VAR[X] + VAR[Y] + 2COV[XY]$$

si capisce anche che la varianza non è additiva

$$\text{e quindi: } VAR[X - Y] = VAR[X] + VAR[Y] - 2COV[XY]$$

da notare che cambia solo nel segno della covarianza

Non è domanda di esame

GENERALIZZAZIONE DELLA SOMMA

usata nella lezioni degli intervalli
di confidenza

NON SONO indipendenti le v.a., se lo fossero allora cadrei in una somma di varianze

$$VAR[X_1 + X_2 + \dots + X_n] = (a+b+c+d)^2 = a^2 + b^2 + c^2 + d^2 + 2ab + 2ac + 2ad + 2bc + 2bd + 2cd$$

di nuovo non serve l'indipendenza!

$$\begin{aligned} &= E\left[((X_1 + X_2 + \dots + X_n) - E[X_1 + X_2 + \dots + X_n])^2 \right] = \\ &= E\left[(X_1 + X_2 + \dots + X_n - E[X_1] - E[X_2] - \dots - E[X_n])^2 \right] = \\ &= E\left[((X_1 - E[X_1]) + (X_2 - E[X_2]) + \dots + (X_n - E[X_n]))^2 \right] = \\ &= VAR[X_1] + \dots + VAR[X_n] + \\ &\quad + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n E[(X_i - E[X_i])(X_j - E[X_j])] \end{aligned}$$

La “media campionaria”

Ricordiamo la media campionaria : $\bar{X}(n) \doteq \frac{1}{n} \sum_{i=1}^n X_i$

Se le X_i sono identicamente distribuite e con la stessa media, $\mu_i = \mu \quad \forall i$

$$\Rightarrow E[\bar{X}(n)] = \frac{1}{n} E\left[\sum_{i=1}^n X_i\right] = \frac{1}{n} \sum_{i=1}^n E[X_i] = \mu$$

Adesso:

$$\begin{aligned} Var[\bar{X}(n)] &= \frac{1}{n^2} E\left[\left(\sum_{i=1}^n X_i - E\left[\sum_{i=1}^n X_i\right]\right)^2\right] = \frac{1}{n^2} E\left[\left(\sum_{i=1}^n (X_i - \mu)\right)^2\right] \\ &= \frac{1}{n^2} E\left\{\sum_{i=1}^n (X_i - \mu)^2 + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n (X_i - \mu)(X_j - \mu)\right\} \\ &= \frac{1}{n^2} \left\{ \sum_{i=1}^n E[(X_i - \mu)^2] + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n E[(X_i - \mu)(X_j - \mu)] \right\} \end{aligned}$$

Allora, se le X_i sono anche indipendenti e con la stessa varianza, $\sigma_i^2 = \sigma^2 \quad \forall i$

$$Var[\bar{X}(n)] = \frac{1}{n^2} \left\{ n\sigma^2 + 0 \right\} = \frac{\sigma^2}{n} \text{ risultato sotto ipotesi di indipendenza}$$

Applicando la generalizzazione della somma calcoliamo la varianza della media campionaria!

ADDITIVITÀ DELLA VARIANZA

abbiamo visto poco fa che non è additiva
ma lo è in tutti i casi? c'è magari un caso in cui
lo è?

partiamo dalla covarianza

$$\begin{aligned} E[(X - E[X]) \cdot (Y - E[Y])] & \quad \text{Punto di partenza} \\ & = E[XY - YE[X] - XE[Y] + E[X]E[Y]] = E[XY] - E[Y]X - E[X]Y + E[X]E[Y] \\ & = E[XY] - E[X]E[Y] \end{aligned}$$

Importante perché collega la covarianza indirettamente a dipendenza tra variabili aleatorie.

Ricordando che l'indipendenza è stata definita sulle funzioni di densità e non sul valore atteso (v.a. indipendenti se le funzioni di densità sono indipendenti)

Osservazione:

Se per ipotesi **X e Y INDIPENDENTI** $\Rightarrow E[X \cdot Y] = E[X] \cdot E[Y]$ e la covarianza si annulla

In definitiva: se vale l'indipendenza sulla forma (funzione di densità e non su gli attributi ovvero i valori attesi) la covarianza si annulla

Dunque:

$$VAR[X + Y] = VAR[X] + VAR[Y] \equiv VAR[X - Y]$$

la varianza della somma e la varianza della differenza con l'ipotesi di indipendenza sono la stessa cosa

L' ADDITIVITÀ DELLA VARIANZA SCATURISCE DALLA INDIPENDENZA

Modulare la varianza attorno alla stessa media

vista anche con la Erlang modulata

siamo in un modello esponenziale

X = “durata di un fenomeno”;

$$X \doteq X_1 + X_2 + \dots + X_n$$

$$X_i \approx \exp(\lambda) \quad i = 1, \dots, n \quad \text{INDIPENDENTI}$$

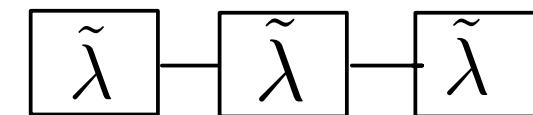
$$E[X] = n E[X_i] = n \frac{1}{\lambda}$$

$$VAR[X] = n VAR[X_i] = n \frac{1}{\lambda^2}$$

l'idea sta nel riscrivere il valore atteso in somma
di n blocchi esponenziali con valori attesi ridotti

Ridefinizione:

$$\tilde{\lambda} \doteq n \lambda$$



quindi il valore atteso non cambia

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

$$VAR[X] = n \left(\frac{1}{n \tilde{\lambda}} \right)^2 \rightarrow \frac{VAR[X]}{n}$$

qui n diventa il fattore
di modulazione

E' stata ridotta la varianza senza alterare la media!

Concetto e formula della covarianza

Definizione:

valore atteso di prodotto di scarti, per noi X input, Y output di un sistema

$$E[(X - E[X])(Y - E[Y])] \hat{=} COV(X, Y)$$

se ad esempio X aumenta rispetto a $E[X]$, ci si aspetta che Y aumenti, quindi correlazione positiva. se $E[X]$ molto vicina a X allora anche lo scarto tra Y e $E[Y]$ sarà basso: correlazione negativa.

Importanza: offre un'indicazione (sia pure abbastanza limitata) di come X e Y varino l'una relativamente all'altra.
ovvero dice il verso ma non la quantità

il nostro wish è di collegare covarianza e dipendenza ma come?

Formula di calcolo

$$COV(X, Y) = E[X \cdot Y] - E[X] \cdot E[Y]$$

qui si apre uno spiraglio per la dipendenza,
perché come visto se X e Y sono
indipendenti, la covarianza è nulla

(Richiede la densità congiunta di X,Y e le due densità marginali!)

per semplicità ci si pone il caso di un input ed un output ma nella realtà vi sono più input e più output

Applicazione:

la covarianza ci deve aiutare ad abbassare la varianza, quindi la dipendenza è buona se abbassa la varianza non se l'aumenta



bisogna capire quale input incide maggiormente sul output!

sarebbe interessante capire se una volta trovato un attributo in comune si ha qualche tipo di dipendenza

DIFFICILE CON L'ANALISI PROBABILISTICA!

Correlazione di una coppia di variabili

$COV(X, Y) \neq 0 \Rightarrow X \text{ e } Y \text{ correlate}$

$COV(X, Y) = 0 \Rightarrow X \text{ e } Y \text{ NON correlate}$

La covarianza nulla implica, (vedi Pearson) che non sono correlate LINEARMENTE!

$X \text{ e } Y \text{ indipendenti} \Rightarrow E[XY] = E[X]E[Y]$ implica covarianza nulla
 \Rightarrow pure non correlate

Se X e Y sono indipendenti o a correlazione nulla

$$\begin{aligned} VAR[X + Y] &= VAR[X] + VAR[Y] + 2COV(X, Y) \\ &= VAR[X] + VAR[Y] \end{aligned}$$

covarianza nulla
non vuol dire che
non ci sia una relazione
di grado più alto!

In pratica Pearson
è il coefficiente di correlazione
che appunto può essere negativa o positiva
essendo $-1 \leq \rho(X, Y) \leq 1$

Il coefficiente di Pearson

da non confondere con il $\frac{\lambda}{\mu} = \rho$ modello P.C.

Correlazione positiva:

una variabile aumenta quando aumenta anche l'altra.

Correlazione negativa:

quando una variabile aumenta, l'altra diminuisce, e viceversa.

$$\rho(X, Y) \doteq \frac{COV(X, Y)}{\sqrt{VAR[X]VAR[Y]}}$$

normalizzo la varianza
per rendere il coefficiente adimensionale

**è una misura, adimensionale, del grado di linearità o di correlazione lineare
della correlazione fra le variabili X e Y.**

pearson cattura bene il caso in cui è molto piccolo e molto vicino a 1, quelli nel mezzo no

(Teorema)

Proprietà n°1: $-1 \leq \rho(X, Y) \leq 1$

importanza: la covarianza ha un range ampio
e Pearson si pone di restringere tale range

(Teorema)

Proprietà n°2: $\rho(X, Y) =$

se rho=1 c'è linearità. e se c'è linearità rho=1
linearità su tutta la forma!

$$\begin{cases} -1 & \text{ses } Y = -aX + b, \quad a > 0 \\ 0 & \text{ses } X, Y \text{ non correlate} \\ +1 & \text{ses } Y = +aX + b, \quad a > 0 \end{cases}$$

Dimostrazione della Proprietà n°1:

da non fare

$$-1 \leq \rho(X, Y) \leq 1$$

Partenza: $\begin{cases} W \text{ è v.a. : } E[W^2] \neq 0 \\ Z \text{ è v.a. : } E[Z^2] \neq 0 \end{cases}$

NOTA: $(aW - Z)^2 \geq 0 \quad (a \in \mathbb{R})$
 $\Rightarrow E[(aW - Z)^2] \geq 0$

allora: $E[(aW - Z)^2] = a^2 E[W^2] - 2a \cdot E[W \cdot Z] + E[Z^2] \geq 0$

Si pone: $a \triangleq E[W \cdot Z] / E[W^2]$

e si osserva: $\frac{(E[W \cdot Z])^2}{(E[W^2])^2} E[W^2] - 2 \frac{E[W \cdot Z]^2}{E[W^2]} + E[Z^2] \geq 0$

$$\Rightarrow -\frac{(E[W \cdot Z])^2}{E[W^2]} + E[Z^2] \geq 0 \quad \Rightarrow \frac{(E[W \cdot Z])^2}{E[W^2] \cdot E[Z^2]} \leq 1$$

Ponendo ancora: $W \triangleq X - \mu_X; Z \triangleq Y - \mu_Y$ risulta $-1 \leq \rho(X, Y) \leq +1$

Dimostrazione della Proprietà n°2

questa serve e dimostra
che la linearità implica rho=1 in modulo

$$\text{Da } \rho(X, Y) \triangleq \frac{\text{COV}[X, Y]}{\sqrt{\text{VAR}[X]\text{VAR}[Y]}}$$

risulta :

partendo dalla covarianza NO INDIPENDENZA

$$\begin{aligned} \text{COV}[X, Y] &= E[X \cdot Y] - E[X] \cdot E[Y] && \text{provo ad usare l'ipotesi} \\ &= E[X \cdot (b + aX)] - E[X] \cdot E[b + aX] \\ &= b \cdot E[X] + a \cdot E[X^2] - E[X] \cdot (b + aE[X]) && = a \cdot (E[X^2] - E[X]^2) = a \cdot \text{VAR}[X] \end{aligned}$$

$$\text{Analogamente: } \text{COV}[X, Y] = a^{-1} \text{VAR}[Y] \quad \text{con } X = \frac{Y + b}{a}$$

$$\text{Dunque: } \text{COV}[X, Y]^2 = a \text{VAR}[X] \cdot a^{-1} \text{VAR}[Y] = \text{VAR}[X] \cdot \text{VAR}[Y]$$

$$\text{ovvero: } \left((\text{COV}[X, Y]) / \sqrt{\text{VAR}[X]\text{VAR}[Y]} \right)^2 = \rho^2 = 1 \Rightarrow |\rho| = 1$$

ipotesi: se c'è linearità allora

Con riferimento alla parte non banale:

$$|\rho(X, Y)| = 1 \Leftrightarrow Y = aX + b,$$

per ora, dimostriamo la sola implicazione \Leftarrow

Dunque se X e Y sono correlate, implica che la covarianza sia diversa da 0 grazie a Pearson

Analisi di correlazione in un modello P-C

VEDI EXCEL

IPOTESI: il produttore immette i prodotti appena realizzati nel buffer in accordo al suo ritmo di produzione , senza tenere conto del ritmo di consumo.

Cosa può accadere? ci può essere congestione

TESI: perché: moltissime operazioni veloci, oppure poche operazioni ma complicatissime

nel caso di produttore consumatore ho ritmo di interarrivi molto alti quindi burst oppure pochi arrivi ma operazioni lunghissime

1) assenza di correlazione fra tempo di produzione e tempo di consumo; non ho quindi dipendenza

2) una forte **correlazione positiva** ^{aumentano assieme quindi} fra tempo di vita dei prodotti e tempo di giacenza è segno che il sistema lavora con livelli e tempi di giacenza alti; ^(molti clienti in attesa) Se la giacenza è quasi uguale al servizio allora probabilmente il server è vuoto

Qual è il punto di saturazione del sistema? Il valore rho visto nel modello P.C.

3) una correlazione più o meno significativa, negativa, fra tempi di produzione collocati in un **determinato intervallo temporale** e tempi di giacenza collocati in un intervallo successivo;

ho un burst di richieste in determinato intervallo quanto tempo devo attendere affinché termini?

Valore atteso condizionato e curva di regressione

a partire da un valore atteso condizionato

$$E[Y | X = x] \hat{=} \int_{y=0}^{\infty} y \cdot f(y | x) \cdot dy$$

in generale è una
Funzione
non lineare!

$$E[Y | X = x], \quad 0 \leq x < \infty$$

al variare di x ho una funzione in particolare una curva

Definisco **La retta di regressione:**

ponendo

$$E[Y | X = x] \hat{=} a \cdot x + b, \quad 0 \leq x < \infty$$

la regressione fa parte
della statistica descrittiva

dove:

coefficiente angolare
della retta dunque
se $\rho=0$ la retta non
ha alcuna inclinazione

$$a = \rho \frac{\sqrt{VAR[Y]}}{\sqrt{VAR[X]}}$$

$$b = E[Y] - \rho \frac{\sqrt{VAR[Y]}}{\sqrt{VAR[X]}} E[X]$$

grazie alla dimostrazione di a
calcoleremo lo stimatore della
covarianza e di Pearson e dimostreremo che $\rho = 1 \implies$ linearità/retta di regressione perfetta

DEVONO ESSERE DEMOSTRATE!

dimostrazione retta di regressione

PROVA:

si vuole dimostrare che un particolare valore atteso condizionato è la retta
dunque parto dal valore atteso condizionato

$$\int_{y=0}^{\infty} y \cdot f(y | x) \cdot dy \stackrel{\uparrow}{=} E[Y | X = x]$$

tengo fermo x, ho una funzione di solo y

dalla condizionata
passo alla congiunta

$$f_{Y/X}(y/x) \stackrel{def}{=} \frac{f_{Y,X}(x,y)}{f_X^{(m)}(x)}$$

$$\frac{1}{f_X^{(m)}(x)} \int_{y=0}^{\infty} y \cdot f_{Y,X}(x,y) \cdot dy = E[Y | X = x]$$

$$\int_0^{\infty} y f_{Y,X}(x,y) dy = (ax + b) f_X^{(m)}(x)$$

sfrutto l'ipotesi

$$\int_{x=0}^{\infty} \int_0^{\infty} y f_{Y,X}(x,y) dy \cdot dx = \int_{x=0}^{\infty} (ax + b) f_X^{(m)}(x) \cdot dx$$

a questo punto passo da congiunta
a marginale e trovo il risultato finale

$$E[Y] = a \cdot E[X] + b$$

si è dimostrato che il valore
atteso è lineare ma la varianza?

OSSERVAZIONE: la relazione lineare tra i valori attesi è cosa ben
più debole della relazione lineare tra le variabili aleatorie X e Y!

che si scopre con la perfetta regressione

seconda parte: si parte dall'ipotesi

$$E[Y | X = x] \doteq a \cdot x + b$$

$$\int_0^\infty x \cdot f_X^{(m)}(x) \cdot E[Y | X = x] \cdot dx = \int_0^\infty x \cdot f_X^{(m)}(x) \cdot (ax + b) \cdot dx$$

$$\int_0^\infty x \cdot f_X^{(m)}(x) \cdot \left\{ \int_0^\infty y \cdot f_{Y|X}(y | x) \cdot dy \right\} \cdot dx = b \cdot E[X] + a \cdot E[X^2]$$

condizionante si rimette la definizione
 condizionata

$$\int_0^\infty \int_0^\infty xy \cdot f_{X,Y}(x, y) \cdot dxdy = b \cdot E[X] + a \cdot E[X^2]$$

congiunta e definizione
del valore atteso del prodotto

$$E[X \cdot Y] = b \cdot E[X] + a \cdot E[X^2]$$

valore atteso del prodotto è la somma tra momento del primo
ordine e momento del secondo ordine?

dalla formula di calcolo della covarianza

$$COV(X, Y) = E[X \cdot Y] - E[X] \cdot E[Y]$$

$$E[X \cdot Y] = b \cdot E[X] + a \cdot E[X^2]$$

terza parte: si parte dal risultato ottenuto per trovare a

$$\begin{aligned} E[X \cdot Y] &= COV[X, Y] + E[X]E[Y] \\ &= \rho \sqrt{VAR[X]} VAR[Y] + E[X]E[Y] \end{aligned}$$

sfruttando la def. di momento del secondo ordine

$$E[X^2] = Var[X] + E[X]^2$$

$$\rho \sqrt{VAR[X]} VAR[Y] + E[X]E[Y] = b \cdot E[X] + a \cdot \{Var[X] + E[X]^2\}$$

$$ricordando: E[Y] = a \cdot E[X] + b \quad \Rightarrow \quad b = E[Y] - a \cdot E[X]$$

primo risultato della dimostrazione

$$\rho \sqrt{VAR[X]} VAR[Y] + E[X]E[Y] = \{E[Y] - aE[X]\} \cdot E[X] + a \cdot \{Var[X] + E[X]^2\}$$

importante per analisi statistica

da cui ricavo
a e di conseguenza
b

$$a = \rho \frac{\sqrt{VAR[Y]}}{\sqrt{VAR[X]}}$$

$$b = E[Y] - \rho \frac{\sqrt{VAR[Y]}}{\sqrt{VAR[X]}} E[X]$$

$$E[Y | X = x] \doteq E[Y] + \rho \frac{\sqrt{VAR[Y]}}{\sqrt{VAR[X]}} (x - E[X]), \quad 0 \leq x < \infty$$

espressione analitica

Stima dei parametri della retta di regressione

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

Supponiamo di avere

Coppie di realizzazioni
delle v.a. X e Y

definiamo una funzione anche facile da derivare

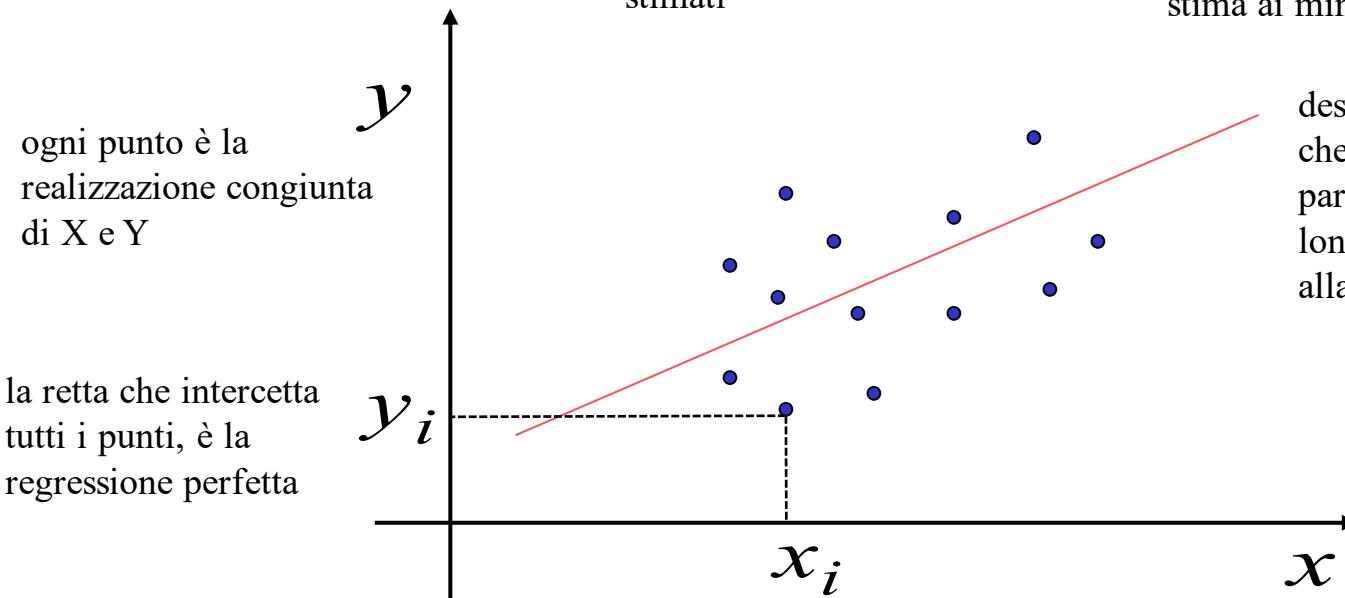
$$si\ pone:\ f(a,b) \hat{=} \sum_{i=1}^n [y_i - (ax_i + b)]^2$$

distanza tra il punto y_i ed il corrisponde punto
sulla retta (scarto, o meglio distanza dalla retta)

si cercano (\hat{a}, \hat{b}) : problema di minimo
valori stimati

$$f(\hat{a}, \hat{b}) = \min_{(a,b)} \sum_{i=1}^n [y_i - (ax_i + b)]^2$$

stima ai minimi quadrati



desidero la retta migliore
che si sistema tra i punti in
particolare: voglio che il punto più
lontano sia il più vicino possibile
alla retta

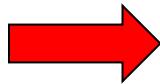
Stima ai minimi quadrati

calcolo le derivate parziali e le pongo pari a 0 e risolvo il sistema

$$\begin{cases} \frac{\partial f(a,b)}{\partial b} = -2 \cdot \sum_{i=1}^n (y_i - b - ax_i) = 0 \\ \frac{\partial f(a,b)}{\partial a} = -2 \cdot \sum_{i=1}^n (y_i - b - ax_i) \cdot x_i = 0 \end{cases}$$



$$\begin{cases} n \cdot b + a \cdot \sum_{i=1}^n x_i - \sum_{i=1}^n y_i = 0 \\ b \cdot \sum_{i=1}^n x_i + a \cdot \sum_{i=1}^n x_i^2 - \sum_{i=1}^n x_i y_i = 0 \end{cases}$$



$$\hat{a} = \frac{n \cdot \sum_{i=1}^n x_i \cdot y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \cdot \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2}$$

$$\hat{b} = \frac{\sum_{i=1}^n y_i}{n} - \hat{a} \cdot \frac{\sum_{i=1}^n x_i}{n}$$

Stima della covarianza

ricavata

da minimi quadrati

qui uso il teorema 2 di Pearson!

$$a = \frac{COV(X, Y)}{VAR[X]} \Rightarrow \hat{COV}(X, Y) = \hat{a} \cdot \hat{VAR}[X] = \frac{1}{n} \sum_{i=1}^n [(x_i - \bar{x}) \cdot (y_i - \bar{y})]$$

valore vero di a mi viene da pensare
che possa valere anche per la stima

parto dal risultato dei minimi quadrati

$$\hat{a} = \frac{n \cdot \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \cdot \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2} = \frac{\left(n \cdot \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i \right) / n^2}{\left[n \cdot \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2 \right] / n^2}$$

svolgo passaggi algebrici

$$= \frac{\frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{x} \cdot \bar{y}}{\frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2} = \frac{\frac{1}{n} \sum_{i=1}^n [(x_i - \bar{x}) \cdot (y_i - \bar{y})]}{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$$

stimatore della varianza
non corretto da MOM

risultato noto ovvero stimatore corretto della varianza

Stima del coefficiente di Pearson

partendo dal risultato vero

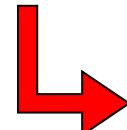
$$\rho \hat{=} \frac{COV(X, Y)}{\sqrt{VAR[X] \cdot VAR[Y]}}$$

grazie alla definizione di a

$$a = \frac{COV(X, Y)}{VAR[X]}$$



$$\rho = a \cdot \frac{\sqrt{VAR[X]}}{\sqrt{VAR[Y]}}$$



utilizzando
la stima ai
minimi quadrati
 $\hat{a} = \dots =$

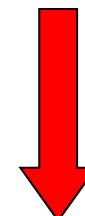
$$\frac{1}{n} \sum_{i=1}^n [(x_i - \bar{x}) \cdot (y_i - \bar{y})]$$

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

$$(i=1 \dots n)$$

**Stima ai minimi
quadrati**

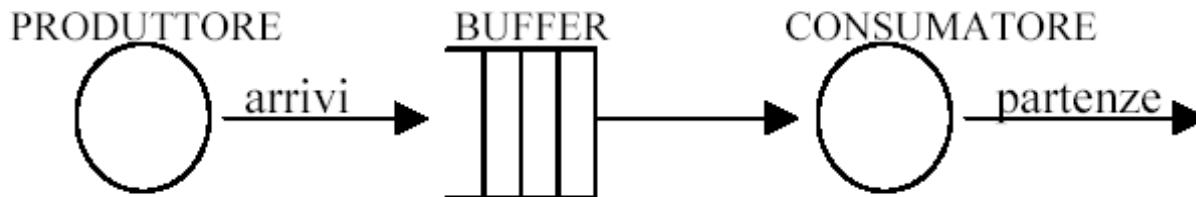
uso lo stimatore di a perchè lo abbiamo trovato!



$$\hat{\rho} = \frac{\frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{x} \cdot \bar{y}}{\sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2} \sqrt{\frac{1}{n} \sum_{i=1}^n y_i^2 - \bar{y}^2}}$$

sostituendo nel coefficiente ottengo la stima

Analisi (puntuale) della correlazione



$P :=$ "tempo fra due produzioni consecutive";
 $C :=$ "tempo fra due consumi consecutivi"; servizi
 $G :=$ "tempo di giacenza nel buffer"; soggiorno
 $V :=$ "tempo di vita (tempo di giacenza + tempo di consumo)"

Per tutte le variabili aleatorie di cui sopra sono state registrate 30 realizzazioni a partire dal prodotto n° 2001 e fino al 2030.

Per la sola variabile P sono state registrate ulteriori 30 osservazioni, riferite al gruppo di prodotti dal n°1971 al 2000, per effettuare analisi di correlazione fra tempi di produzione e tempi di giacenza temporalmente sfalsati.

Analisi (puntuale) della correlazione

Risultati delle 3 domandi
precedenti della lezione 24

STIME CON:

meglio vedere EXCEL

$$\hat{\rho} = \frac{\frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{x} \cdot \bar{y}}{\sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2} \sqrt{\frac{1}{n} \sum_{i=1}^n y_i^2 - \bar{y}^2}}$$

$$\hat{\rho}(P, C) = 7.37 \times 10^{-5}$$

Assenza di correlazione fra produzione e consumo.

$$\hat{\rho}(V, G) = 0.993$$

Fortissima correlazione fra tempi di vita e tempi di giacenza (forte congestione)

$$\hat{\rho}(P_{[1971-2000]}, G) = -0.433$$

Correlazione negativa fra tempi di produzione e tempi di giacenza

$$\hat{\rho}(P_{[2001-2030]}, G) = -0.095$$

Debole correlazione fra tempi di produzione e tempi di giacenza

Che relazione c'è tra **Retta di regressione e varianza degli scarti**

scarti

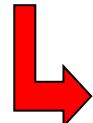
$$\delta_i \hat{=} |y_i - (a \cdot x_i + b)|, \quad i = 1, 2, \dots n$$

realizzazioni di Δ , $\Delta \hat{=} Y - (aX + b)$
distribuite come una Gaussiana

vedo

δ_i è la realizzazione i-ma della var al Δ (scarto)
ci interessano valore atteso e varianza

Serve per dimostrare che
 $\rho = 1 \implies$ linearità
quindi i punti stanno sulla retta



$$E[\Delta] = E[Y] - (a \cdot E[X] + b) = 0$$

(Dalla diapositiva n°9)

$$VAR[\Delta] = E^2[\Delta] - (E[\Delta])^2 = E^2[\Delta]$$

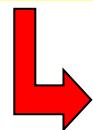
grazie alla dimostrazione di \hat{a}

momento del secondo ordine

Col metodo dei minimi quadrati: tornando indietro

$$f(a, b) \hat{=} \sum_{i=1}^n [y_i - (ax_i + b)]^2 = \sum_{i=1}^n \delta_i^2$$

ricorda un po'
lo stimatore del
momento del
secondo ordine



$$\text{attraverso} \quad \text{minimi quadrati} \iff E^2[\Delta] = \left(\sum_{i=1}^n \delta_i^2 \right) / n$$

La retta di regressione determinata col metodo dei minimi quadrati è quella di minima varianza

se la retta di regressione è quella di minima varianza la migliore è quindi quella con varianza nulla

Regressione lineare perfetta (a varianza nulla)

(i punti stanno tutti sulla retta)

La regressione lineare è perfetta quando risulta:

$$y_i = a \cdot x_i + b, \quad i = 1, 2, \dots n$$

Coefficiente di Pearson e perfetta regressione lineare

$$|\rho| = 1 \Rightarrow Y = aX + b$$

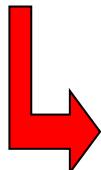
$$\rho = a \cdot \frac{\sqrt{VAR[X]}}{\sqrt{VAR[Y]}}$$

ricordando che

DIMOSTRAZIONE:

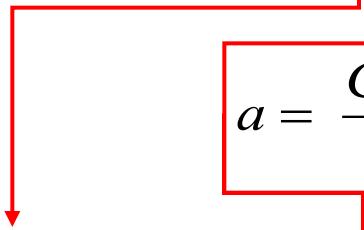
partiamo dallo scarto

$$\Delta \hat{=} Y - (aX + b)$$



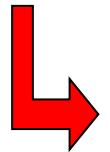
calcoliamo la varianza per arrivare alla varianza nulla

$$a = \frac{COV[X, Y]}{VAR[X]}$$



$$\begin{aligned} VAR[\Delta] &= VAR[Y] + a^2 VAR[X] - 2a \cdot COV[X, Y] \\ &= VAR[Y] + \rho^2 VAR[X] - 2\rho^2 \cdot VAR[X] \\ &= VAR[Y] \cdot (1 - \rho^2) \quad \text{sfrutto l'ipotesi che } |\rho| = 1 \end{aligned}$$

La varianza degli scarti e la varianza della Y ?



$$|\rho| = 1 \Rightarrow \rho^2 = 1 \Rightarrow VAR[\Delta] = 0$$

Test d'ipotesi sulla forma (basato sulla chi-quadrato)

fa parte della statistica inferenziale come gli intervalli di confidenza e stimatori

statistica in questo caso.

- ✓ “testare se la forma della distribuzione sperimentale (istogramma) dei dati provenienti dalla variabile aleatoria che rappresenta il tempo di attesa dei clienti nel buffer di un modello P-C è assimilabile ad una forma scientificamente nota”.

(differenza nulla tra le forme)

L'ipotesi formulata è detta (storicamente) ipotesi *nulla* (H_0) ed è quella che ha un fondamento di convinzione scientifica. La negazione di questa è detta ipotesi *alternativa* (H_1).

Alla base del test c'è la variabile aleatoria detta “statistica del test”, ricavata dal matematico. Il test si esegue per tentare di rigettare l'ipotesi nulla e se questo non riesce, si conclude che l'ipotesi è accettata, sì, ma solo fino a nuovo (eventuale) test!

Il test superato non vuol dire che la forma è buona per sempre, ma fino a che non si faccia un test che lo smentisca.

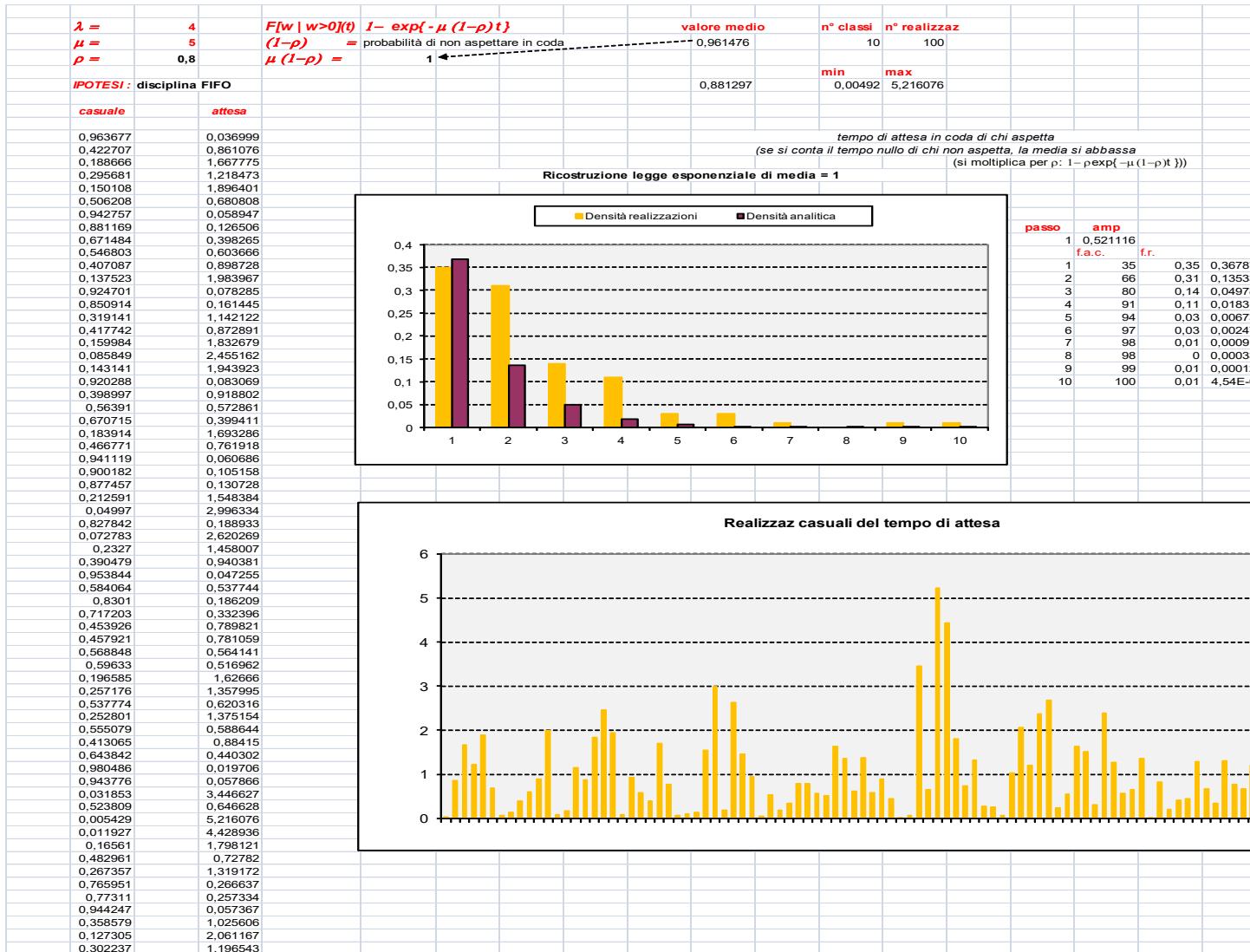
Le statistiche dei test interessanti sono variabili aleatorie con funzioni di densità/distribuzione nuove e non facili da ricavare!

Nel modello client-server faccio i test per protocolli complicati (Go back-n) non su protocolli FIFO.

Supponiamo di avere la statistica. Vediamo come procedere

File Excel: “Monte Carlo per attesa nel modello P-C” 27.0

possibile esercizio usare un parametro stimato invece di quello vero



Formulazione dell'ipotesi nulla - test della chi-quadrato

Raggruppando "n" realizzazioni indipendenti (osservazioni reali) su "k" intervalli (di ampiezza uguale o diversa) e costruendo il relativo istogramma di probabilità, si vuole testare l'ipotesi di "bontà della forma" di una fissata densità di probabilità con parametri noti perché già stimati (per ipotesi).
prima del test
tramite i metodi visti (MLE)

k è un parametro soggettivo,
lo scelgo io

p_j probab. calcolata per il j-esimo dei k intervalli

$$p_j = \int_{a_{j-1}}^{a_j} \hat{f}(x)dx = \hat{F}(a_j) - \hat{F}(a_{j-1}) \stackrel{\text{probabilità calcolata}}{=} \frac{1}{k} = \frac{j}{k} - \frac{(j-1)}{k}$$

equiprobabilità (per sistemare l'istogramma ed i suoi intervalli mi muovo su passi fissi sull'ordinata)

$$\Rightarrow a_j = \hat{F}^{-1}(j/k), \quad j = 1, \dots, k$$

oppure si possono utilizzare passi fissi sull'ascissa dunque equidistanza tra gli intervalli, quello che si è sempre fatto su Excel con il metodo Monte carlo

\hat{p}_j probab. stimata $\hat{=} N_j / n$

Nj= n° di realizzazioni in [aj-aj-1]

appunto realizzazioni che cadono nell'intervallo

IPOTESI NULLA:

$$H_0: p_1 = \hat{p}_1 \cap p_2 = \hat{p}_2 \cap \dots \cap p_k = \hat{p}_k$$

\forall coppia di probabilità, (reale e stimata), la differenza tra valore reale e stimato è nullo

Tipi di ERRORE E POTENZA del TEST

valgono anche per i test sulla media

- considerare falsa una ipotesi nulla che, in realtà, è vera
(errore di prima specie); avevi ragione ma il test ti dà torto
- considerare vera una ipotesi nulla che, in realtà, è falsa
(errore di seconda specie). hai torto ed il test dà ragione

$$\Pr(\text{errore di I specie}) = \Pr(\text{rigettare } H_0 \mid H_0 \text{ è vera}) = \alpha$$

$$\Pr(\text{errore di II specie}) = \Pr(\text{accettare } H_0 \mid H_0 \text{ è falsa}) = \beta$$

“*p-value*” di un test è il più piccolo livello di significatività (α), a partire dal quale verrebbe rigettata l’ipotesi nulla.
(è il rumore nel test. la soglia di errore che posso tollerare)

La *potenza* di un test d’ipotesi è la probabilità di rifiutare una ipotesi falsa, ossia $1 - \beta$ (complemento a 1 del errore di seconda specie)

Lo studio teorico delle probabilità di errore e della potenza del test non sarà affrontato ... solo valutazioni empiriche con Excel.

Teorema di Bernoulli nel test della chi-quadrato

(convergenza debole)

-) X_1, X_2, \dots var. al. ind. BERNOULLIANE

ovvero: $X_i = \begin{cases} 1 & \text{con prob} = p \\ 0 & \text{con prob} = 1 - p \end{cases}, \quad i = 1, 2, \dots, n$

X_i = Esito controllo di ciascuna realizzazione

$N_j \hat{=} \sum_{i=1}^n X_i$ è binomiale con: $E[N_j] = np_j, \quad VAR[N_j] = np_j(1 - p_j)$

n resta lo stesso perché la prova di Bernoulli ogni volta resetta tutto. Ad esempio sistemo 10 palline in intervalli. Ogni volta ripredendo le palline sistemate in un intervallo.

$$\Rightarrow \lim_{n \rightarrow \infty} \Pr \left\{ \left| \frac{N_j}{n} - p_j \right| < \delta \right\} = 1; \quad \text{segue} \quad \frac{N_j}{n} \hat{=} \hat{p}_j$$

Dimensione del campione!

STATISTICA chi-quadrato E TEOREMA

il test della chi-square si fa con molti dati a differenza della statistica

la difficoltà sta nel valutare T
unica incognita p_j

statistica $T \hat{=} \sum_{j=1}^k \frac{(N_j - np_j)^2}{np_j}$

sta per test

np_j è il valore atteso

$D \xrightarrow{\quad}$ significa che converge in distribuzione

$\rightarrow \chi_{k-1}^2$ $n \rightarrow \infty$

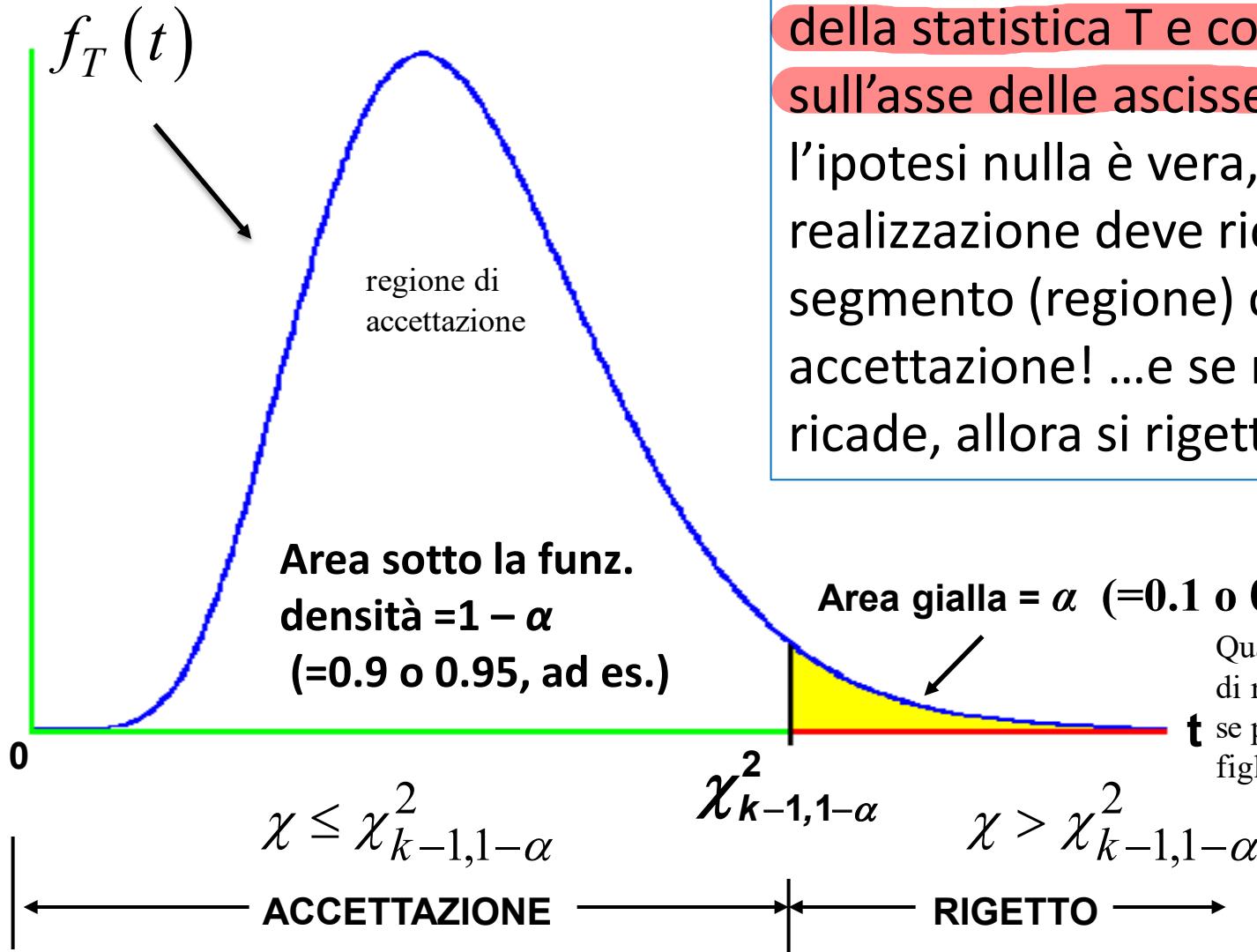
quello che ti trovi meno quello che ti aspetti

chi-square con $k-1$ gradi di libertà

con le osservazioni ricavo la realizzazione della variabile aleatoria statistica del test

Idea alla base del test della “chi-quadrato”

Eseguire il test



Produrre una realizzazione

della statistica T e collocarla
sull'asse delle ascisse: se

l'ipotesi nulla è vera, allora la
realizzazione deve ricadere nel
segmento (regione) di
accettazione! ...e se non ci
ricade, allora si rigetta l'ipotesi!

Quando cade nella regione
di rigetto, non si sa
se per caso, oppure perché
figlia di un'altra forma

se il numerato di T è 0 (grazie al teorema di Bernoulli) accetto il risultato del test

File Excel: “Test sulla forma (chi-quadrato)”

la distribuzione del tempo di attesa nel modello P-C non è un'esponenziale a causa di rho

	lambda =	media campionaria =	num realizzazioni =					
Realizzazioni di una esponenziale	0,5	2,047351785	100					
0,075346194			punto	valore punto				
2,707563293			num intervalli =	10				
0,232322275					0	0		
6,091987717					1	0,210721031	13	0,13
0,18924669					2	0,446287103	15	0,15
2,657011249					3	0,713349888	7	0,07
2,002398131					4	1,021651248	13	0,13
0,534989814					5	1,386294361	4	0,04
0,101932062					6	1,832581464	7	0,07
2,048269332					7	2,407945609	7	0,07
0,441871935					8	3,218875825	7	0,07
0,17681158					9	4,605170186	15	0,15
3,53604033					10		12	0,12
3,875885756							Controllo:	100
5,197645844								
7,267218052								
4,234475964								
0,215612772								
3,354254216								
0,224015794								
1,530687184								
0,440618103								
4,125106229								
0,278879725								
0,00498668								
0,040349821								
0,181563164								
5,129938918								
0,011654979								
0,763143499								
6,88274026								
1,12415307								
1,38082721								
0,031085778								
1,179165859								
4,724754296								
3,551444248								
4,777239159								
0,993702835								
0,397989517								
0,238267236								
2,290257657								
0,077515777								
3,2992926346								
0,325035554								
1,622442561								
0,273631468								
2,502733204								
0,491817623								
0,406004688								
3,980632873								
3,951417086								
0,87101393								
3,400963965								
0,233619852								
5,45602897								
1,31721751								
0,541354889								
3,825265355								
0,62036042								
2,363440276								
0,784221155								
0,984464318								
1,477481261								
0,481311859								
1,657093044								
0,862111436								
0,921162349								
0,877200088								

Abbiamo partitionato lo spazio dei risultati di questa esponenziale in dieci intervalli. Si è stati attenti affinché questi intervalli siano equiprobabili e nella colonna “valore punto” si trova il valore degli estremi di questi intervalli. Questi intervalli sono stati calcolati come :

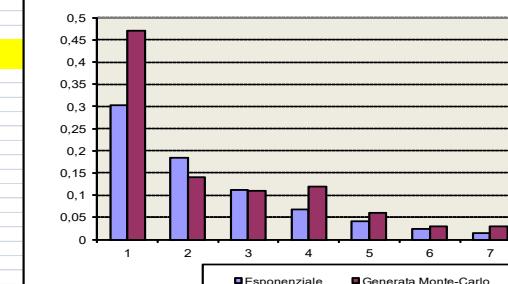
$p_j = \int_{a_{j-1}}^{a_j} f(x)dx = F(a_j) - F(a_{j-1}) \hat{=} 1/k$. In questo caso abbiamo 10 intervalli quindi $p_j = 0,1$. Possiamo porre $\frac{1}{k} = \frac{j}{k} - \frac{(j-1)}{k}$, da cui ricavare che $a_j = F^{-1}(j/k)$, con $j=1, \dots, k$. Abbiamo che $F^{-1}(x) = -\frac{1}{\lambda} \ln(1-x)$, quindi $a_j = F^{-1}(j/k) = -\frac{1}{\lambda} \ln(1-j/k)$ ed è proprio questa espressione che è stata usata per individuare gli intervalli. Notiamo che, per costruzione, gli intervalli non hanno la stessa ampiezza ma sono equiprobabili.

Per il teorema di Pearson sappiamo che $T \hat{=} \sum_{j=1}^k \frac{(N_j - np_j)^2}{np_j} \rightarrow \chi^2_{k-1}$, $n \rightarrow \infty$, (k indica il numero di intervalli utilizzati). Perciò i punti tali per cui $T > \chi^2_{k-1, 1-\alpha}$ rientrano nella regione di rigetto con livello α .

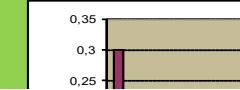
passo =	1
valore	1
1	0,30326533
2	0,18393972
3	0,11156508
4	0,06766764
5	0,0410425
6	0,02489353
7	0,01509869
8	0,00915782
9	0,0055545
10	0,00336897

oltre 10:	
Controllo:	

Risultato test:
OK



1/4 del valore atteso per rappresentare istogramma	valore	frequenze osservate
0,5	0,3894	0,194700196
1	0,303265	0,151632665
1,5	0,236183	0,118091638
2	0,18394	0,091969886
2,5	0,143252	0,071626199



ESEMPIO su TEST CHI QUADRATO

Verificare se la variabile aleatoria X , di cui 50 realizzazioni sono in tabella, si distribuisce con legge esponenziale.

Realizzazioni X									
79,919	3,027	6,769	18,387	144,695	0,941	0,624	0,590	7,004	3,217
3,081	6,505	59,899	0,141	2,663	0,878	5,380	1,928	31,764	14,382
0,062	0,021	1,192	43,565	17,967	3,371	3,148	0,300	1,005	1,008
1,961	0,013	34,760	24,420	0,091	2,157	7,078	0,002	1,1147	2,336
5,845	0,123	5,009	0,433	9,003	7,579	23,960	0,543	0,219	4,562

TEST CHI QUADRATO

k è soggettivo

Si raccomanda $k \geq 3$, $n * p_i \geq 5$ per ogni i . Inoltre, nel continuo, per la scelta del numero di intervalli, si faccia riferimento ai seguenti *range*.

Numerosità campione (n)	Numero intervalli (k)
20	non usare CQ
50	5 - 10
100	10 - 20
>100	$\sqrt{n} - n/5$

TEST CHI QUADRATO

Calcolo *upper bound* del numero di intervalli

$$k \leq 50/5=10$$

Scelta del numero di intervalli si arriva a k=8

$$k = 8$$

Calcolo p dei k intervalli (equiprobabili)

$$p = 1/8= 0,125$$

Stima di λ

$$\hat{\lambda} = 1/\text{media} = 1/11,893 = 0,084$$

Se fosse una Weibull la stima dei parametri risulterebbe problematica

TEST CHI QUADRATO

Calcolo estremi intervalli a_i a partire dalla funzione di distribuzione di probabilità si potrebbe usare una qualsiasi distribuzione con cui si può attuare il metodo della trasformazione inversa

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

$$ip = 1 - e^{-\lambda a_i} \quad \text{quindi intervelli equiprobabili}$$

$$e^{-\lambda a_i} = 1 - ip$$

$$a_i = -\frac{1}{\lambda} \ln(1 - ip)$$

Per qualsiasi valori di λ , gli estremi inferiore e superiori sono

$$a_0 = 0, a_k = \infty$$

TEST CHI QUADRATO

Calcolo estremi intervalli a_i per $i = 1, \dots, 7$

$$a_1 = -\frac{1}{\lambda} \ln(1 - 0,125) = 1,588$$

$$a_2 = -\frac{1}{\lambda} \ln(1 - 0,250) = 3,422$$

$$a_3 = -\frac{1}{\lambda} \ln(1 - 0,375) = 5,590$$

$$a_4 = -\frac{1}{\lambda} \ln(1 - 0,500) = 8,244$$

$$a_5 = -\frac{1}{\lambda} \ln(1 - 0,625) = 11,665$$

$$a_6 = -\frac{1}{\lambda} \ln(1 - 0,750) = 16,488$$

$$a_7 = -\frac{1}{\lambda} \ln(1 - 0,875) = 24,732$$

TEST CHI QUADRATO

numero osservazioni
nell'intervallo

↓

<i>Intervallo</i>	O_i	E_i	$(O_i - E_i)^2 / E_i$
[0, 1588)	19	6,25	26,01
[1,588, 3,422)	10	6,25	2,25
[3,422, 5,590)	3	6,25	1,69
[5,590, 8,244)	6	6,25	0,01
[8,244, 11,665)	1	6,25	4,41
[11,665, 16,488)	1	6,25	4,41
[16,488, 24,732)	4	6,25	0,81
[24,732, ∞)	6	6,25	0,01

con $E_i = n * p_i = 50 * 0,125 = 6,25$ per $i = 1, \dots, 8$.

in questo caso equiprobabili sono gli intervalli

TEST CHI QUADRATO

Sommando i valori dell'ultima colonna risulta $\chi^2_0 \leftarrow 39,6$.
realizzazione della statistica

Dalla tabella dei valori critici del test, per un livello di significatività $\alpha = 0,05$ risulta un valore di $\chi^2_{k-1, \alpha}$ pari a 12,6.

Essendo

$$\chi^2_0 = 39,6 > 12,6 = \chi^2_{k-1, \alpha}$$

si rigetta l'ipotesi nulla secondo la variabile aleatoria X si distribuisce con legge esponenziale.

E per un livello di significatività $\alpha = 0,01$?

**Statistica
e teorema**

$$T \doteq \sum_{j=1}^k \frac{(N_j - np_j)^2}{np_j} \xrightarrow{D} \chi^2_{k-1} \quad n \rightarrow \infty$$

Ipotesi

$\{B_j ; j = 1, \dots, k\}$ v.a. binomiali indipendenti
 perché osservazioni indipendenti
 con $E[B_j] = np_j$ e $Var[B_j] = np_j(1 - p_j)$

Binomiale standardizzata. $Y(n)$ perchè il campione è importante

Prova

$$Y(n) \doteq (B_1 - np_1) / \sqrt{np_1(1 - p_1)}$$

$$Y^2(n) \xrightarrow{D} \chi^2_1 \quad \text{per } n \rightarrow \infty$$

la binomiale al quadrato converge in forma ad una chi-squared al quadrato con 1 grado di libertà per n tendente all'infinito

B_j è il numero di realizzazioni che ricadono nell'intervallo j

$$F_{Y^2(n)}(z) \doteq \Pr(Y^2(n) \leq z) = \Pr(-\sqrt{z} \leq Y(n) \leq \sqrt{z})$$

grazie alla funzione di v.a. esempio 3.8

$$= F_{Y(n)}(\sqrt{z}) - F_{Y(n)}(-\sqrt{z})$$

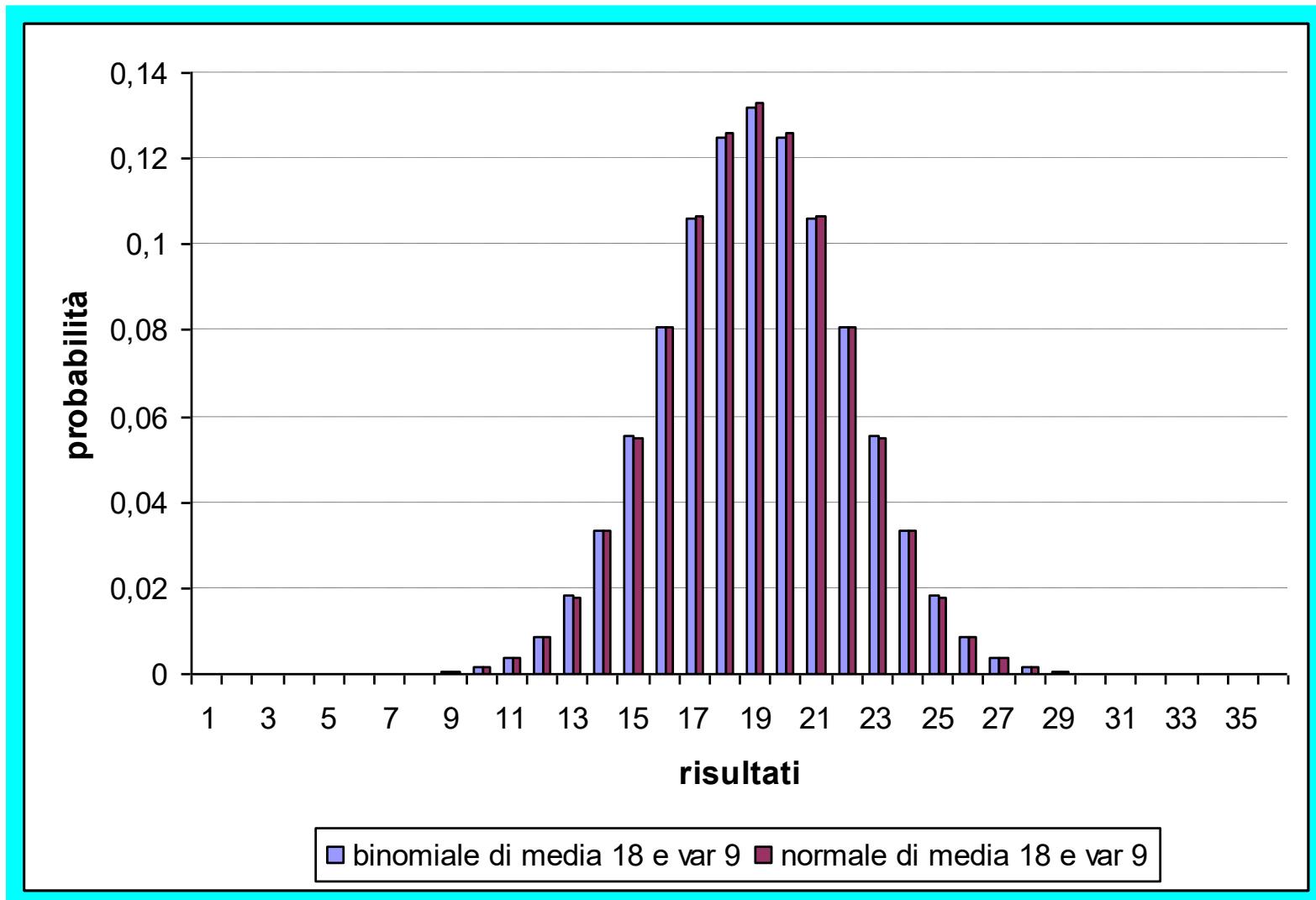
$$\rightarrow F_Z(\sqrt{z}) - F_Z(-\sqrt{z})$$

$$Y(n) \xrightarrow{D} Z \equiv N(0,1) \quad \text{per } n \rightarrow \infty$$

Teorema Limite Centrale

Illustrazione del T. L. C. per la Y(n) non standardizzata

risultato visto in precedenza



Dimostrazione già vista nella lezione delle funzioni variabili aleatorie

$$F_{Y^2(n \rightarrow \infty)}(z) \rightarrow F_Z(\sqrt{z}) - F_Z(-\sqrt{z}) =$$

$$= 2 \int_0^{\sqrt{z}} \left(1/\sqrt{2\pi}\right) \cdot \exp(-u^2/2) \cdot du$$

$$= 2 \int_0^{\sqrt{z}} \left(1/\Gamma(1/2)\sqrt{2}\right) \cdot \exp(-u^2/2) \cdot du$$

$$(\text{ con } u \hat{=} \sqrt{v} \Rightarrow du = v^{-1/2} \cdot dv = v^{1/2-1} \cdot dv)$$

$$= \int_0^z \left(1/\Gamma(1/2)2^{1/2}\right) \cdot v^{1/2-1} \cdot \exp(-v/2) \cdot dv$$

ricordando densità chi-quadro

$$f_{\chi_\gamma^2}(x) \hat{=} \left(1/\Gamma(\gamma/2)2^{\gamma/2}\right) \cdot x^{\gamma/2-1} \cdot e^{-x/2} \quad x \geq 0$$

$$\Rightarrow \lim_{n \rightarrow \infty} F_{Y^2(n)}(z) = F_{\chi_1^2}(z) = F_{Z^2}(z)$$

Risultato
intermedio
importante

Il quadrato di normale standard corrisponde a chi-quadro con un grado di libertà

Ritorniamo a $\{B_j ; j = 1, \dots, k\}$ v.a. binomiali indip.

Siano: $X_1 \hat{=} B_1$ e $X_2 \hat{=} n - X_1$, con $p_2 \hat{=} 1 - p_1$

X_1, X_2, \dots, X_{k-1} multinomiali con $E[X_{k-1}] = np_{k-1}$

$X_k = n - (X_1 + X_2, \dots, + X_{k-1})$; $p_k = 1 - (p_1 + \dots + p_{k-1})$

Ancora: $Y^2(n) = \left((B_1 - np_1) / \sqrt{np_1(1-p_1)} \right)^2 \hat{=} Q_1(n)$

$$\begin{aligned} Q_1(n) &= \frac{(X_1 - np_1)^2}{np_1 \cdot (1-p_1)} = \frac{(X_1 - np_1)^2}{np_1} + \frac{(X_1 - np_1)^2}{n(1-p_1)} \\ &= \frac{(X_1 - np_1)^2}{np_1} + \frac{(X_2 - np_2)^2}{np_2} \quad \approx \chi_1^2 \quad \text{per } n \gg 1 \end{aligned}$$

$$\text{nota: } \frac{(X_1 - np_1)^2}{n \cdot (1-p_1)} = \frac{(n - X_2 - n \cdot (1-p_2))^2}{n \cdot (1-p_1)} = \frac{(-X_2 + np_2)^2}{n \cdot p_2}$$

$$Q_{k-1}(n) \hat{=} \sum_{j=1}^k \frac{(X_j - np_j)^2}{np_j} \quad \approx \chi_{k-1}^2 \quad \text{per } n \gg 1$$

Il risultato del teorema è che la somma di quadrati di chi-quadrato è ancora una chi-quadrato

vista anche negli
intervalli di confidenza

La distribuzione “T di Student” (1/5)

risultato di Gosset

$$f_T(t) = \frac{\Gamma((\gamma+1)/2)}{\Gamma(\gamma/2)\sqrt{\pi\gamma}} \left(1 + \frac{t^2}{\gamma}\right)^{-(\gamma+1)/2} \quad -\infty < t < +\infty$$

La v.a. $T \equiv T_\gamma$ è detta "T di Student"

con γ "gradi di libertà" e corrisponde a:

Z è una Gaussiana standard

$$T_\gamma \doteq \frac{Z}{\sqrt{X_\gamma^2 / \gamma}}$$

distribuzione
del quoziente fra v.a.
ed una funzione della chi-square

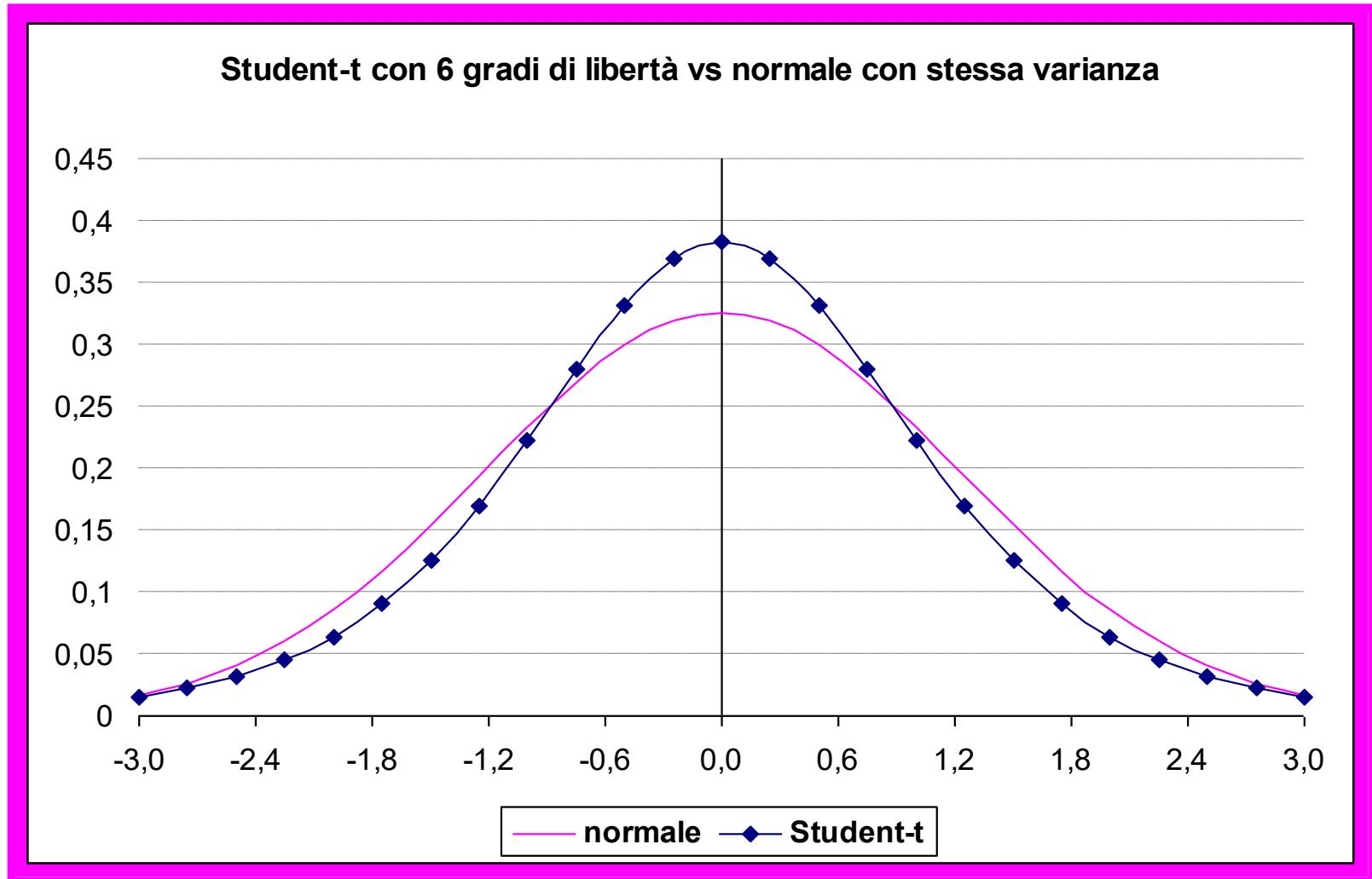
INDIPENDENZA

- **MEDIA E VARIANZA:**

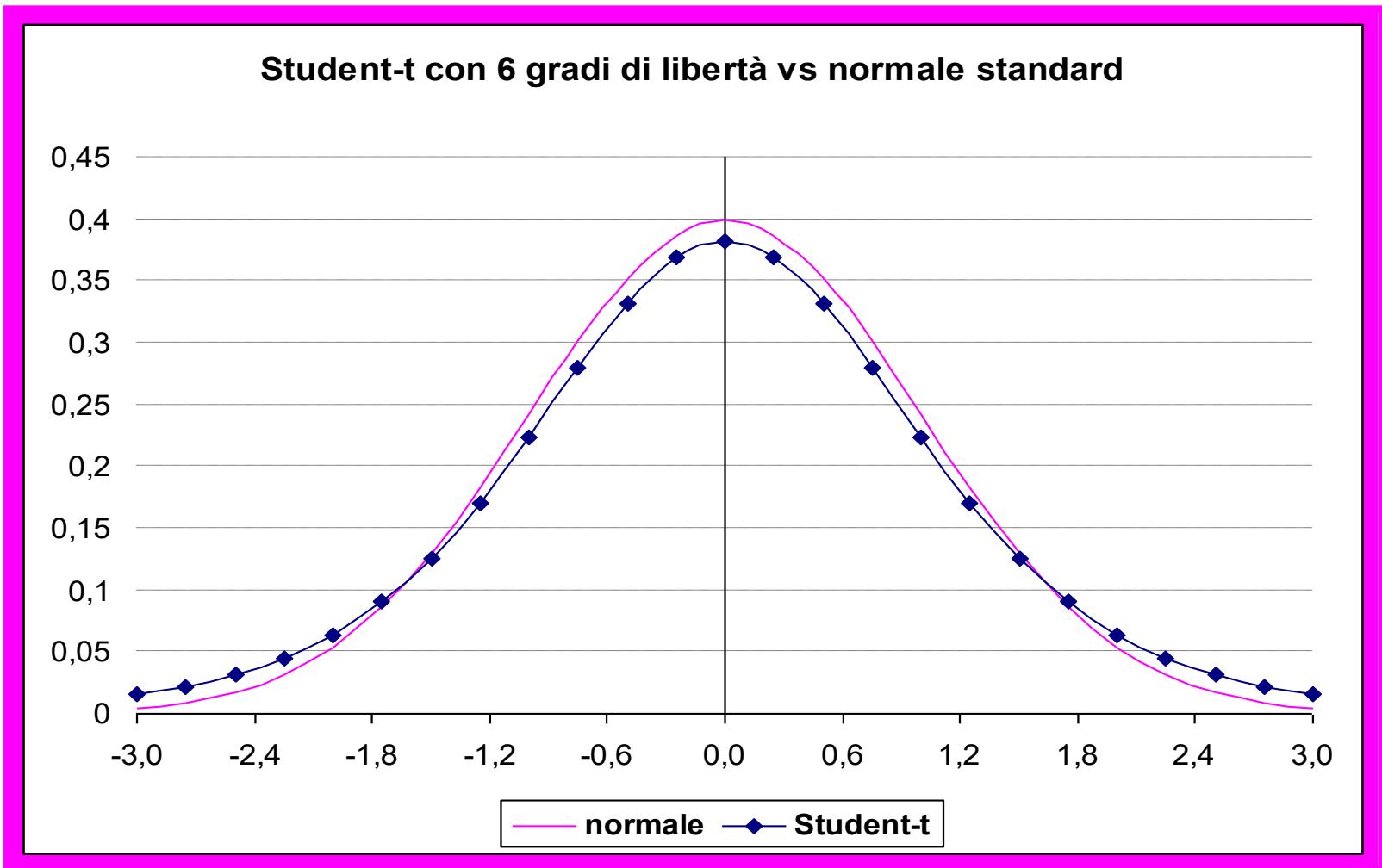
$$E[T] = 0 \quad VAR[T] = \gamma / (\gamma - 2) \quad \left\{ \rightarrow 1 \text{ per } \gamma \rightarrow \infty \right\}$$

come Gauss

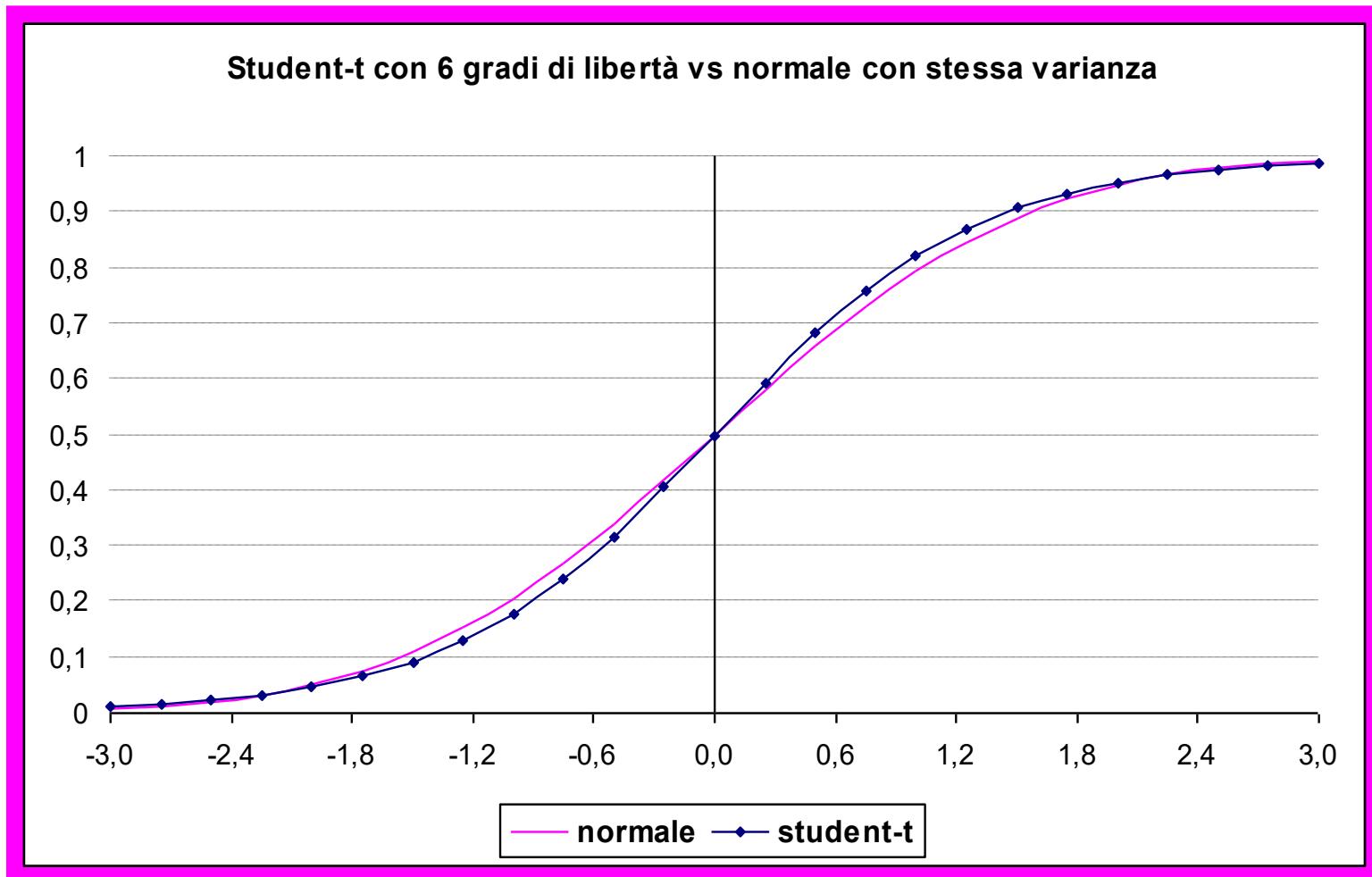
La distribuzione “T di Student” (3/5)



La distribuzione “T di Student” (4/5)



La distribuzione “T di Student” (5/5)



La distribuzione “T di Student” (2/5)

$$T_\gamma \triangleq \frac{Z}{\sqrt{\chi_\gamma^2 / \gamma}}$$

UTILE nella costruzione degli intervalli di confidenza per la media campionaria di osservazioni normali e indipendenti grazie ai seguenti risultati:

$$1) \quad \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} = Z \quad ;$$

no statistica

$$2) \quad \frac{(n-1)S^2}{\sigma^2} = X_{\gamma=n-1}^2$$

statistica vista negli intervalli di confidenza

$$\frac{\sqrt{n}(\bar{X} - \mu) / \sigma}{\sqrt{S^2 / \sigma^2}} = T_{\gamma=n-1}$$



$$\frac{\bar{X} - \mu}{S / \sqrt{n}} = T_{\gamma=n-1}$$

statistica



$$\left[\bar{X} - \frac{S}{\sqrt{n}} t_{\alpha/2}, \bar{X} + \frac{S}{\sqrt{n}} t_{\alpha/2} \right]$$

IMPORTANZA: per produrre un intervallo di confidenza per la media di un campione “piccolo” ($10 < n < 30$) di osservazioni indipendenti estratte da una normale.

per $n > 30$ il quantile della t-student è pari a quello della Gaussiana

Dall'intervallo di confidenza per la media al test di ipotesi

Vale lo stesso discorso fatto con il test con la chi-square

valore atteso

Siano X_1, X_2, \dots, X_n I.I.D. come $\approx N(\mu, \sigma^2)$ o con buona approx grazie al TLC

$$\text{Allora : } \bar{X}(n) \hat{=} \frac{\sum_{i=1}^n X_i}{n} \approx N(\mu, \sigma^2 / n)$$

Teorema di riproducibilità della normale oppure teorema limite centrale con buona approx

Statistica del test: $Z \hat{=} \frac{\bar{X}(n) - \mu}{\sigma / \sqrt{n}} \approx N(0,1)$
del intervallo (quando ne era uno)

siccome conosco la distribuzione Z posso scrivere

$$\Pr[-z_{\alpha/2} \leq Z \leq z_{\alpha/2}] = 1 - \alpha$$

$$\begin{aligned} \text{con: } \alpha/2 &\hat{=} \int_{-\infty}^{-z_{\alpha/2}} f_Z(z) dz \\ &\equiv \int_{z_{\alpha/2}}^{+\infty} f_Z(z) dz \end{aligned}$$

uso per intervallo

quindi in un intervallo

ovvero

uso la v.a. mentre nel test

una sua realizzazione

realizzazione

media

esecuzione:

uso per il test

$\frac{\bar{X}(n) - \mu}{\sigma / \sqrt{n}} \in [-z_{\alpha/2}, z_{\alpha/2}]$ con prob $(1 - \alpha)$

verifico se il valore calcolato finisce tra i quantili

(ovvero con S la sua stima)

stavolta

$1 - \alpha$ è il livello di

significatività del test

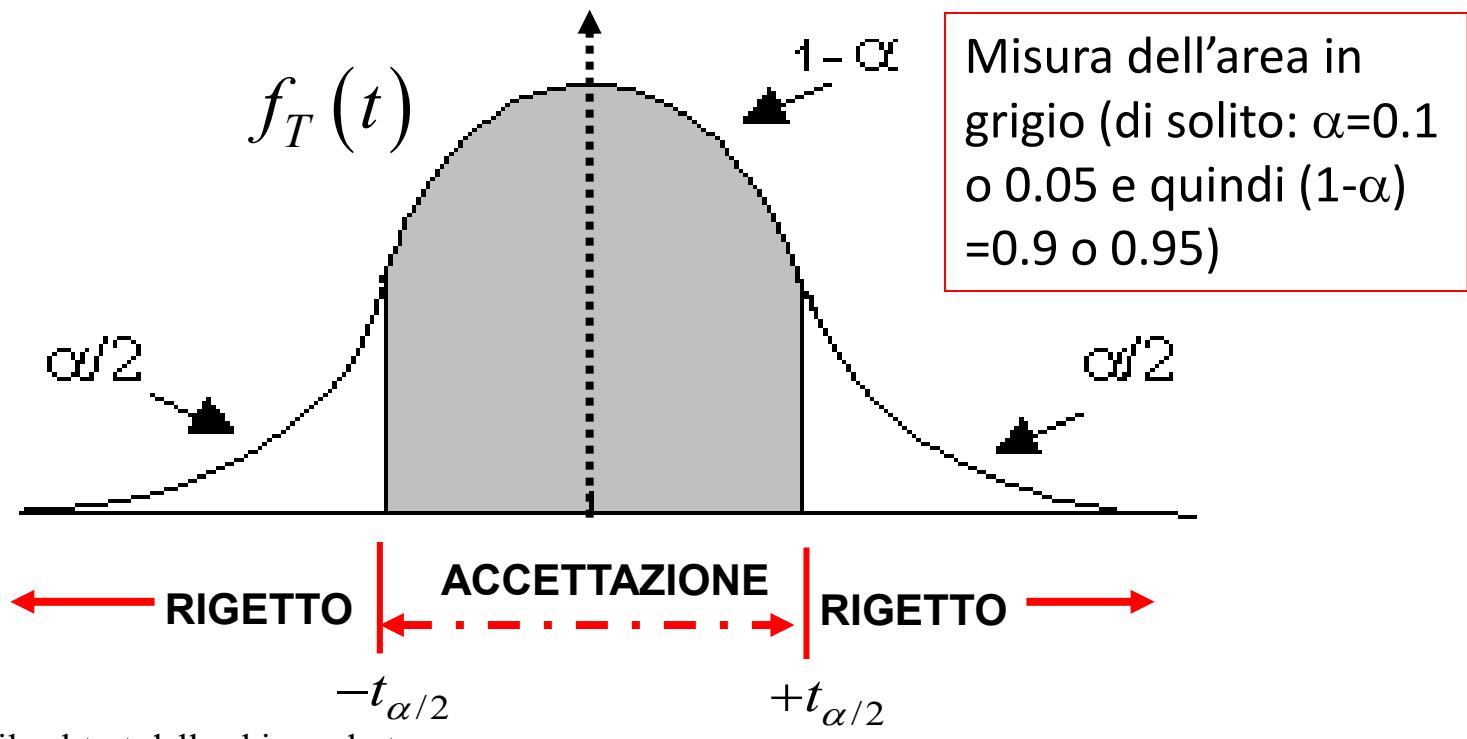
$\bar{x}(n)$ è una media aritmetica da dati reali o da risultati (stima) di simulazione Monte Carlo

Statistica del test: $T \hat{=} \frac{\bar{X}(n) - \mu}{S / \sqrt{n}} \approx Student-T_{\gamma=n-1}$
o dell'intervallo (con varianza stimata)

esecuzione: $\frac{\bar{x}(n) - \mu}{s / \sqrt{n}} \in [-t_{n-1; \alpha/2}, t_{n-1; \alpha/2}]$ con prob $(1 - \alpha)$

ESECUZIONE DEL TEST

Produrre una realizzazione della statistica T e collocarla sull'asse delle ascisse: se l'ipotesi nulla è vera, allora la realizzazione deve ricadere nel segmento (regione) di accettazione! ...e se non ci ricade, allora si rigetta l'ipotesi!



Tipi di ERRORE E POTENZA del TEST

come nel test sulla forma

- considerare falsa una ipotesi nulla che, in realtà, è vera (*errore di prima specie*);
- considerare vera una ipotesi nulla che, in realtà, è falsa (*errore di seconda specie*).

$$\Pr(\text{errore di I specie}) = \Pr(\text{rigettare } H_0 \mid H_0 \text{ è vera}) = \alpha$$

$$\Pr(\text{errore di II specie}) = \Pr(\text{accettare } H_0 \mid H_0 \text{ è falsa}) = \beta$$

La *potenza* di un test d'ipotesi è la probabilità di rifiutare una ipotesi falsa, ossia $1 - \beta$

Lo studio teorico delle probabilità di errore e della potenza del test non sarà affrontato ... solo valutazioni empiriche con Excel.

Significativa' fissata e potenza garantita

Fissato α , quale n garantisce un β voluto ?

Dimostrazione

$$\beta = \Pr(\text{accettare } H_0 \mid H_0 \text{ è falsa})$$

valore ipotesi reale

$$= \Pr\left(-z_{1-\alpha/2} \leq \frac{\bar{X} - w_0}{\sigma/\sqrt{n}} \leq z_{1-\alpha/2} \mid w_{re} - w_0 \hat{=} \delta > 0\right)$$

differenza tra nota e stima

$$= \Pr\left(-z_{1-\alpha/2} \leq \frac{\bar{X} - w_{re}}{\sigma/\sqrt{n}} + \frac{w_{re} - w_0}{\sigma/\sqrt{n}} \leq z_{1-\alpha/2}\right)$$

$$= \Pr\left(-z_{1-\alpha/2} \leq Z + \frac{\sqrt{n}\delta}{\sigma} \leq z_{1-\alpha/2}\right)$$

$$= \Pr\left(-z_{1-\alpha/2} - \frac{\sqrt{n}\delta}{\sigma} \leq Z \leq z_{1-\alpha/2} - \frac{\sqrt{n}\delta}{\sigma}\right)$$

$$= F\left(z_{1-\alpha/2} - \frac{\sqrt{n}\delta}{\sigma}\right) - F\left(-z_{1-\alpha/2} - \frac{\sqrt{n}\delta}{\sigma}\right)$$

per n tendente ad infinito diventa molto piccolo quindi trascurabile
cosa accade se utilizzo la T-student invece della Gaussiana?

ipotesi

Normalità
Varianza nota
Media reale nota

$$n \approx \frac{\sigma^2}{\delta^2} (z_{1-\alpha/2} + z_{1-\beta})^2$$

$$z_{1-\beta} = -z_{1-\alpha/2} + \frac{\sqrt{n}\delta}{\sigma}$$

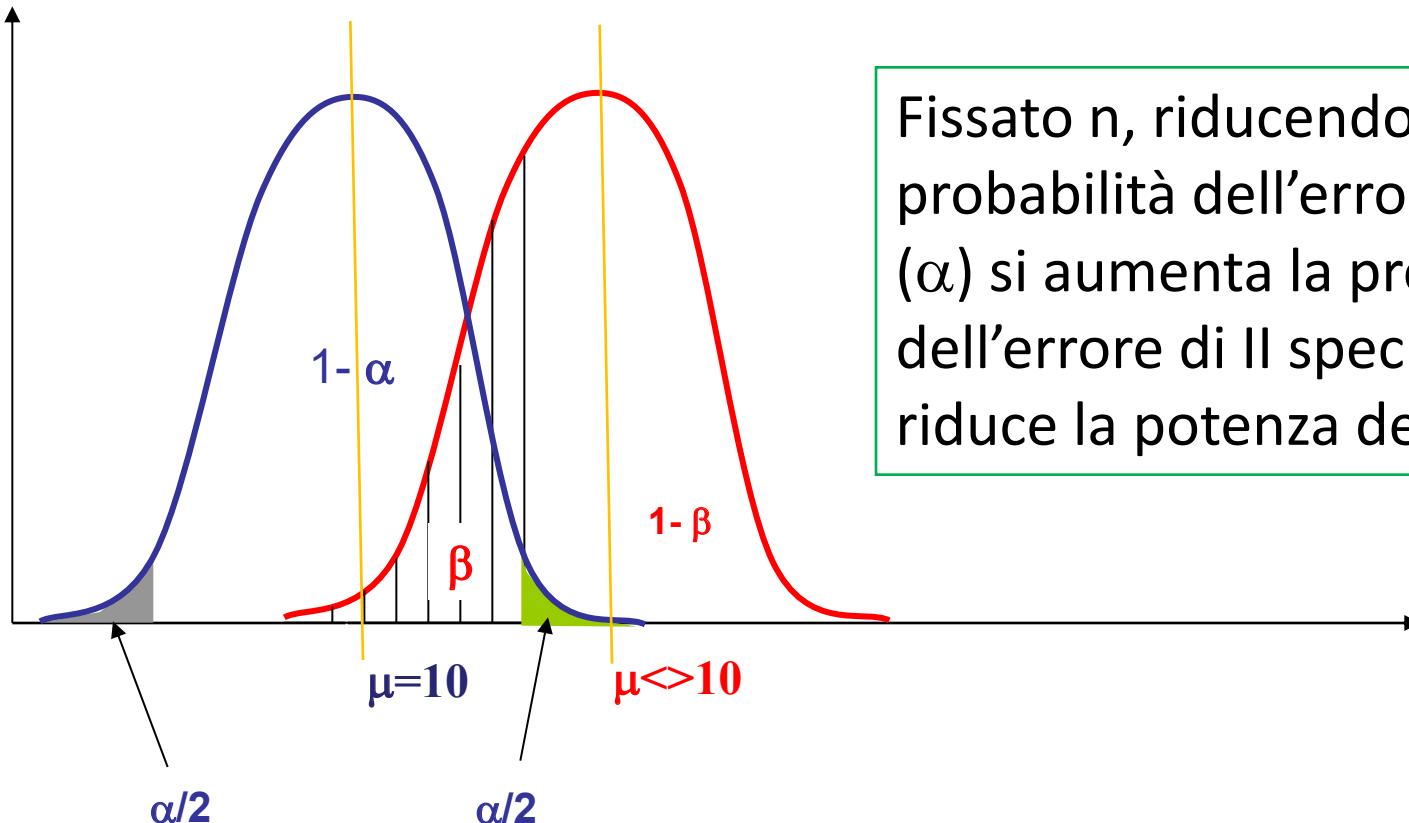
$$1 - \beta \cong F\left(-z_{1-\alpha/2} + \frac{\sqrt{n}\delta}{\sigma}\right)$$

$$\beta \cong F\left(z_{1-\alpha/2} - \frac{\sqrt{n}\delta}{\sigma}\right)$$

SIGNIFICATIVITÀ, POTENZA e «p-value»

$H_0: \mu = 10$ (funzione in blu) versus $H_1: \mu \neq 10$ (funzione in rosso)

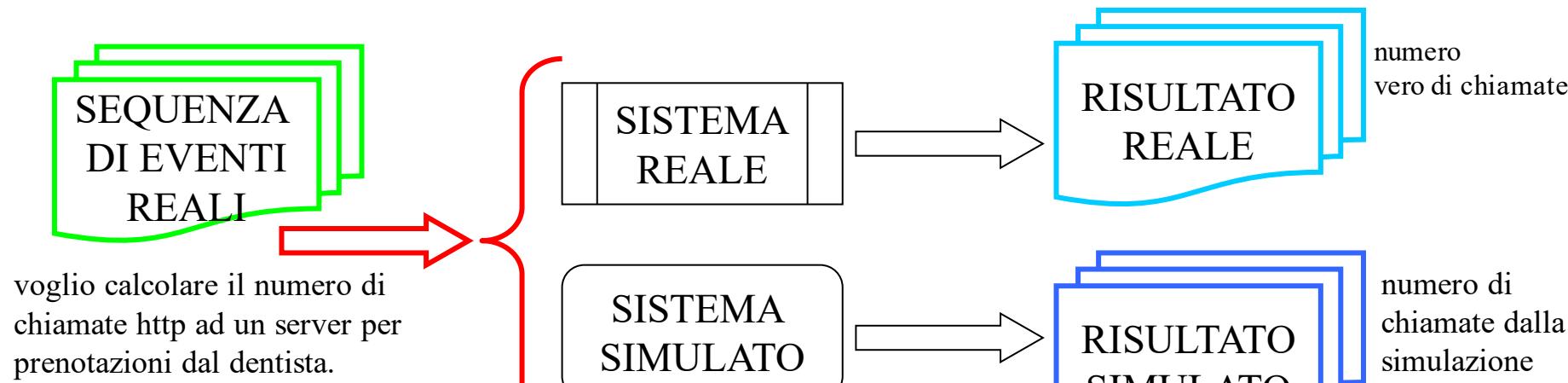
per ipotesi quella reale (che non si può conoscere)
distante non perché casuale



Fissato n, riducendo la probabilità dell'errore di I specie (α) si aumenta la probabilità dell'errore di II specie (β) e si riduce la potenza del test ($1-\beta$).

“*p-value*” di un test è il più piccolo livello di significatività a partire dal quale verrebbe rigettata l’ipotesi nulla

Esperimenti di confronto tra un P-C simulato e uno reale



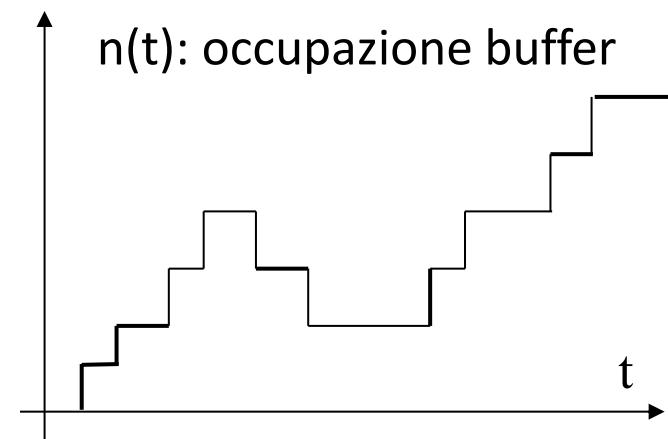
entrambi i risultati hanno sicuramente una media. ma tale media magari non nasce allo stesso modo. magari la simulazione ha tutte le chiamate fatte di sera (quando il dentista non lavora!) si capisce quindi che si necessita correlazione

coppia di "n" sequenze: ("n" ris reali (X), "n" ris simulati (Y))

$X_i, (Y_i) = \text{MEDIA DELLA TRAIETTORIA "i"}$
REALE, (SIMULATA)

$$X_i \doteq \frac{1}{T} \int_{t=0}^T n_{rea}(t) dt$$

$$Y_i \doteq \frac{1}{T} \int_{t=0}^T n_{sim}(t) dt$$



Il test per dati appaiati (PAIRED-T TEST)

EXTRA DA 28 IN SU

X_i e Y_i , ^{devono essere} correlate $\forall i \Rightarrow D_i \hat{=} X_i - Y_i, i = 1, \dots, n$

IPOTESI NULLA: sistema simulato riproduce fedelmente sistema reale
punto per punto non in media

Allora è lecito assumere che le $D_i, i=1, \dots, n$ siano tutte gaussiane identiche e indipendenti, di valore atteso nullo

TEST

$$\mu_D \hat{=} \mu_X - \mu_Y = 0 ?$$

STATISTICA

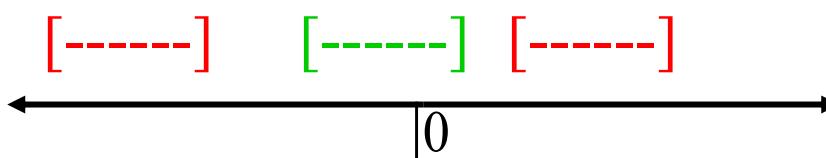
$$\frac{\bar{D}(n) - 0}{S / \sqrt{n}} \approx T_{n-1}$$

REGIONE DI ACCETTAZIONE

$$|T_{n-1}| \leq t_{n-1; 1-\alpha/2}$$

per cosa usarlo? intervallo di confidenza o test?

Osservazione_1: questo test corrisponde a I.C. al $100(1-\alpha)\%$:

$$0 \in \bar{D}(n) \pm t_{\alpha/2} (S / \sqrt{n}) ?$$


Osservazione_2: se si aumenta troppo la dimensione del campione, il test tenderà a rigettare l'ipotesi nulla!

Il “POOLED T-TEST” nell’ipotesi di varianza (reale) identica

$X_i, \quad i = 1, \dots, n_x$

indip. e

id. distr.

$Y_j, \quad j = 1, \dots, n_y$

$\bar{X}(n_x)$ e $\bar{Y}(n_y)$ gaussiane,
almeno con buona approx.

ESEMPIO: UN MODELLO P-C CON DISCIPLINA FIFO E L’ALTRO CON DISCIPLINA RANDOM (il servente preleva dal buffer a caso il cliente da servire)

TEST

$\bar{X}(n_x) = \bar{Y}(n_y)$?

STATISTICA del test

$$\frac{(\bar{X}(n_x) - \bar{Y}(n_y)) - 0}{S_p \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} \approx T_\gamma$$

*REGIONE DI
ACCETTAZIONE*
 $|T_{n-1}| \leq t_{n-1; 1-\alpha/2}$

$$S_p^2 \triangleq \frac{(n_x - 1)S_x^2 + (n_y - 1)S_y^2}{n_x + n_y - 2} ; \quad \gamma = n_x + n_y - 2$$

NON TROPPO
DIFFICILE DA
DIMOSTRARE!

$$S_p^2 \triangleq \frac{(n_x - 1)S_x^2 + (n_y - 1)S_y^2}{n_x + n_y - 2} ; \quad \gamma = n_x + n_y - 2$$

Prova della statistica del POOLED T-TEST:

$$(n_x - 1)S_x^2 / \sigma^2, \approx \chi_{n_x - 1}^2 ; (n_y - 1)S_y^2 / \sigma^2, \approx \chi_{n_y - 1}^2$$

$$\Rightarrow (n_x - 1)S_x^2 / \sigma^2 + (n_y - 1)S_y^2 / \sigma^2 =$$

$$= \frac{1}{\sigma^2} [(n_x - 1)S_x^2 + (n_y - 1)S_y^2] \approx \chi_{n_x + n_y - 2}^2$$

$$\bar{X}, \approx N(\mu_x, \sigma^2 / n_x) ; \bar{Y}, \approx N(\mu_y, \sigma^2 / n_y)$$

$$\Rightarrow \bar{X} - \bar{Y}, \approx N(\mu_x - \mu_y, \sigma^2 / n_x + \sigma^2 / n_y)$$

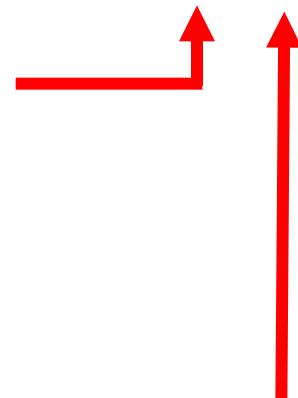
$$\Rightarrow \frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{\sigma^2 / n_x + \sigma^2 / n_y}} = \frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sigma \cdot \sqrt{1/n_x + 1/n_y}} \approx N(0, 1) \equiv Z$$

STATISTICA

$$\frac{(\bar{X} - \bar{Y}) - 0}{S_p \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} \approx T_\gamma$$

senza Gosset non si dimostrava Z

$$T_\gamma \triangleq \frac{Z}{\sqrt{\chi_\gamma^2 / \gamma}}$$



TEST SULLA DIFFERENZA FRA DUE MEDIE

CASO ASSAI PARTICOLARE: NORMALITA' e VARIANZA NOTA

$$\frac{(\bar{X}(n_x) - \bar{Y}(n_y)) - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}} \approx Z \quad |Z| \leq z_{1-\alpha/2}$$

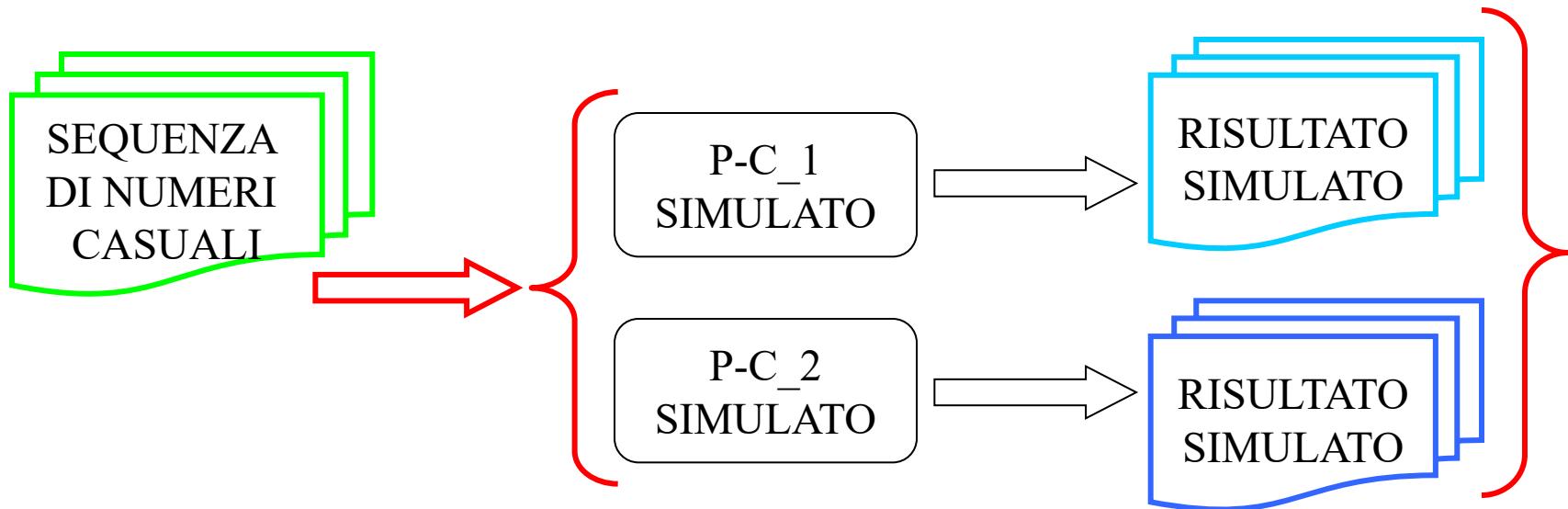
per gli esercizi dei professori

$$H_0: \mu_x = \mu_y$$

CASO GENERALE: NORMALITA' e VARIANZA INCOGNITA e diversa

$$\frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{\frac{S_x^2}{n_x} + \frac{S_y^2}{n_y}}} \approx T_\gamma \quad \gamma = \frac{\left(\frac{S_x^2}{n_x} + \frac{S_y^2}{n_y} \right)^2}{\left(\frac{S_x^2}{n_x} \right)^2 / (n_x + 1) + \left(\frac{S_y^2}{n_y} \right)^2 / (n_y + 1)} - 2$$
$$|T_\gamma| \leq t_{\gamma-1; 1-\alpha/2}$$

Un test su politiche di servizio alternative: “approximate T-test”



$$\frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{\frac{S_x^2}{n_x} + \frac{S_y^2}{n_y}}} \approx T_\gamma$$

$$|T_\gamma| \leq t_{\gamma; 1-\alpha/2}$$

ESEMPIO:
tempo medio di attesa secondo politica
“shortest job first” oppure “FIFO”

$$\gamma = \frac{\left(S_x^2 / n_x + S_y^2 / n_y \right)^2}{\left(S_x^2 / n_x \right)^2 / (n_x + 1) + \left(S_y^2 / n_y \right)^2 / (n_y + 1)} - 2$$

ABbastanza difficile da dimostrare!