Paolo Cremonesi

Impianti Informatici

POLITECNICO DI MILANO

Affidabilità
Cenni di statistica

- eventi indipendenti
- eventi mutuamente esclusivi
- probabilità condizionata
- teorema di Bayes
- distribuzione cumulativa
- densità di probabilità
- distribuzione esponenziale
- distribuzione di Weibull

Approccio probabilistico

- I meccanismi che determinano il comportamento di un impianto informatico sono complessi e dipendono da fattori diversi:
  - umani, tecnologici, ambientali, ...
- Le grandezze relative all'affidabilità e alle prestazioni di un sistema non posso essere descritte da leggi deterministiche
  - si è obbligati a considerare queste grandezze in termini statistici
  - le leggi che le governano sono leggi di tipo probabilistico

I meccanismi che causano guasti in un sistema sono molto complessi e dipendono da fattori diversi sia umani, che tecnologici, che ambientali.

•Le grandezze relative all'affidabilità di un sistema (ad esempio, il tempo di vita di un disco, o il numero di bug in un'applicazione) non posso essere descritte da leggi deterministiche.

Si è obbligati a considerare queste grandezze in termini statistici. Ad esempio, "esiste il 50% di probabilità che il disco si guasti entro 1000 giorni".

Inoltre, le leggi che governano queste grandezze sono leggi di tipo probabilistico. Ad esempio, "la probabilità che il server si guasti cresce linearmente con il numero di CPU".

Vedremo adesso alcune definizioni ed alcune proprietà elementari della statistica e del calcolo delle probabilità che ci potranno essere utili ai fini dello studio dell'affidabilità dei sistemi informativi.

Probabilità: eventi indipendenti

Indichiamo con E_1 e E_2 due eventi **indipendenti**

Indichiamo

- con **P**(E_1) la probabilità che si verifichi l'evento E_1
- con **P**(E_2) la probabilità che si verifichi l'evento E_2

La probabilità che i due eventi si verifichino **contemporaneamente** è data da

- $\mathbf{P}(E_1 \cap E_2) = \mathbf{P}(E_1) \cdot \mathbf{P}(E_2)$

Vediamo innanzitutto alcune proprietà del calcolo delle probabilità.

•Indichiamo con E_1 e E_2 due eventi **indipendenti**

•Indichiamo con **P**(E_1) la probabilità che si verifichi l'evento E_1 e con **P**(E_2) la probabilità che si verifichi l'evento E_2

•La probabilità che due eventi indipendenti si verifichino entrambi è pari al prodotto delle loro probabilità.

> Ad esempio, lancio una moneta e desidero calcolare la probabilità che esca testa due volte di seguito.
>
> La probabilità che esca testa una volta è pari al 50%.
>
> La probabilità che esca testa due volte di seguito è uguale alla probabilità che esca testa al primo lancio x probabilità che esca testa al 2° lancio, quindi è pari al 25%

Probabilità: eventi mutuamente esclusivi

Indichiamo con E_1 e E_2 due eventi **mutuamente esclusivi**

Indichiamo

- con $\mathbf{P}(E_1)$ la probabilità che si verifichi l'evento E_1
- con $\mathbf{P}(E_2)$ la probabilità che si verifichi l'evento E_2

La probabilità che si verifichi uno qualsiasi dei due eventi è data da

- $\mathbf{P}(E_1 \cup E_2) = \mathbf{P}(E_1) + \mathbf{P}(E_2)$

Indichiamo adesso con E_1 e E_2 due eventi **mutuamente esclusivi**

•Indichiamo sempre con $\mathbf{P}(E_1)$ la probabilità che si verifichi l'evento E_1 con $\mathbf{P}(E_2)$ la probabilità che si verifichi l'evento E_2

•La probabilità che si verifichi uno qualsiasi di due eventi mutuamente esclusivi è pari alla somma delle loro probabilità.

Ad esempio, lancio un dado e desidero conoscere la probabilità che esca 1 o 2. La probabilità che esca uno qualsiasi dei 6 numeri è pari ad 1/6. La probabilità che esca 1 o 2 è pari a 1/6+1/6=1/3.

Si noti che

- due eventi mutuamente esclusivi non sono mai indipendenti

- due eventi indipendenti, non sono mai mutuamente esclusivi

Probabilità condizionata e teorema di Bayes

Indichiamo con E_1 e E_2 due eventi **qualsiasi**

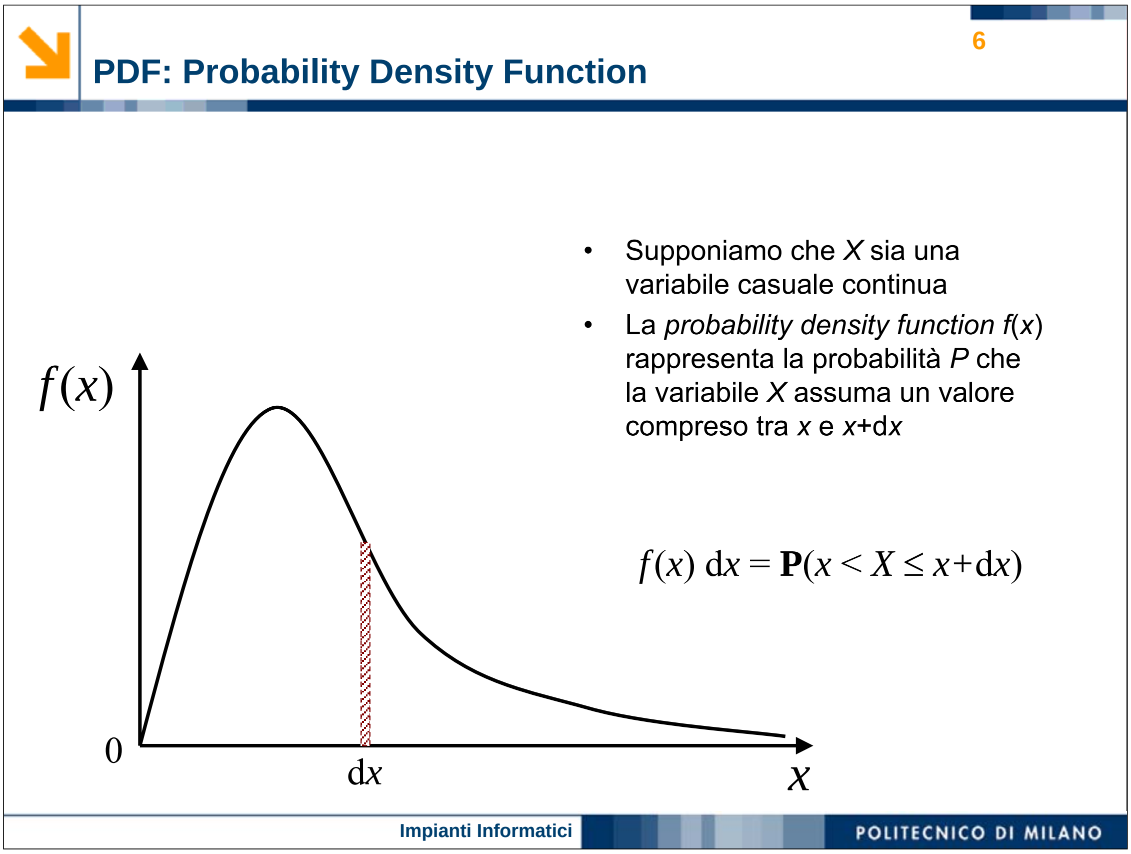
- La probabilità che si verifichi l'evento E_1 nel caso in cui si sia verificato l'evento E_2 è data da
  - $\mathbf{P}(E_1|E_2) = \mathbf{P}(E_1 \cap E_2) / \mathbf{P}(E_2)$
- Si noti che **se due eventi sono indipendenti**
  - $\mathbf{P}(E_1 \cap E_2) = \mathbf{P}(E_1) \cdot \mathbf{P}(E_2)$
- quindi
  - $\mathbf{P}(E_1|E_2) = \mathbf{P}(E_1)$

Indichiamo adesso con E_1 e E_2 due eventi **qualsiasi**

•La probabilità che si verifichi l'evento E_1 nel caso in cui si sia verificato l'evento E_2 è data da

$$\mathbf{P}(E_1|E_2) = \mathbf{P}(E_1 \cap E_2) / \mathbf{P}(E_2)$$

•Due eventi sono indipendenti quando la probabilità del primo condizionata al secondo è uguale alla probabilità del primo non condizionata (e viceversa).



Iniziamo dalla definizione di *densità di probabilità*.

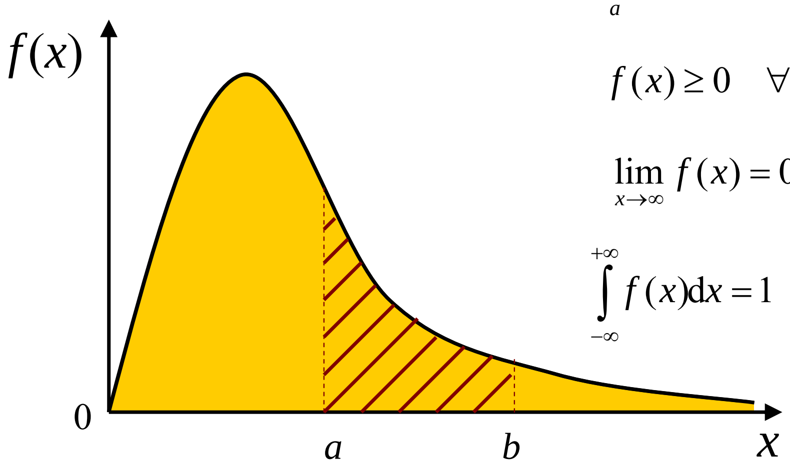
•Supponiamo che X sia una variabile casuale continua, ossia una variabile che può assumere un qualsiasi valore reale.

> Ad esempio, la variabile può rappresentare il numero che esce lanciando un dando.
>
> Facendo un esempio più vicino all'affidabilità, la variabile può rappresentare l'istante di tempo in cui si guasta un disco.

•La *probability density function* $f(x)$ X ci permette di calcolare la probabilità P che la variabile X assuma un valore compreso tra x e $x+\mathrm{d}x$.

•Questa probabilità è data dall'area sottesa dalla funzione $f(x)$ nell'intervallo compreso tra x e $x+\mathrm{d}x$

Proprietà della PDF



$$\int_a^b f(x)\mathrm{d}x = \mathrm{P}(a < X \le b)$$

$$f(x) \ge 0 \quad \forall x$$

$$\lim_{x\to\infty} f(x) = 0$$

$$\int_{-\infty}^{+\infty} f(x)\mathrm{d}x = 1 \qquad \int_{0}^{+\infty} f(t)\mathrm{d}t = 1$$

La *probability density function*, detta anche PDF, ha alcune proprietà che è importante sottolineare.

•La prima proprietà della PDF stabilisce che l'area sottesa tra due punti *a* e *b* rappresenta la probabilità che la variabile casuale *X* assuma un valore compreso tra *a* e *b*.

•La seconda proprietà ci dice che la PDF è sempre positiva.

•La terza proprietà afferma che la probabilità di avere un valore infinitamente grande è zero. In altre parole, la variabile casuale *X* assumerà sempre un valore finito.

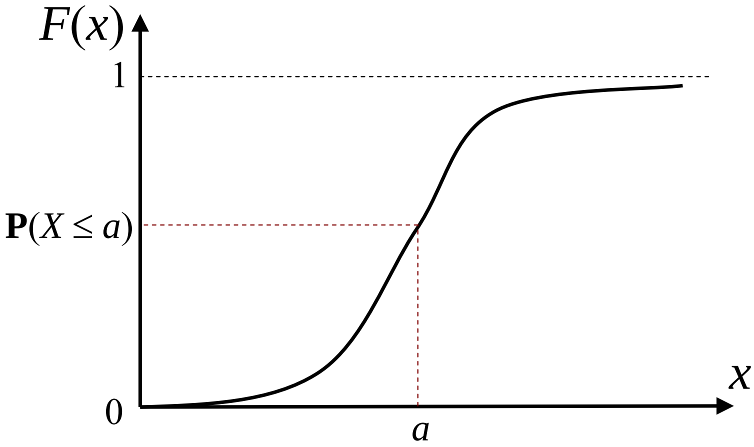
•La quarta proprietà ci dice che l'area sottesa da tutta la curva vale uno. In altre parole, la variabile assume sicuramente **un** valore.

•Nel caso in cui la variabile casuale rappresenti il tempo, come avviene spesso nei problemi legati all'affidabilità, tutte le proprietà indicate per $x \to -\infty$ valgono per x=0.

CDF: Cumulative Distribution Function

$$F(x) \equiv \mathbf{P}(X \leq x)$$

- La *cumulative distribution function* $F(x)$ rappresenta la probabilità P che la variabile X assuma un valore minore o uguale a x



$$f(x) = \frac{\mathrm{d}F(x)}{\mathrm{d}x}$$

$$\int_0^x f(t)\mathrm{d}t = F(x)$$

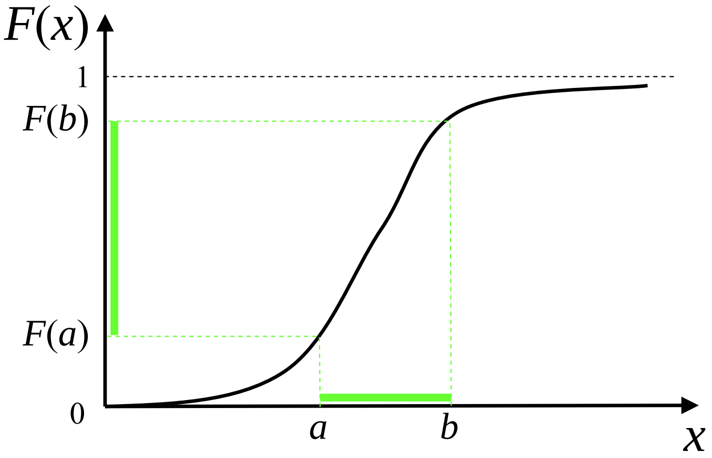
La *cumulative distribution function*, detta anche CDF, $F(x)$ rappresenta la probabilità P che la variabile X assuma un valore minore o uguale a x.

•In base alle definizioni, PDF e CDF sono legate tra di loro, nel senso che la PDF è la derivata rispetto ad x della CDF.

•Anche in questo caso, se la variabile casuale X rappresenta una grandezza positiva, la CDF viene definita per x maggiore o uguale a 0.

Proprietà della CDF

$$F(b) - F(a) = \mathrm{P}(a < X \leq b)$$



$$\lim_{x \to -\infty} F(x) = 0$$

$$\lim_{x \to +\infty} F(x) = 1$$

Anche la CDF ha delle proprietà che è importante sottolineare.

•La prima proprietà permette di calcolare la probabilità che la variabile casuale *X* assuma un valore compreso tra *a* e *b*, come differenza tra *F(b)* ed *F(a)*.

•Le successive due proprietà affermano che la variabile *X* assume un valore finito.

Nel caso in cui la variabile casuale sia sempre positiva, tutte le proprietà indicate per $x \to -\infty$ valgono per $x=0$. Ad esempio F(0)=0.

Media e Percentile

- Il valore medio μ è calcolato come media della variabile *x* pesata con la probabilità f

$$\mu = \mathrm{E}[X] = \int_{-\infty}^{-\infty} x f(x)\mathrm{d}x$$

- Il percentile μ_p è il valore assunto dalla variabile *x* al di sotto del quale cade una determinata percentuale *p* della distribuzione

$$F(\mu_p) = p$$

Vediamo adesso come si calcolano media e percentile di una distribuzione.

•Il valore medio μ della variabile casuale *X* è chiamato spesso valore atteso di *X* ed è indicata con E[*X*]. Questo valore atteso è calcolato come media pesata di tutti i possibili valori della variabile *X*, dove il peso è dato dalla CDF.

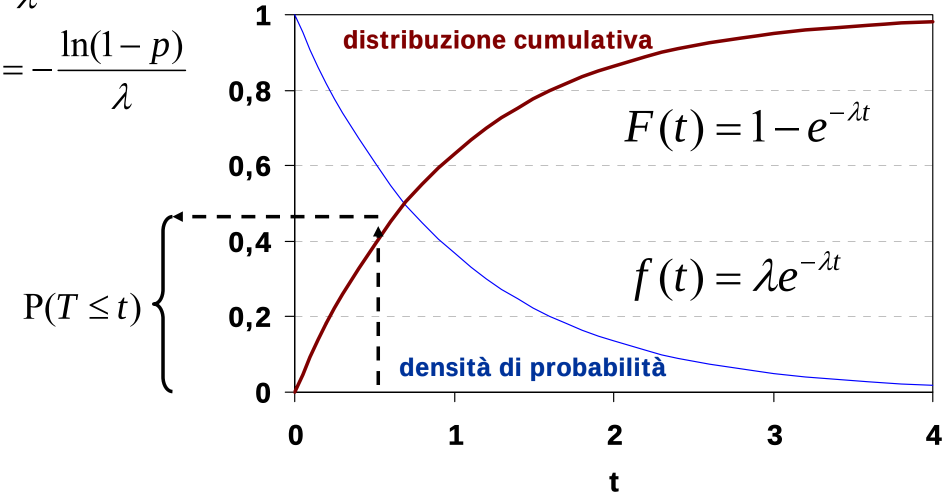
•Il *percentile* μ_p indica invece quel particolare valore assunto dalla variabile *x* al di sotto del quale cade una certa percentuale *p* della distribuzione.

Ad esempio, se la variabile casuale rappresenta la vita di un server, il 75-esimo percentile della variabile indica dopo quanto tempo avrò il 75% di probabilità che il server si sia guastato.

Distribuzione esponenziale

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

$$E_p[t] = -\frac{\ln(1-p)}{\lambda}$$



Vediamo adesso alcune distribuzioni statistiche interessanti per il calcolo dell'affidabilità.

•Partiamo dalla distribuzione esponenziale. Questa distribuzione è definita a meno di un parametro λ (il parametro della distribuzione) che è il reciproco del valor medio. Tra le tante proprietà della distribuzione esponenziale sicuramente la più interessante è quella della mancanza di memoria

Proprietà memoryless dell'esponenziale

Il futuro non dipende dal passato

$$P(T > t + \Delta t \mid T > t) = \frac{P(T > t + \Delta t \cap T > t)}{P(T > t)} =$$

$$= \frac{P(T > t + \Delta t)}{P(T > t)} = \frac{1 - F(t + \Delta t)}{1 - F(t)} =$$

$$= \frac{e^{-\lambda(t+\Delta t)}}{e^{-\lambda t}} = e^{-\lambda \Delta t} = 1 - F(\Delta t) = P(T > \Delta t)$$

Un oggetto che segue una distribuzione esponenziale non ha memoria di quanto tempo ha funzionato, cioè non è soggetto a invecchiamento

In base a questa proprietà, un fenomeno che segue la distribuzione esponenziale è caratterizzato dal fatto che il futuro non dipende dal passato.

•Vediamo di darne una dimostrazione. Facciamo prima un esempio pratico. Arrivo alla fermata dell'autobus e aspetto *t* minuti. Mi chiedo a questo punto quale è la probabilità che l'autobus passi entro Δt minuti. Più formalmente, vogliamo calcolare la probabilità che un certo evento *T* accada entro $t+\Delta t$ unità di tempo, probabilità condizionata dal fatto che sono già trascorsi *t* unità di tempo.

•Applicando la definizione di proprietà condizionata, si vede come questa probabilità sia uguale alla probabilità di aspettare un autobus per Δt minuti, indipendentemente dal tempo già trascorso alla fermata.

In altre parole, arrivo alla fermata dell'autobus e mi chiedo: "in media quanto aspetterò prima che passi un autobus?" Supponiamo che la risposta sia "10 minuti". Trascorsi 5 minuti mi richiedo: "In media quanto aspetterò ancora?". La risposta sarà sempre "10 minuti".

Distribuzione di Weibull a tre parametri

13

Generalizzazione dell'esponenziale

$$f(t) = \frac{\beta}{\alpha}\left(\frac{t-\gamma}{\alpha}\right)^{\beta-1} e^{-\left(\frac{t-\gamma}{\alpha}\right)^{\beta}} \qquad t \geq \gamma$$

$$F(t) = 1 - e^{-\left(\frac{t-\gamma}{\alpha}\right)^{\beta}}$$

- α è detto parametro di scala o vita caratteristica
- β è detto parametro di forma
- γ è detto vita minima

Impianti Informatici POLITECNICO DI MILANO

Passiamo adesso ad analizzare le caratteristiche e le proprietà di un'altra distribuzione molto importante per lo studio dei fenomeni legati all'affidabilità, la distribuzione di Weibull.

•La distribuzione di Weibull è una generalizzazione della distribuzione esponenziale. Sia la DPF che la CDF dipendono tra tre parametri positivi

α è detto parametro di scala o vita caratteristica e corrisponde al reciproco del parametro λ dell'esponenziale

β è detto parametro di forma e la sua funzione sarà chiarita nel lucido successivo

γ è detto vita minima e permette di traslare la distribuzione lungo l'asse delle ascisse

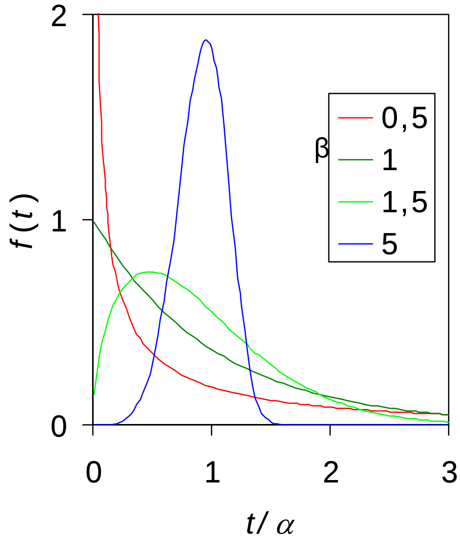
Si noti che quando β è uguale ad uno e γ è uguale a zero, la distribuzione di Weibull si riconduce ad una esponenziale con valor medio pari ad α

Distribuzione di Weibull (2)

$$E[t] = \gamma + \alpha\,\Gamma\left(\frac{1}{\beta} + 1\right)$$

$$E_p[t] = \gamma + \alpha[-\ln(1-p)]^{1/\beta}$$

$$\Gamma(x) = \int_0^{+\infty} e^{-t} t^{x-1} \mathrm{d}t$$



Anche per la distribuzione di Weibull è possibile calcolare media e percentili, anche se il calcolo della media non può essere svolto in modo esplicito, dato che contiene al suo interno la funzione Gamma il cui integrale non è risolvibile in forma analitica.

•E' interessante studiare il comportamento della distribuzione di Weibull in funzione del parametro β. Il grafico la PDF nel caso in cui γ sia pari a zero.

•Per β<1 la distribuzione è sempre decrescente ed è molto schiacciata contro gli assi cartesiani

•Per β=1 la distribuzione si comporta come un esponenziale

•Per β>1 la distribuzione presenta una gobba

•e la gobba si accentua all'aumentare di β.

L'andamento della distribuzione di Weibull in funzione del parametro β è molto importante perché, come vedremo in seguito, rispecchia le diversi fasi della vita di un componente informatico.