

Funzioni di variabili aleatorie

Abbiamo visto durante il corso le distribuzioni: esponenziale, di Weibull e Gaussiana. I matematici con il concetto di funzione di funzione, sono riusciti a costruire dei modelli diversi. Ma che vuol dire funzione di variabili aleatorie?

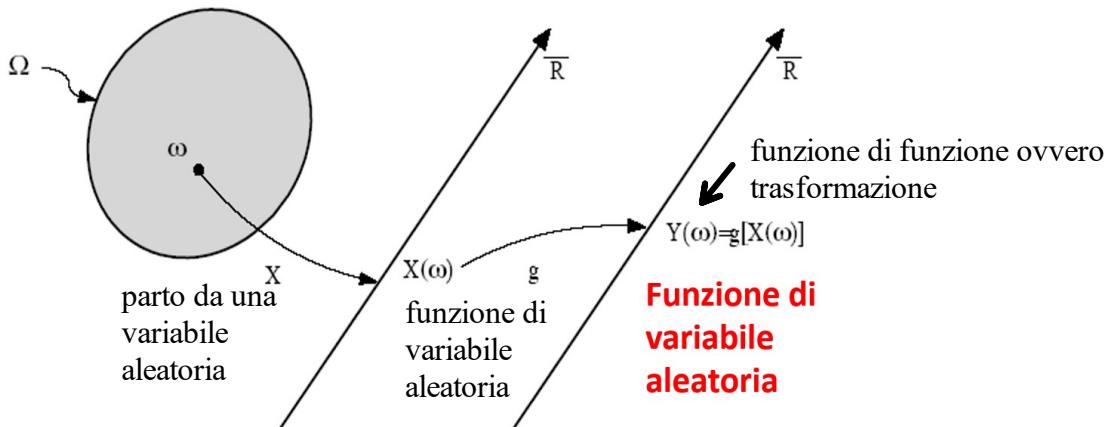
All'interno della letteratura scientifica si usa il termine trasformazioni di una variabile aleatoria. Nel nostro corso utilizzeremo il termine funzione al posto di trasformazione. Una trasformazione consiste in un'operazione algebrica di variabili aleatorie

Dal libro del prof. Gelli, "Probabilità e informazione"

86

Trasformazioni di una variabile aleatoria

"Trasformazione" = "funzione"



Il caso più facile è quello in cui g è funzione differenziabile e strettamente monotona (invertibile su tutti il dominio)

Il caso più difficile è quello in cui g è funzione differenziabile con punti a derivata nulla e interi intervalli a derivata nulla!

A ω risultato elementare, viene associato un numero e dopo tramite g se ne associa un altro. La domanda che nasce è può cambiare la probabilità di ω ?

La risposta è no, non cambia semplicemente la si rappresenta diversamente. Vediamo ciò tramite la proposizione di Miller seguente:

4.6 Transformations of Random Variables

$$F_Y(y) = \Pr(g(X) \leq y) = \Pr(X \leq g^{-1}(y)) = F_X(g^{-1}(y)). \quad (4.20)$$

probabilità che Y sia

minore o uguale ad un valore y

Note that this can also be written as

probabilità che X sia

minore o uguale all'inverso

della trasformazione applicata

a y

$$F_X(x) = F_Y(g(x)). \text{ "Si cambia la forma
ma non la sostanza"}$$

Differenziale?

Si ma non detto
esplicitamente

(4.21)

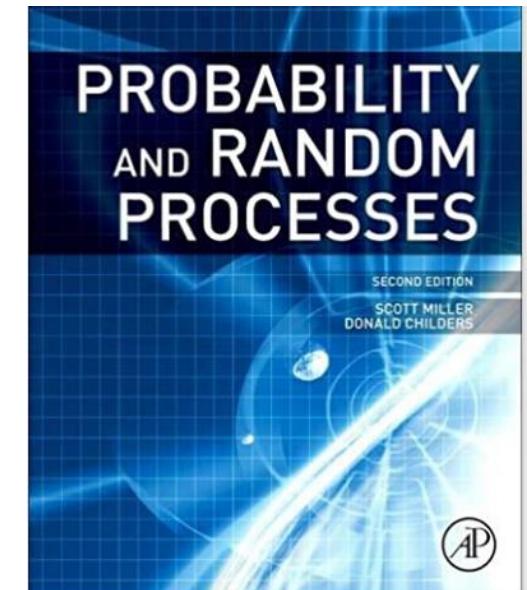
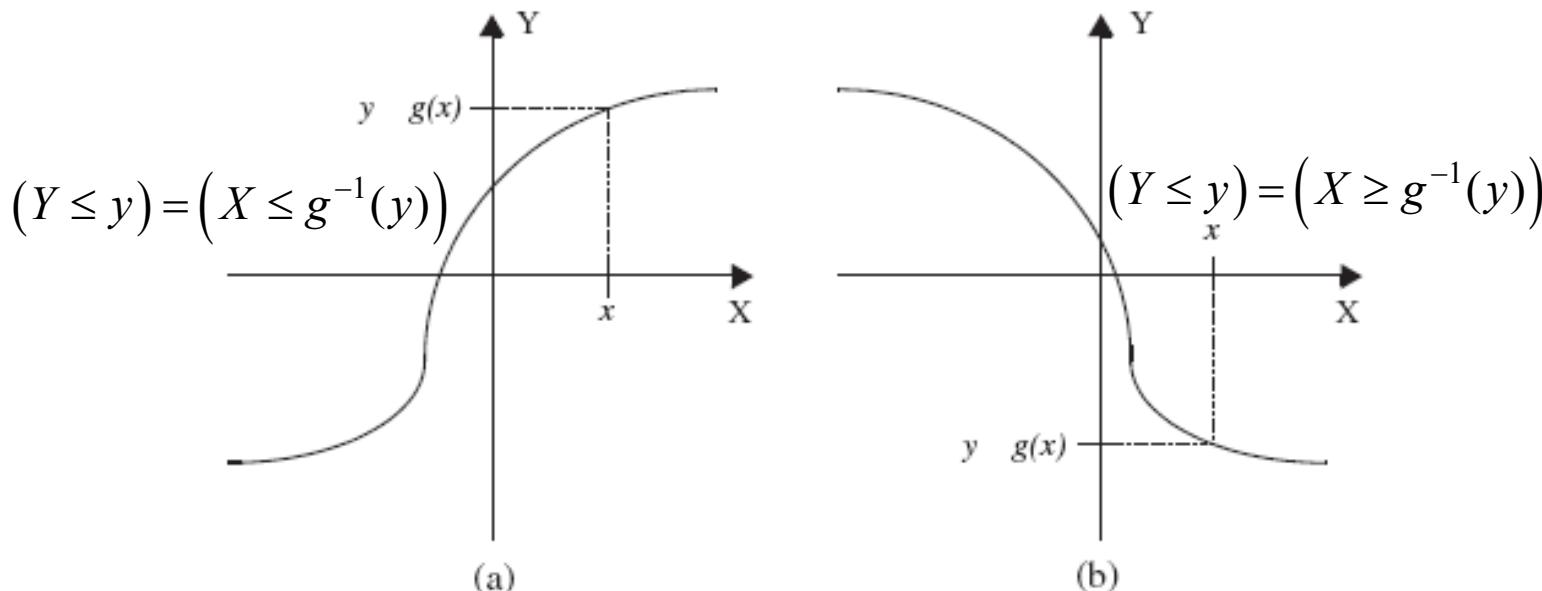


Figure 4.1 A monotonic increasing function (a) and a monotonic decreasing function (b).

Per funzioni **monotone crescenti** ottengo la densità di una variabile trasformata con le regole di derivazione delle funzioni composte

Differentiating Equation 4.20 with respect to y produces

densità di probabilità
della variabile trasformata Y

$$f_Y(y) = f_X(g^{-1}(y)) \frac{dg^{-1}(y)}{dy} = f_X(x) \frac{dx}{dy} \Big|_{x=g^{-1}(y)},$$

La densità di probabilità di Y in un punto y
è pari alla densità di X in un punto x
(4.22)

while differentiating Equation 4.21 with respect to x gives

Da questa slide si capisce che
la funzione è differenziabile!

$$f_X(x) = f_Y(g(x)) \frac{dy}{dx} \Rightarrow f_Y(y) = \frac{f_X(x)}{\frac{dy}{dx}} \Big|_{x=g^{-1}(y)} \quad \leftarrow \text{come ottenere
la densità di Y
a partire dalla
densità di X}$$
(4.23)

Per funzioni **monotone decrescenti** si parte dalla distribuzione

La direzione della diseguaglianza
nella probabilità arancio a differenza
della funzione crescente si inverte
per la natura decrescente della
trasformazione

Differentiating with respect to y gives

$$f_Y(y) = -f_X(x) \frac{dx}{dy} \Big|_{x=g^{-1}(y)} \quad \begin{array}{l} \text{usa la complementare} \\ \text{segno negativo perché la} \\ \text{funzione è decrescente!} \end{array} \quad (4.25)$$

(4.24) Spiegazione evidenziato
arancione:
la probabilità che Y sia $\leq y$
corrisponde alla probabilità che
X è $\geq g^{-1}(y)$ perché la funzione
è decrescente, quindi i valori
più grandi di y corrispondono a
valori più piccoli di x

Similarly, writing $F_Y(g(x)) = 1 - F_X(x)$ and differentiating with respect to x results in

$$f_Y(y) = -\frac{f_X(x)}{\frac{dy}{dx}} \Big|_{x=g^{-1}(y)} . \quad (4.26)$$

Dal libro del prof. Gelli, "Probabilità e informazione"

► Esempio 4.7. Consideriamo nuovamente la trasformazione lineare

$$Y = aX + b, \quad \text{ricordando che } X \text{ è la variabile originale e } Y \text{ è la variabile ricavata}$$

Qualunque sia $y \in \mathbb{R}$, e per ogni $a \neq 0$, l'equazione $y = g(x) = ax + b$ ammette l'unica soluzione

$$X = \frac{y - b}{a}, \quad \text{ricavata tramite differenziabilità}$$

ed inoltre risulta

$$|g'(x)| = |a|, \quad \text{per cui: } f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y - b}{a}\right) \quad \text{permette di ricavare la non standard a partire dalla standard}$$

Si può procedere anche diversamente, ottenendo la distribuzione prima e derivando poi:

Nel caso $a > 0$, si ha:

$$F_Y(y) = P(Y \leq y) = P(aX + b \leq y) = P\left(X \leq \frac{y - b}{a}\right) = F_X\left(\frac{y - b}{a}\right). \quad \text{usando la 4.22 per funzioni monotone crescenti}$$

Per $a < 0$, il verso della diseguaglianza si inverte,

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(aX + b \leq y) = P\left(X \geq \frac{y - b}{a}\right) \\ &= 1 - P\left(X < \frac{y - b}{a}\right) = 1 - F_X\left(\frac{y - b}{a}\right). \end{aligned}$$

la probabilità che Y sia \leq di un valore y corrisponde a calcolare la probabilità che X sia maggiore uguale al valore x

otteniamo lo stesso risultato per $a > 0$ e $a < 0$

$$\frac{d}{dy} F_Y(y) = \frac{1}{|a|} \cdot f_X\left(\frac{y - b}{a}\right)$$

Giustifico la forma della densità Gaussiana non standard a partire dalla standard

Valore atteso della funzione di una variabile aleatoria (non negativa)

ricorda $g(x) = y = ax + b$

Il valore atteso di Y $E[Y] \hat{=} \int_{y=0}^{\infty} y \cdot f_Y(y) dy$, con $Y \hat{=} a \cdot X$, $a > 0$ (esempio)
quindi $b=0$

$$E[aX] \hat{=} \int_{ax=0}^{\infty} ax \cdot f(ax) d(ax)$$

utilizzando $\frac{d}{dy} F_Y(y) = \frac{1}{|a|} \cdot f_X\left(\frac{y-b}{a}\right)$ con $b=0$

$$= \int_{x: ax=0}^{\infty} ax \cdot (1/a) f(y/a) \cdot d(ax) = \int_{x: ax=0}^{\infty} ax \cdot (1/a) f(ax/a) \cdot adx$$

$$= \int_{x=0}^{\infty} ax \cdot f(x) \cdot dx$$

Generalizzazione, con $g(x)$ idonea:

per idonea si intende monotona non decrescente e differenziabile

$$E[g(X)] \hat{=} \int_{x: g(x)=0}^{\infty} g(x) \cdot f(x) dx, \quad g(x) \geq 0$$

è esattamente la definizione del valore atteso di X per la costante a

Theorem: pdf for a transformed RV

Sia X : variabile aleatoria continua con densità f_X che non è 0 su un sottoinsieme I di numeri reali [i.e., $f_X(x) > 0, x \in I$ and $f_X(x) = 0, x \notin I$]. I può essere un punto!

Sia g : funzione monotona differenziabile con dominio I e immagine l'insieme dei reali.

Allora $Y = g(X)$: variabile aleatoria continua con densità f_Y definita come :

Densità di X valutata rispetto alla trasformazione inversa g nel punto y

$$f_Y(y) = \begin{cases} f_X[g^{-1}(y)][|(g^{-1})'(y)|], & y \in g(I) \\ 0, & \text{altrimenti} \end{cases}$$

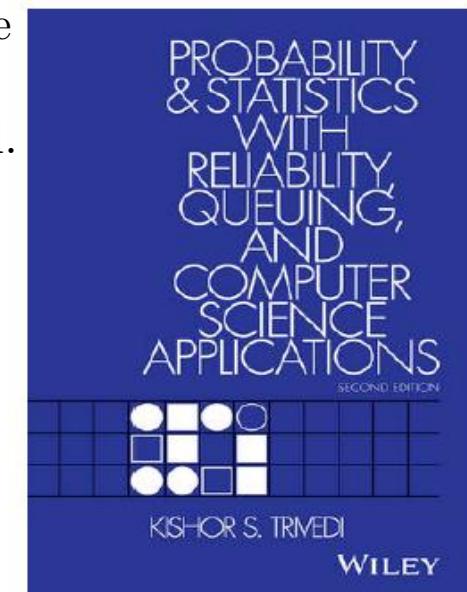
la derivata è la pendenza della funzione. Il valore assoluto mi dice di quanto sto "stirando" o "comprimendo" la distribuzione di probabilità

Prova:

Derivando ed utilizzando la regola di derivazione delle funzioni composte si ottiene che

$$F_Y(y) = P(Y \leq y) = P[g(X) \leq y] = P[X \leq g^{-1}(y)] = F_X[g^{-1}(y)]$$

Conosco la distribuzione di X e
 la funzione che trasforma X in Y ,
 quindi calcolo la distribuzione in Y



Consideriamo adesso la funzione quadratica: (che è monotona crescente strettamente)

Example 3.8

<https://ece.duke.edu/faculty/kishor-trivedi>

Distribuzione per $Y = g(X) = X^2$ Ricorda che
 $F_Y(y) = P(Y \leq y)$

$F_Y(y) = 0$, per $y \leq 0$ $P(Y \leq y)$ è zero! Essendo $Y = X^2 \implies X^2 \leq y$ ovvero $X^2 \leq$
di una quantità \leq di zero si risolve appunto solo
invece nel caso $y > 0$: per $X=0$

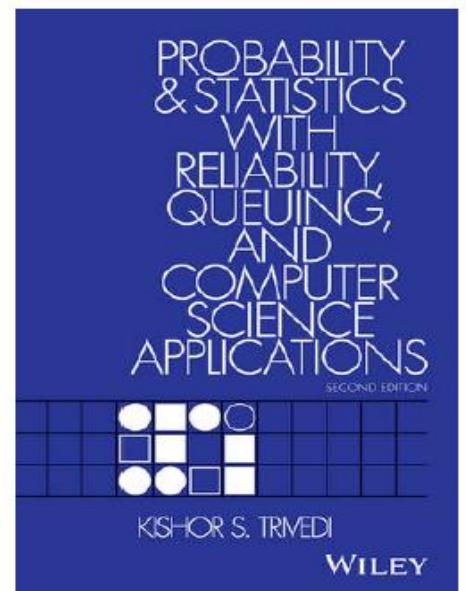
$$F_Y(y) = P(Y \leq y) = P(X^2 \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y})$$

Esprimo $X \in [-\sqrt{y}, \sqrt{y}]$ come differenza esplicitando la definizione di distribuzione

La densità di Y , f_Y si ottiene con la regola di derivazione per funzioni composte,

$$f_Y(y) = \frac{dF_Y(y)}{dy} = \begin{cases} \frac{1}{2\sqrt{y}} [f_X(\sqrt{y}) + f_X(-\sqrt{y})], & y > 0, \\ 0, & \text{altrimenti.} \end{cases}$$

Quando elevo al quadrato una variabile aleatoria sto "comprimendo" la parte negativa della distribuzione e sto "stirando" quella positiva



Per chiarezza, vi anticipo la densità della V.A. gamma, che sarà dimostrata più avanti nel corso come generalizzazione del modello di Erlang:

https://en.wikipedia.org/wiki/Agner_Krarup_Erlang

Variabile aleatoria Gamma

A partire dal secondo integrale di Eulero si definisce la funzione
portando fuori $\Gamma(\alpha)$ di densità Gamma
perchè costante
ottengo $\frac{1}{\Gamma(\alpha)} \cdot \Gamma(\alpha) = 1$

$$f(t) = \frac{\lambda^\alpha t^{\alpha-1} e^{-\lambda t}}{\Gamma(\alpha)}, \quad \alpha > 0, \quad t > 0$$

$$\int_0^\infty \frac{\hat{f}(t)}{\Gamma(\alpha)} dt = 1$$

appunto l'integrale
di una densità deve
essere pari a 1

dove

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx, \quad \alpha > 0$$

- Show the recurrence for the gamma function:
 $\Gamma(\alpha) = (\alpha-1) \Gamma(\alpha-1)$; and show that $\Gamma(1/2) = \sqrt{\pi}$
- Because $\Gamma(1) = 1$, it follows that for an integer r , $\Gamma(r) = (r-1) \Gamma(r-1) = \dots = (r-1)!$
- So gamma with a positive integer valued shape parameter is the Erlang random variable

La variabile aleatoria Gamma con parametro di forma $\alpha = \frac{n}{2}$,

parametro di scala $\lambda = \frac{1}{2}$ è nota come variabile aleatoria "chi-square", (chi-quadrato) con n gradi di libertà

Se integro $\hat{f}(t)$

ottengo

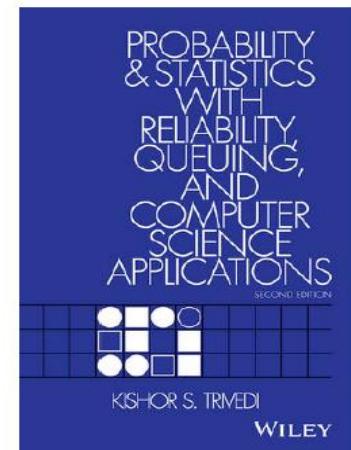
$$\Gamma(\alpha) = \int_{t=0}^{\infty} (\lambda t)^{\alpha-1} \cdot e^{-\lambda t} d(\lambda t), \quad \lambda > 0$$

sostituendo con

$$x \triangleq \lambda t \Rightarrow dx = \lambda dt$$

$$\Rightarrow x^{\alpha-1} \cdot dx = \lambda \cdot (\lambda t)^{\alpha-1}$$

Al crescere di alfa,
la distribuzione è
più simmetrica a destra
Al crescere di lambda
la distribuzione è
concentrata attorno a 0



IMPORTANTE
NOTIZIA

Caso particolare famoso (della funzione quadratica):

Quadrato della Gaussiana standard $N(0,1)$ Come calcoliamo le caratteristiche di densità e distribuzione del quadrato? Applicando il risultato del esempio 3.8 evidenziato in verde!

Example 3.9

In Example 3.8, assume X to be $N(0,1)$:

quindi si sa che la densità è $f_X(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$, $-\infty < x < \infty$.

- Using result from Example 3.8:

$$f_Y(y) = \begin{cases} \frac{1}{2\sqrt{y}} \left(\frac{1}{\sqrt{2\pi}}e^{-y/2} + \frac{1}{\sqrt{2\pi}}e^{-y/2} \right), & y > 0, \\ 0, & y \leq 0, \end{cases}$$

or, $f_Y(y) = \begin{cases} \frac{1}{\sqrt{2\pi y}}e^{-y/2}, & y > 0, \\ 0, & y \leq 0. \end{cases}$

i.e.,

Y has a *gamma* distribution with $\alpha = 1/2$ and $\lambda = 1/2$ e $y \equiv t$

- Which is also known as chi-square distribution with 1 degree of freedom

$$\Gamma(\alpha) = \int_{t=0}^{\infty} (\lambda t)^{\alpha-1} \cdot e^{-\lambda t} d(\lambda t), \lambda > 0$$

Passaggi algebrici

$$\begin{aligned} f(y) &= \frac{1}{\Gamma(\frac{1}{2})} \cdot \frac{1}{2} \cdot \left(\frac{y}{2}\right)^{-\frac{1}{2}} \cdot e^{-\frac{y}{2}} \\ &= \frac{1}{\Gamma(\frac{1}{2})} \cdot \frac{1}{\sqrt{2} \cdot \sqrt{2}} \cdot \frac{\sqrt{2}}{\sqrt{y}} \cdot e^{-\frac{y}{2}} \\ &= \frac{1}{\sqrt{\pi}} \cdot \frac{1}{\sqrt{2}} \cdot \frac{1}{\sqrt{y}} \cdot e^{-\frac{y}{2}} \end{aligned}$$

Il quadrato di una standard porta ad una funzione gamma particolare con alfa e lambda particolari detta chi-quadrato che servirà per la statistica della varianza

Una funzione logaritmo (ad hoc) su una var. al. X, UNIFORME in (0,1):

FINO A QUI DOMANDE DI ESAME!

Logaritmo è sempre
monotona e differenziabile come
vuole l'ipotesi ipotesi (slide 1)

$$Y \doteq -\lambda^{-1} \cdot \ln(1-X)$$

X distribuito uniformemente da 0 a 1 è un
generatore di numeri casuali

Example 3.10

- Let X be uniformly distributed, $\text{Unif}(0,1)$
- Then, $Y = -\lambda^{-1} \ln(1-X)$ is $\text{EXP}(\lambda)$.

$$\text{for } y \leq 0, F_Y(y) = 0$$

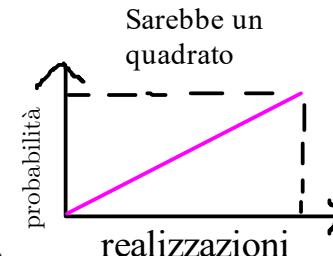
for $y > 0$,

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P[-\lambda^{-1} \ln(1-X) \leq y] \\ &= P[\ln(1-X) \geq -\lambda y] \\ &= P[(1-X) \geq e^{-\lambda y}] \quad (\text{since } e^x \text{ is an increasing function of } x,) \\ &= P(X \leq 1 - e^{-\lambda y}) \\ &= F_X(1 - e^{-\lambda y}). \end{aligned}$$

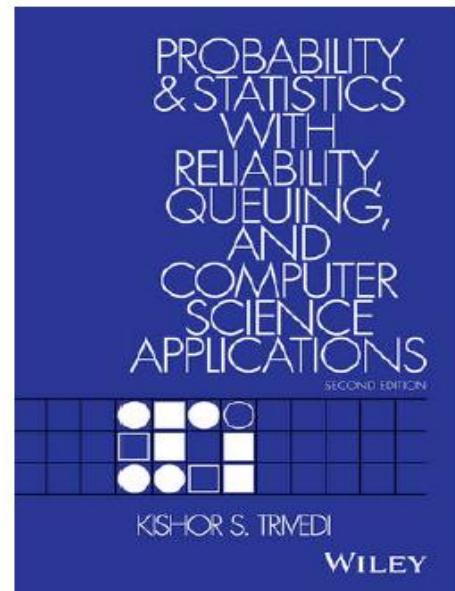
Se X è uniforme distribuito tra 0 e 1
la densità sarà costante dunque
la distribuzione che è la funzione l'integrale
della densità è lineare. Appunto integrale di una
costante funzione lineare da analisi 1

Since X is $\text{U}(0,1)$, $F_X(x) = x, 0 \leq x \leq 1$. Therefore,

$$F_Y(y) = 1 - e^{-\lambda y} \Rightarrow Y \text{ is } \text{EXP}(\lambda)$$



la realizzazione è
uguale alla probabilità
con cui si osserva la
realizzazione. Quindi
la distribuzione è la
diagonale del quadrato
vedi Monte Carlo



IN PARTICOLARE
La funzione (inversa) quantile è la
trasformazione che permette di
produrre realizzazioni a partire da
una funzione di distribuzione
invertibile (Esponenziale, Weibull ecc)

IDEA: avendo a disposizione un generatore di numeri casuali uniformemente distribuiti in (0,1) potremmo generare realizzazioni della V.A. esponenziale di parametro λ sarà il laboratorio di Excel

Example 8.1.2 (Log-Normal PDF). Let $X \sim \mathcal{N}(0, 1)$, $Y = e^X$. In Chapter 6 we named the distribution of Y the Log-Normal, and we found all of its moments using the MGF of the Normal distribution. Now we can use the change of variables formula to find the PDF of Y , since $g(x) = e^x$ is strictly increasing. Let $y = e^x$, so $x = \log y$ and $dy/dx = e^x$. Then

$$f_Y(y) = f_X(x) \left| \frac{dx}{dy} \right| = \varphi(x) \frac{1}{e^x} = \varphi(\log y) \frac{1}{y}, \quad y > 0.$$

Note that after applying the change of variables formula, we write everything on the right-hand side in terms of y , and we specify the support of the distribution. To determine the support, we just observe that as x ranges from $-\infty$ to ∞ , e^x ranges from 0 to ∞ .

We can get the same result by working from the definition of the CDF, translating the event $Y \leq y$ into an equivalent event involving X . For $y > 0$,

$$F_Y(y) = P(Y \leq y) = P(e^X \leq y) = P(X \leq \log y) = \Phi(\log y),$$

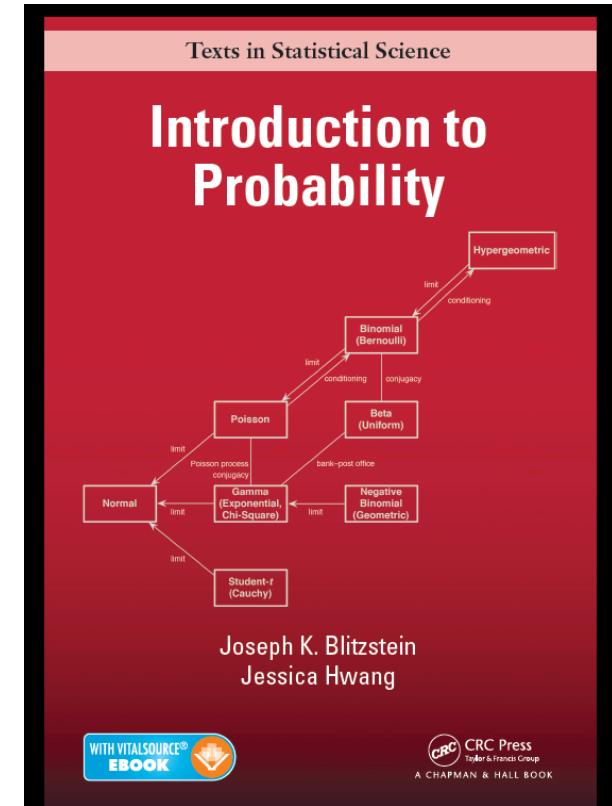
so the PDF is again

$$f_Y(y) = \frac{d}{dy} \Phi(\log y) = \varphi(\log y) \frac{1}{y}, \quad y > 0.$$

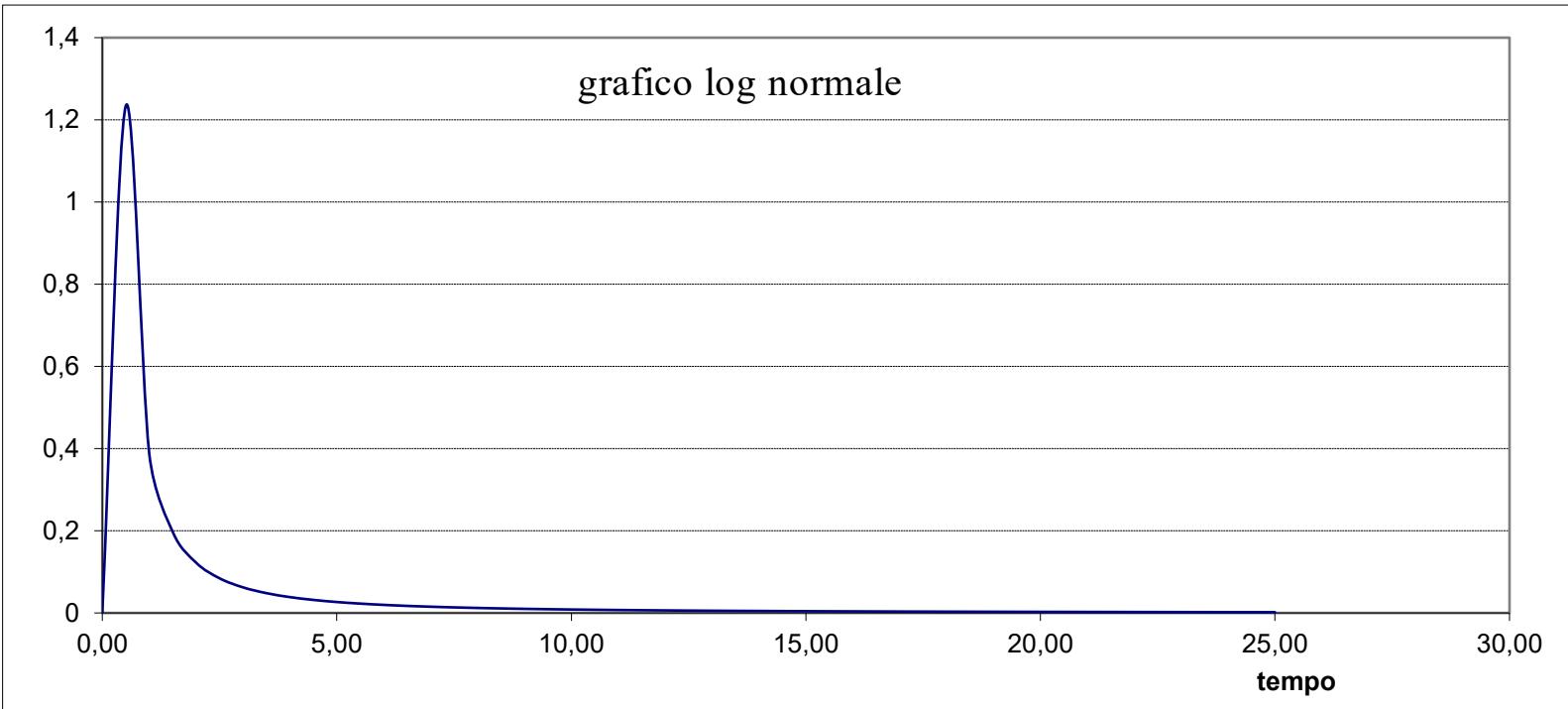
log normale *monotone* *gaussiana* *log normale*

distribuzione gaussiana

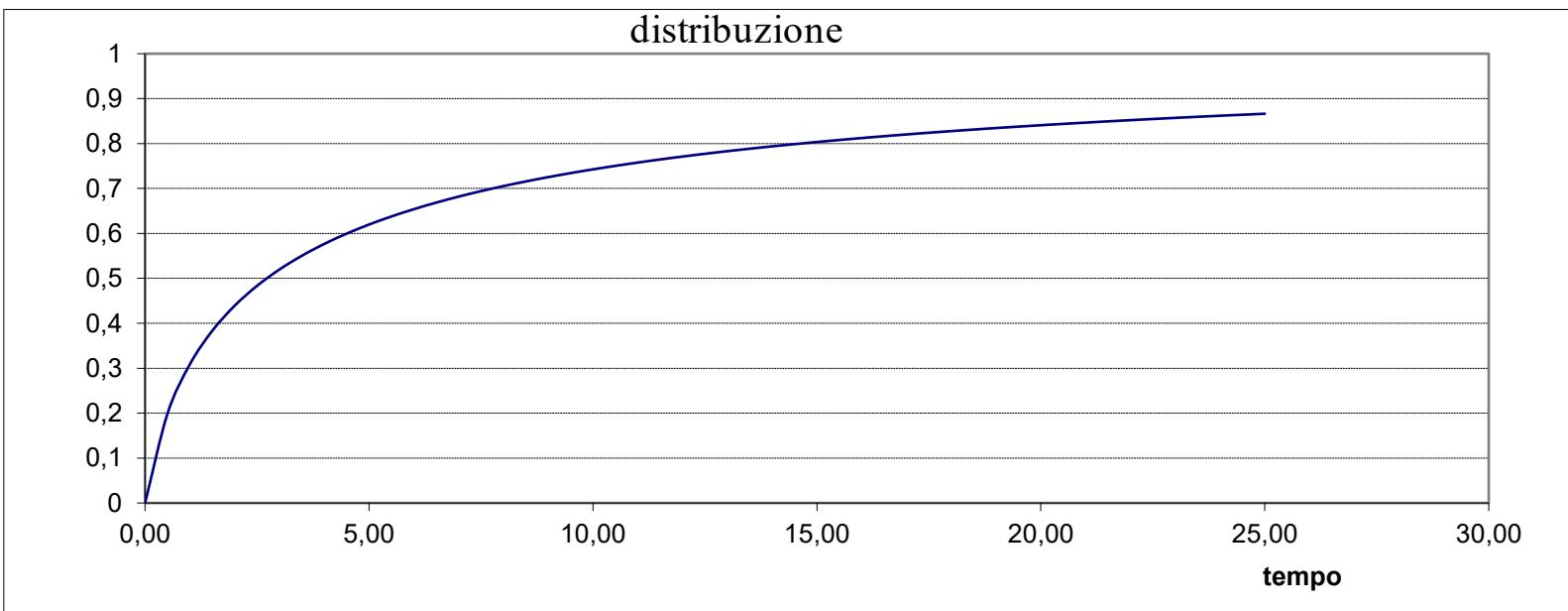
Φ è densità Gauss
Φ è distrib. Gauss



Take a look at my excel file «Modello lognormale (Legato)», please!



simile alla Weibull
ma qui la discesa/decrescita
è molto più ripida/veloce



From WIKIPEDIA

inverte la x con y

X=lognorm Y=norm

Notation	$\text{Lognormal}(\mu, \sigma^2)$
Parameters	$\mu \in (-\infty, +\infty)$, $\sigma > 0$
Support	$x \in (0, +\infty)$
PDF	$\frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right)$
CDF	$\frac{1}{2} + \frac{1}{2} \operatorname{erf}\left[\frac{\ln x - \mu}{\sqrt{2}\sigma}\right]$
Quantile	$\exp(\mu + \sqrt{2\sigma^2} \operatorname{erf}^{-1}(2F - 1))$
Mean	$\exp\left(\mu + \frac{\sigma^2}{2}\right)$
Median	$\exp(\mu)$
Mode	$\exp(\mu - \sigma^2)$
Variance	$[\exp(\sigma^2) - 1] \exp(2\mu + \sigma^2)$
Skewness	$(e^{\sigma^2} + 2)\sqrt{e^{\sigma^2} - 1}$
Ex. kurtosis	$\exp(4\sigma^2) + 2\exp(3\sigma^2) + 3\exp(2\sigma^2) - 6$

$\Phi \triangleq \text{GAUSS DISTRIBUTION}$

$$x = e^y \implies y = \ln x$$

dimostrazione (non richiesta)

$$\begin{aligned} f_X(x) &= \frac{d}{dx} \Pr(X \leq x) = \frac{d}{dx} \Pr(\ln X \leq \ln x) \\ &= \frac{d}{dx} \Phi\left(\frac{\ln x - \mu}{\sigma}\right) \\ &= \varphi\left(\frac{\ln x - \mu}{\sigma}\right) \frac{d}{dx} \left(\frac{\ln x - \mu}{\sigma}\right) \\ &= \varphi\left(\frac{\ln x - \mu}{\sigma}\right) \frac{1}{\sigma x} \\ &= \frac{1}{x} \cdot \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right). \end{aligned}$$

normale
standardizzata

$$F_X(x) = \Phi\left(\frac{(\ln x) - \mu}{\sigma}\right)$$

$$\frac{1}{2} \left[1 + \operatorname{erf}\left(\frac{\ln x - \mu}{\sigma\sqrt{2}}\right) \right] = \frac{1}{2} \operatorname{erfc}\left(-\frac{\ln x - \mu}{\sigma\sqrt{2}}\right)$$

where $\operatorname{erf}(z) \triangleq \frac{2}{\sqrt{\pi}} \int_{t=0}^z e^{-t^2} dt$

and $\operatorname{erfc}(z) \triangleq 1 - \operatorname{erf}(z)$

EXTRA

Appendice I

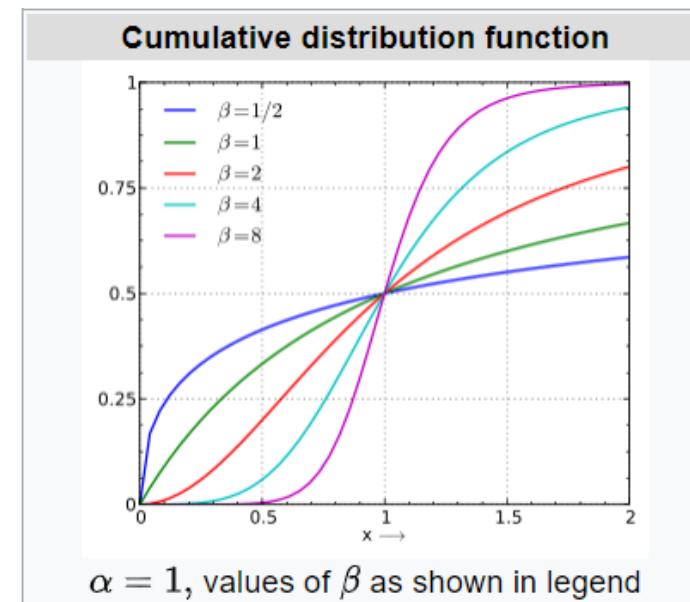
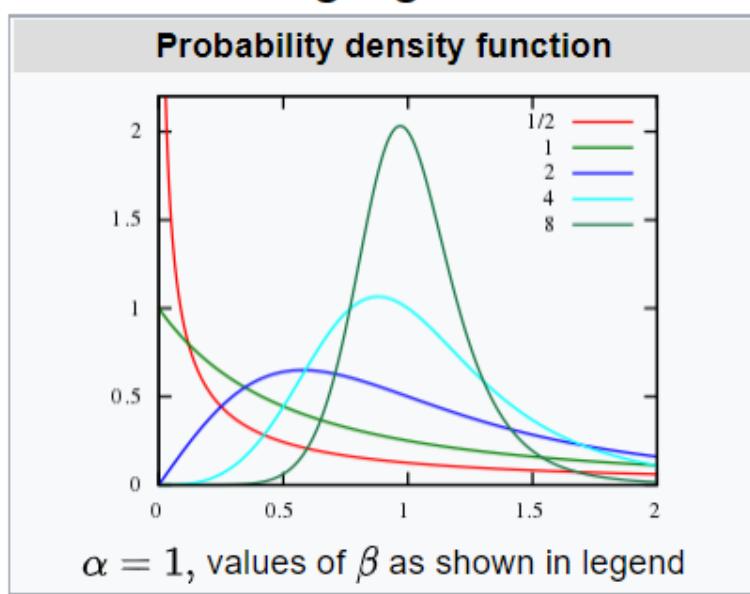
La funzione di distribuzione
Log-Logistic

Log-logistic distribution

From Wikipedia, the free encyclopedia

In probability and statistics, the **log-logistic distribution** (known as the **Fisk distribution** in economics) is a continuous probability distribution for a non-negative random variable. It is used in **survival analysis** as a parametric model for events whose rate increases initially and decreases later, as, for example, mortality rate from cancer following diagnosis or treatment. It has also been used in **hydrology** to model stream flow and **precipitation**, in **economics** as a simple model of the **distribution of wealth or income**, and in **networking** to model the transmission times of data considering both the network and the software.

The log-logistic distribution is the probability distribution of a random variable whose **logarithm** has a **logistic distribution**. It is similar in shape to the **log-normal distribution** but has **heavier tails**. Unlike the log-normal, its **cumulative distribution function** can be written in **closed form**.



There are several different parameterizations of the distribution in use. The one shown here gives reasonably interpretable parameters and a simple form for the cumulative distribution function.^{[3][4]} The parameter $\alpha > 0$ is a scale parameter and is also the median of the distribution. The parameter $\beta > 0$ is a shape parameter. The distribution is unimodal when $\beta > 1$ and its dispersion decreases as β increases.

The cumulative distribution function is

$$\begin{aligned} F(x; \alpha, \beta) &= \frac{1}{1 + (x/\alpha)^{-\beta}} \\ &= \frac{(x/\alpha)^\beta}{1 + (x/\alpha)^\beta} \\ &= \frac{x^\beta}{\alpha^\beta + x^\beta} \end{aligned}$$

where $x > 0, \alpha > 0, \beta > 0$.

The probability density function is

$$f(x; \alpha, \beta) = \frac{(\beta/\alpha)(x/\alpha)^{\beta-1}}{(1 + (x/\alpha)^\beta)^2}$$

Attenzione:
questa volta α è usato per indicare
il fattore di scala e β il fattore di
forma!

Quantiles [edit]

The quantile function (inverse cumulative distribution function) is :

$$F^{-1}(p; \alpha, \beta) = \alpha \left(\frac{p}{1-p} \right)^{1/\beta}.$$

Può essere usata per generare realizzazioni!

It follows that the median is α , the lower quartile is $3^{-1/\beta}\alpha$ and the upper quartile is $3^{1/\beta}\alpha$.

Se $F(\bar{x}_p) = p$, $p \in (0,1)$
allora \bar{x}_p è detto
 p -simo quantile della F
 $p \doteq 0.5$ è detto mediana

Moments [edit]

The k th raw moment exists only when $k < \beta$, when it is given by^{[5][6]}

$$E(X^k) = \alpha^k B(1 - k/\beta, 1 + k/\beta)$$

$$= \alpha^k \frac{k\pi/\beta}{\sin(k\pi/\beta)}$$

where B is the beta function. Expressions for the mean, variance, skewness and kurtosis can be derived from this. Writing $b = \pi/\beta$ for convenience, the mean is

$$E(X) = \alpha b / \sin b, \quad \beta > 1,$$

and the variance is

$$\text{Var}(X) = \alpha^2 (2b / \sin 2b - b^2 / \sin^2 b), \quad \beta > 2.$$

Explicit expressions for the skewness and kurtosis are lengthy.^[7] As β tends to infinity the mean tends to α , the variance and skewness tend to zero and the excess kurtosis tends to 6/5 (see also related distributions below).

Per esercizio, potete graficare densità e distribuzione del modello "Log-logistic"!

Survival analysis [edit]

(RELIABILITY)

The log-logistic distribution provides one parametric model for survival analysis. Unlike the more commonly used Weibull distribution, it can have a non-monotonic hazard function: when $\beta > 1$, the hazard function is unimodal (when $\beta \leq 1$, the hazard decreases monotonically). The fact that the cumulative distribution function can be written in closed form is particularly useful for analysis of survival data with censoring.^[8] The log-logistic distribution can be used as the basis of an accelerated failure time model by allowing α to differ between groups, or more generally by introducing covariates that affect α but not β by modelling $\log(\alpha)$ as a linear function of the covariates.^[9]

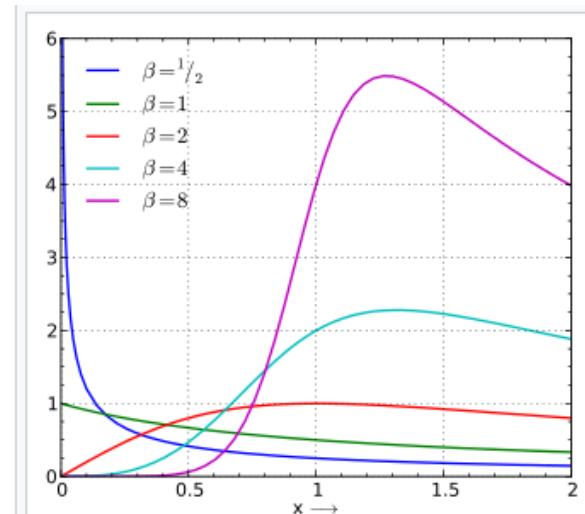
The survival function is

$$S(t) = 1 - F(t) = [1 + (t/\alpha)^\beta]^{-1},$$

and so the hazard function is

$$h(t) = \frac{f(t)}{S(t)} = \frac{(\beta/\alpha)(t/\alpha)^{\beta-1}}{1 + (t/\alpha)^\beta}.$$

A anche questi grafici
potete riprodurre
con Excel !



Hazard function. $\alpha = 1$, values of β as shown in legend

EXTRA

Appendice II

Un teorema per:
**FUNZIONI DI VARIABILI ALEATORIE
strettamente monotone a tratti ...**

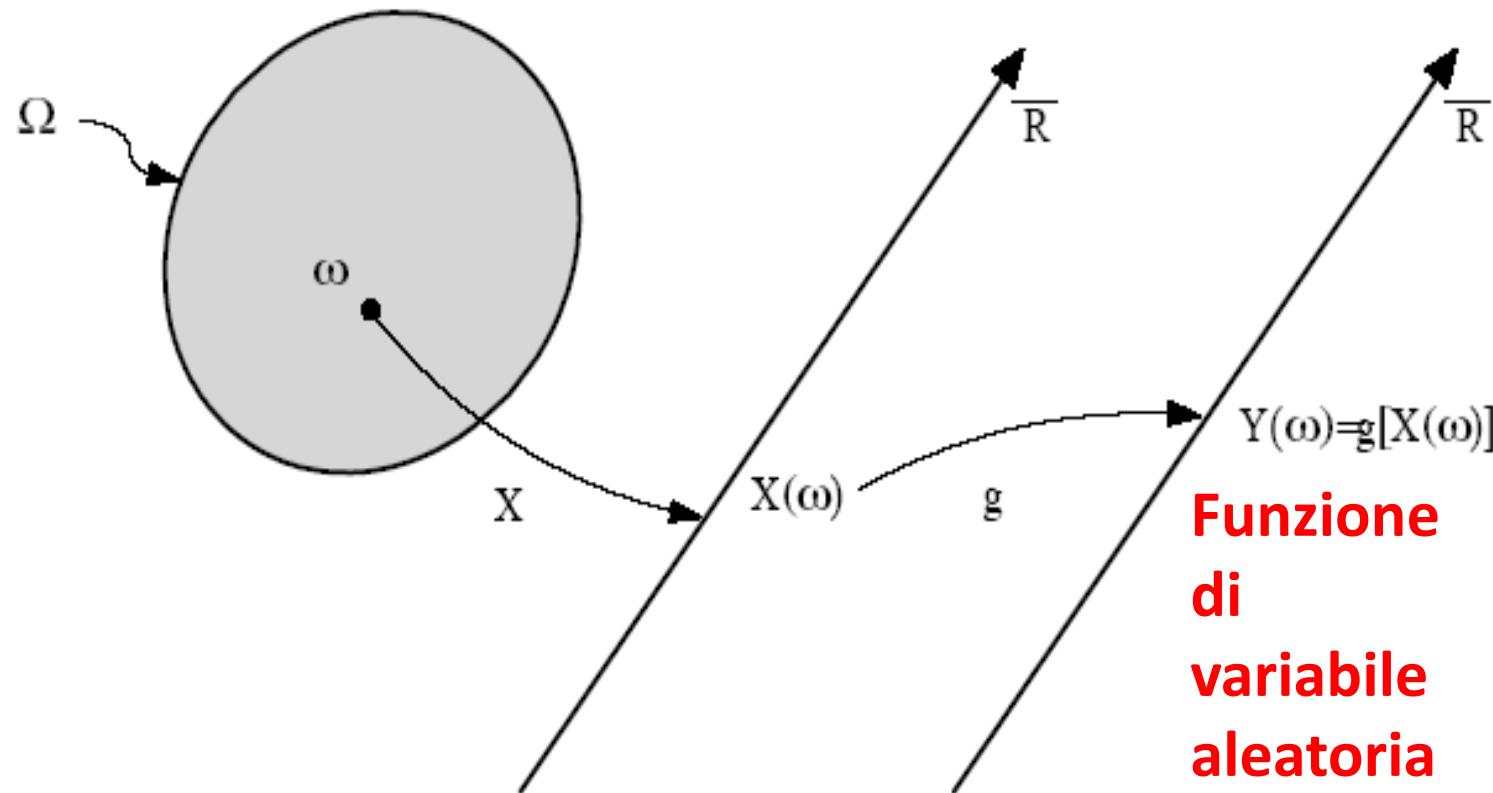
Un teorema per: FUNZIONI DI VARIABILI ALEATORIE

Dal libro del prof. Gelli, "Probabilità e informazione"

86

Trasformazioni di una variabile aleatoria

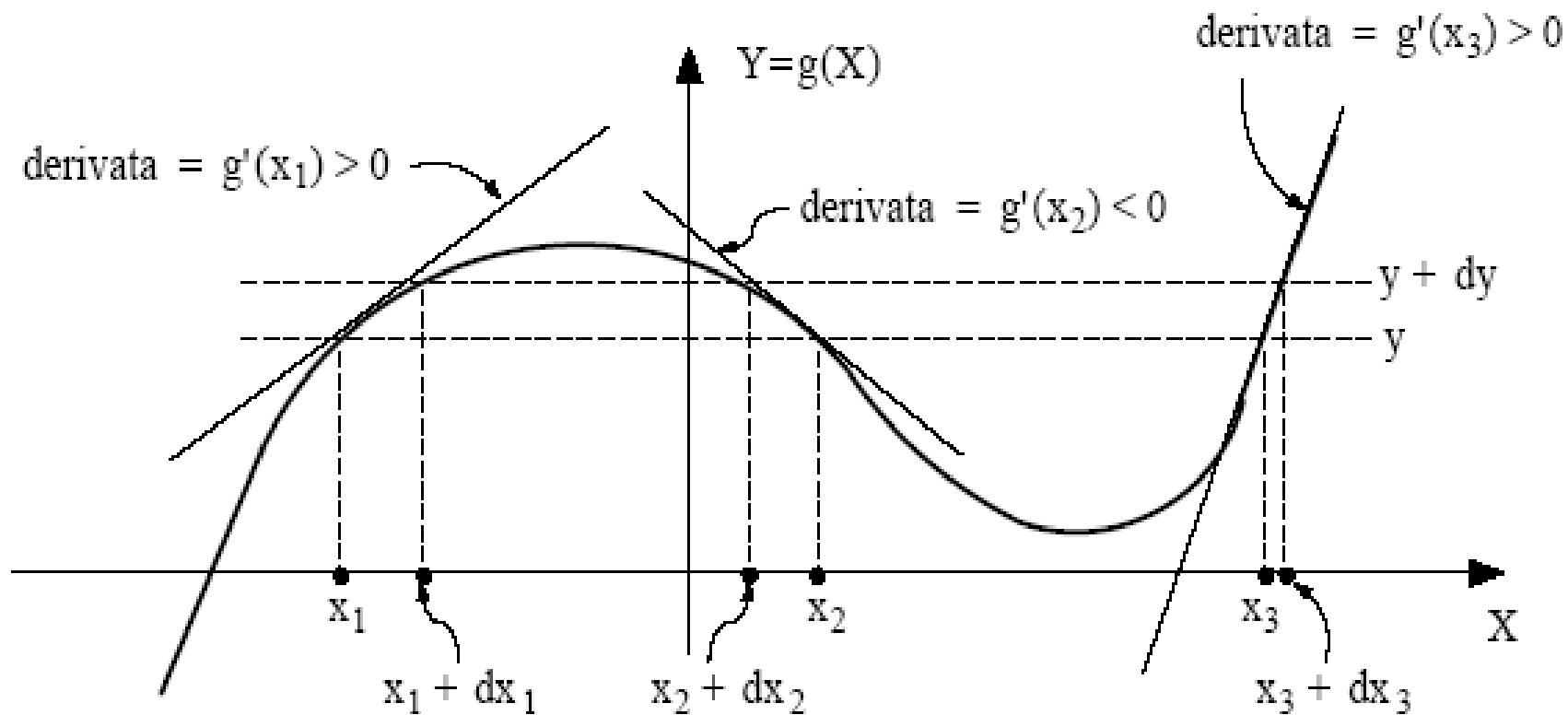
"Trasformazione" = "funzione"



Trattiamo il caso
«moderatamente
difficile»:

G è funzione
differenziabile e
«strettamente
monotona a tratti»,
con punti a derivata
nulla, che separano i
tratti,

ma senza interi
intervalli a derivata
nulla!



Definizione (trasformazione di una variabile aleatoria). Sia X una variabile aleatoria definita sullo spazio di probabilità (Ω, \mathcal{S}, P) , e $g(x)$ una funzione definita in \mathbb{R} e a valori in \mathbb{R} , tale che l'insieme di definizione di $g(x)$ contenga il codominio \mathcal{X} della funzione $X(\omega)$. La trasformazione $Y = g(X)$ definisce una nuova variabile aleatoria ottenuta associando a $\omega \in \Omega$ il valore $Y(\omega) = g[X(\omega)] \in \mathbb{R}$.

4.2.3 Calcolo della pdf di $Y = g(X)$ **$g(x)$ differenziabile e non monotona**

Affrontiamo adesso il problema di determinare la pdf di $Y = g(X)$ in funzione della pdf di X . Di importanza fondamentale è il seguente teorema, nel quale $g'(x)$ indica la derivata prima di $g(x)$:

Teorema 4.1 (teorema fondamentale sulle trasformazioni di variabili aleatorie). Sia X una variabile aleatoria avente pdf $f_X(x)$, e si consideri la trasformazione $Y = g(X)$; la pdf di Y è data da:

$$f_Y(y) = \begin{cases} 0, & \text{se l'equazione } y = g(x) \text{ non ammette soluzioni;} \\ \sum_I \frac{f_X(x_I)}{|g'(x_I)|}, & \text{dove } x_I \text{ è una soluzione dell'equazione } y = g(x). \end{cases}$$

Corollario: Se $g(x)$ è strettamente monotona, allora esiste una soluzione unica di $y=g(x)$

Caso di particolare interesse

$$Y = aX + b, \quad x = \frac{y - b}{a}, \quad |g'(x)| = |a|, \quad \text{per cui:} \quad f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y - b}{a}\right)$$

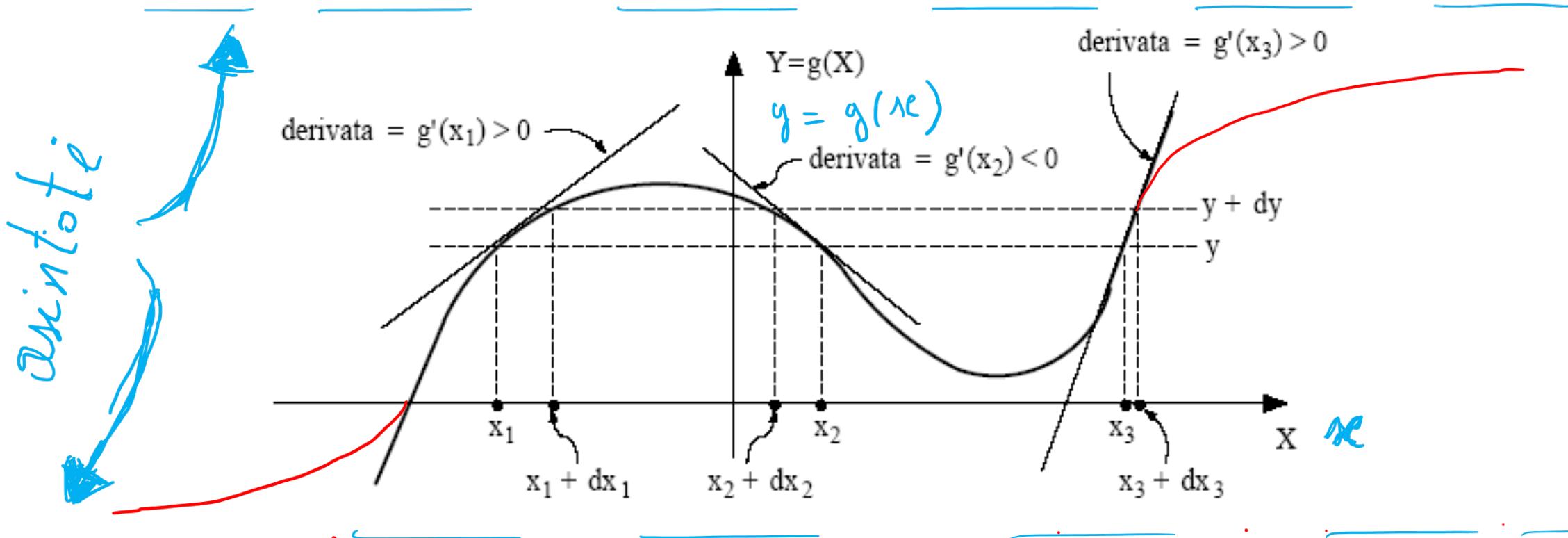


Fig. 4.14. Dimostrazione del teorema fondamentale sulle trasformazioni di variabili aleatorie. Le soluzioni dell'equazione $y = g(x)$ sono x_1 , x_2 , ed x_3 .

Se y è un valore tale che l'equazione $g(x) = y$ non ammette soluzioni, allora $f_Y(y) = 0$. Infatti, se y non appartiene alla frontiera del codominio di $g(x)$, è possibile scegliere dy sufficientemente piccolo tale che

$$\{y < g(x) \leq y + dy\} = \emptyset \Rightarrow f_Y(y) = 0.$$

Se invece y appartiene alla frontiera del codominio di $g(x)$, posso comunque porre $f_Y(y) = 0$.

Viceversa, si consideri il caso in cui y appartenga al codominio di $g(x)$,

$$f_Y(y) dy = P(y < Y \leq y + dy) = P(x_1 < X \leq x_1 + dx_1) + P(x_2 + dx_2 < X \leq x_2) + P(x_3 < X \leq x_3 + dx_3),$$

dove $dx_1 > 0$, $dx_2 < 0$, $dx_3 > 0$. (Fig. 4.14) e, poiché dy è infinitesimo, i tre insiemi cui appartiene X sono mutuamente esclusivi. Poichè:

$$P\{x_1 < X \leq x_1 + dx_1\} = f_X(x_1) dx_1; \quad dx_1 = dy/g'(x_1);$$

$$P\{x_2 + dx_2 < X \leq x_2\} = f_X(x_2) |dx_2|; \quad \text{ed inoltre} \quad dx_2 = dy/g'(x_2);$$

$$P\{x_3 < X \leq x_3 + dx_3\} = f_X(x_3) dx_3; \quad dx_3 = dy/g'(x_3);$$

dove (Fig. 4.14) $g'(x_1) > 0$, $g'(x_2) < 0$, e $g'(x_3) > 0$, risulta

$$f_Y(y) dy = \frac{f_X(x_1)}{g'(x_1)} dy + \frac{f_X(x_2)}{|g'(x_2)|} dy + \frac{f_X(x_3)}{g'(x_3)} dy,$$

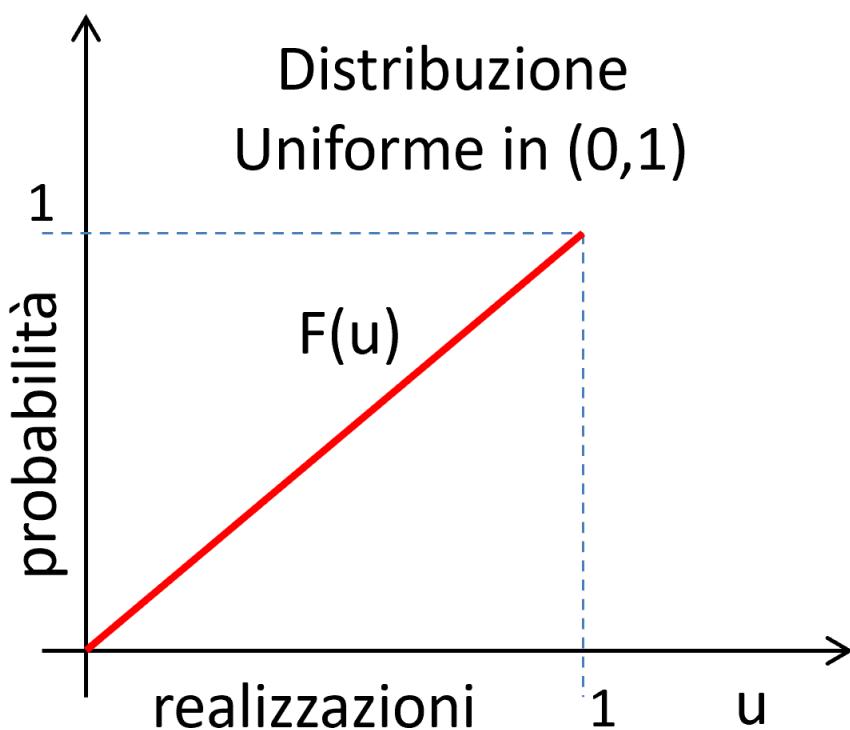
ed eliminando dy , si ha l'asserto.

Prob. di unione di eventi disgiunti

eventi

IMPORTANTE

Come si è visto precedentemente, la funzione (inversa) quantile è la trasformazione che permette di produrre realizzazioni a partire da una funzione di distribuzione invertibile (Esponziale, Weibull ecc)



In generale dunque si ha che:

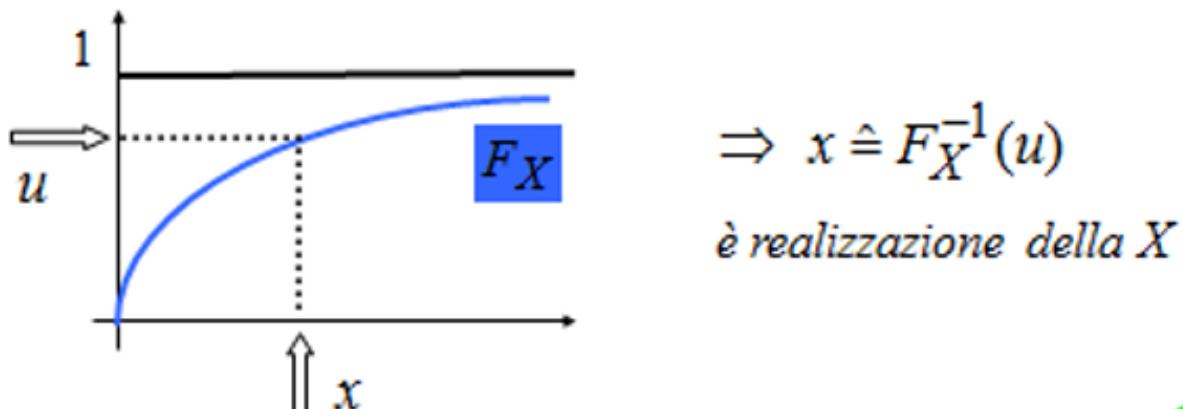
Dai numeri casuali alle realizzazioni di v.a.

$\{u_1, u_2, \dots, u_n, \dots\}$ realizzazioni indipendenti di $U\{0,1\}$

Generazione di realizzazioni delle Var. Al. d'ingresso

Metodo della trasformazione inversa
dimostrazione

$$\Pr\{X \leq x\} \hat{=} F_X(x) = F_X(F_X^{-1}(u)) = u = \Pr\{U \leq u\}$$



A parole: presa una funzione di **distribuzione invertibile** $F_x(x)$ allora si ha che

Il metodo non è utilizzabile per tutte le distribuzioni

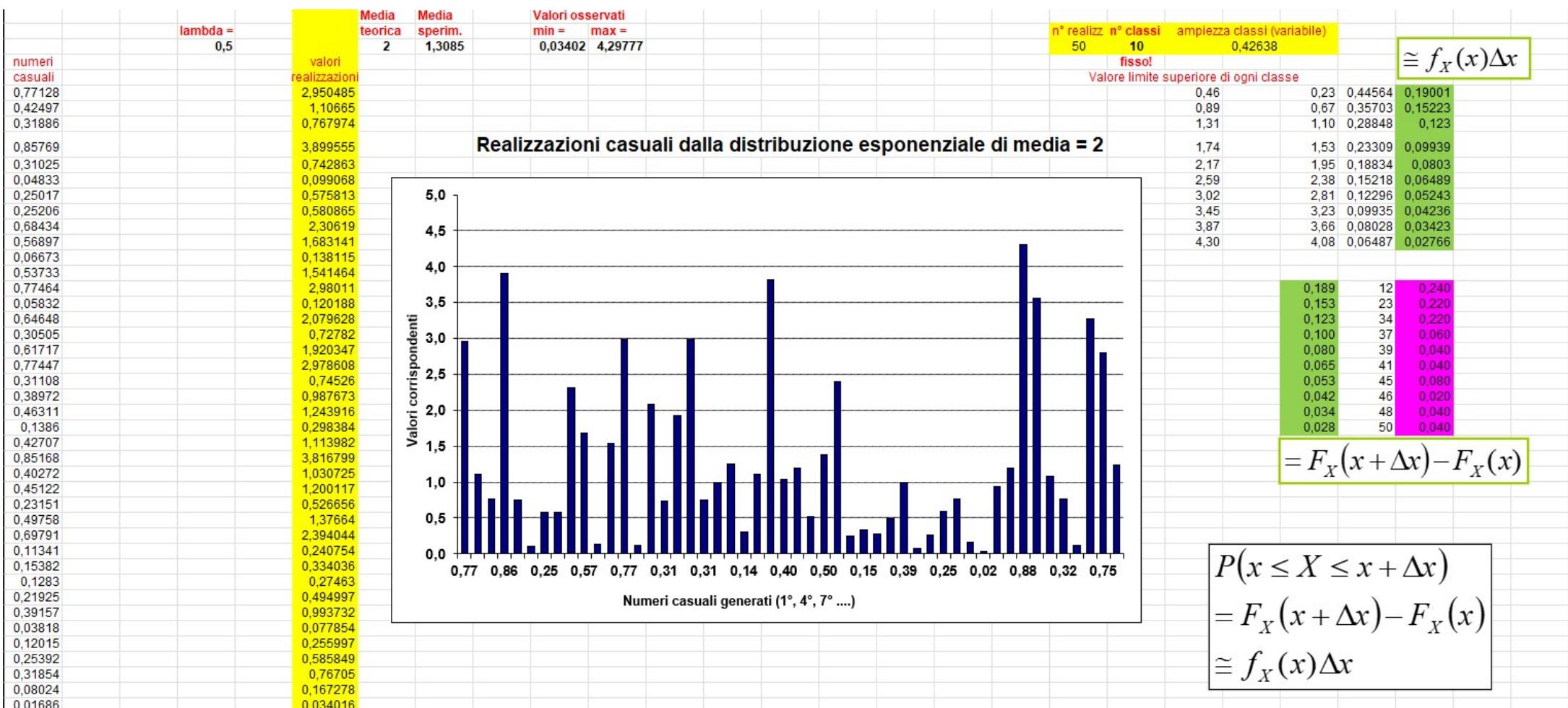
$F_X(F_X^{-1}(u)) = u$ dunque la funzione di distribuzione nell'inversa della distribuzione in u ritorna u. Esempio esponenziale con $\lambda = 1 \Rightarrow F_X(x) = 1 - e^{-x} \Rightarrow F_X^{-1}(x) = -\ln(1 - x)$

$F_X(F_X^{-1}(u)) = 1 - e^{-\ln(1+u)} = 1 - (1 - u) = u$ Quindi come abbiamo visto in

precedenza la realizzazione è uguale alla probabilità con cui si osserva la realizzazione

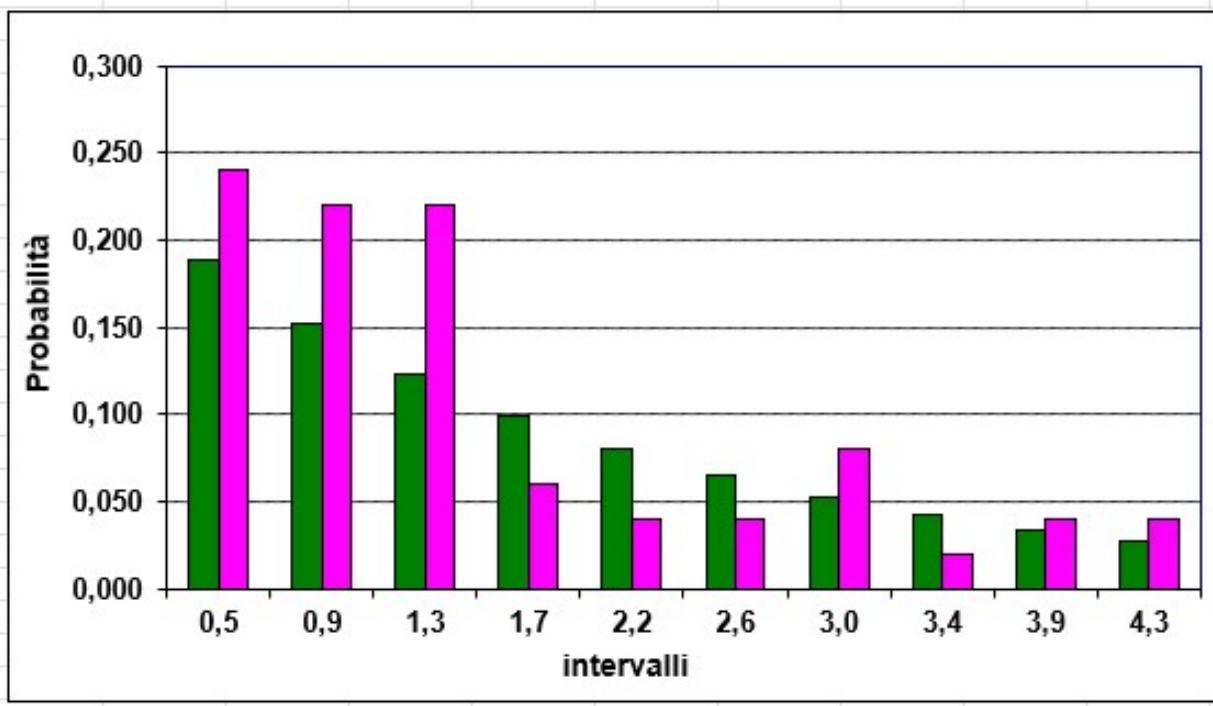
L'unica distribuzione che ha ascissa pari a ordinata è $F_X(x) = x$ appunto la diagonale del quadrato.

Ci mettiamo nell'ipotesi di modelli in cui la distribuzione è invertibile.



Grazie al **metodo della trasformazione inversa** si sono generate molte realizzazioni, che appunto nella cella mostrano la formula inversa. La media sperimentale sarebbe la classica media aritmetica e mostra nella simulazione che facendo più prove, tende al valore della media teorica. La media teorica sarebbe il valore atteso appunto. Quindi si capisce e vedremo in futuro che per n tendente all'infinito la media aritmetica tenda al valore atteso. Facendo simulazioni si ottengono valori che sono più piccoli o più grandi del valore atteso. L'oscillazione dei risultati a parità di dimensione del campione, è dovuta alla varianza! Non solo per la casualità. (Non si sa ancora di quanto si sbaglia) Si hanno poi i valori min e max, che mostrano quanto ad esempio un tempo di esecuzione può essere piccolissimo oppure molto grande. Il grafico ci potrebbe ad esempio mostrare il fenomeno di burst all'interno di una coda. Ad esempio vado dall'elettrauto e devo cambiare la lampadine e sto 0.3 secondi, mentre può capitare che arriva quello che deve cambiare la batteria ed i cavi annessi e sta 4.3 secondi (ovviamente non impiega secondi un elettrauto). Riordinare i valori non basta per ricostruire la legge esponenziale di media. Oltre ai valori di min e max occorre fissare un numero di classi, ovvero un numero di raggruppamenti dove i valori possono finire, un numero di realizzazione da prendere in esame e l'ampiezza delle classi, che sarebbe la differenza tra max e min diviso il numero di classi fissato.

Ricostruzione legge esponenziale di media = 2



lizz	n° classi	ampiezza classi (variabile)	
	10	0,42638	
fisso!			
Valore limite superiore di ogni classe			
2			
	0,46	0,23	0,44564
	0,89	0,67	0,35703
	1,31	1,10	0,28848
	1,74	1,53	0,23309
	2,17	1,95	0,18834
	2,59	2,38	0,15218
	3,02	2,81	0,12296
	3,45	3,23	0,09935
	3,87	3,66	0,08028
	4,30	4,08	0,06487
			0,19001
			0,15223
			0,123
			0,09939
			0,0803
			0,06489
			0,05243
			0,04236
			0,03423
			0,02766

$$\cong f_X(x)\Delta x$$

Il limite superiore di ogni classe si ottiene partendo da 0 e sommando l'ampiezza della classe al valore minimo. L'ultimo limite sarà il valore di max.

$f_X(x)\Delta x$ corrisponde alla percentuale di valori che ci si aspettano in una classe. (19% nello screen)
Ma tale percentuale si può ottenere anche con la distribuzione

$$\cong f_X(x)\Delta x$$

classe

0,23	0,44564	0,19001
0,67	0,35703	0,15223
1,10	0,28848	0,123
1,53	0,23309	0,09939
1,95	0,18834	0,0803
2,38	0,15218	0,06489
2,81	0,12296	0,05243
3,23	0,09935	0,04236
3,66	0,08028	0,03423
4,08	0,06487	0,02766

0,189	12	0,240
0,153	23	0,220
0,123	34	0,220
0,100	37	0,060
0,080	39	0,040
0,065	41	0,040
0,053	45	0,080
0,042	46	0,020
0,034	48	0,040
0,028	50	0,040

$$= F_X(x + \Delta x) - F_X(x)$$

Guardando il grafico con 15 classi si nota come il modello teorico approssima ancora meglio la legge esponenziale, mentre la simulazione sballa di molto! Da notare che la media sperimentale nel esempio è di 1.9835! MOLTO VICINA A 2. Quindi la casualità anche se c'è non porta solo lei a sbagliare! Quanto si peggiora per caso dipende dalla varianza

$$\text{In particolare si sa che: } F_X(x + \Delta x) - F_X(x) \cong f_X(x)\Delta x$$

per il concetto di limite di rapporto incrementale. Dunque guardando le due colonne in verde le percentuali a meno di qualche decimale si approssima abbastanza bene. Ovviamente per applicazioni critiche ad esempio: percentuale di sopravvivenza di un uomo, i decimali (delta x) fanno la differenza. In conclusione l'istogramma è diviso in due: la parte verde è il modello teorico, la parte fucsia è il modello ricavato dalla simulazione. Il numero di classi ovviamente ha un ruolo nel mostrare il modello. Magari più classi mostrano meglio o magari no se la varianza è molto alta. Sotto un esempio con 15 classi:

1° classi ampiezza classi (variabile)

15 0,41535

fisso!

re limite superiore di ogni classe

0,42	0,21	0,45039	0,18707
0,83	0,63	0,36568	0,15189
1,25	1,04	0,29711	0,1234
1,66	1,46	0,24139	0,10026
2,08	1,87	0,19612	0,08146
2,49	2,29	0,15935	0,06618
2,91	2,70	0,12946	0,05377
3,33	3,12	0,10519	0,04369
3,74	3,53	0,08546	0,0355
4,16	3,95	0,06943	0,02884
4,57	4,36	0,05641	0,02343
4,99	4,78	0,04583	0,01904
5,40	5,19	0,03724	0,01547
5,82	5,61	0,03026	0,01257
6,23288	6,03	0,02458	0,01021

$$\cong f_X(x)\Delta x$$

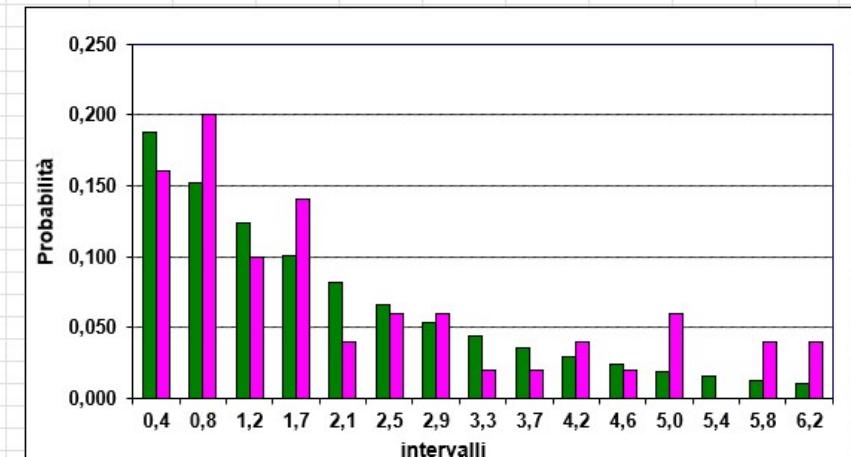
$$P(x \leq X \leq x + \Delta x)$$

$$= F_X(x + \Delta x) - F_X(x)$$

$$\cong f_X(x)\Delta x$$

$$= F_X(x + \Delta x) - F_X(x)$$

Ricostruzione legge esponenziale di media = 2



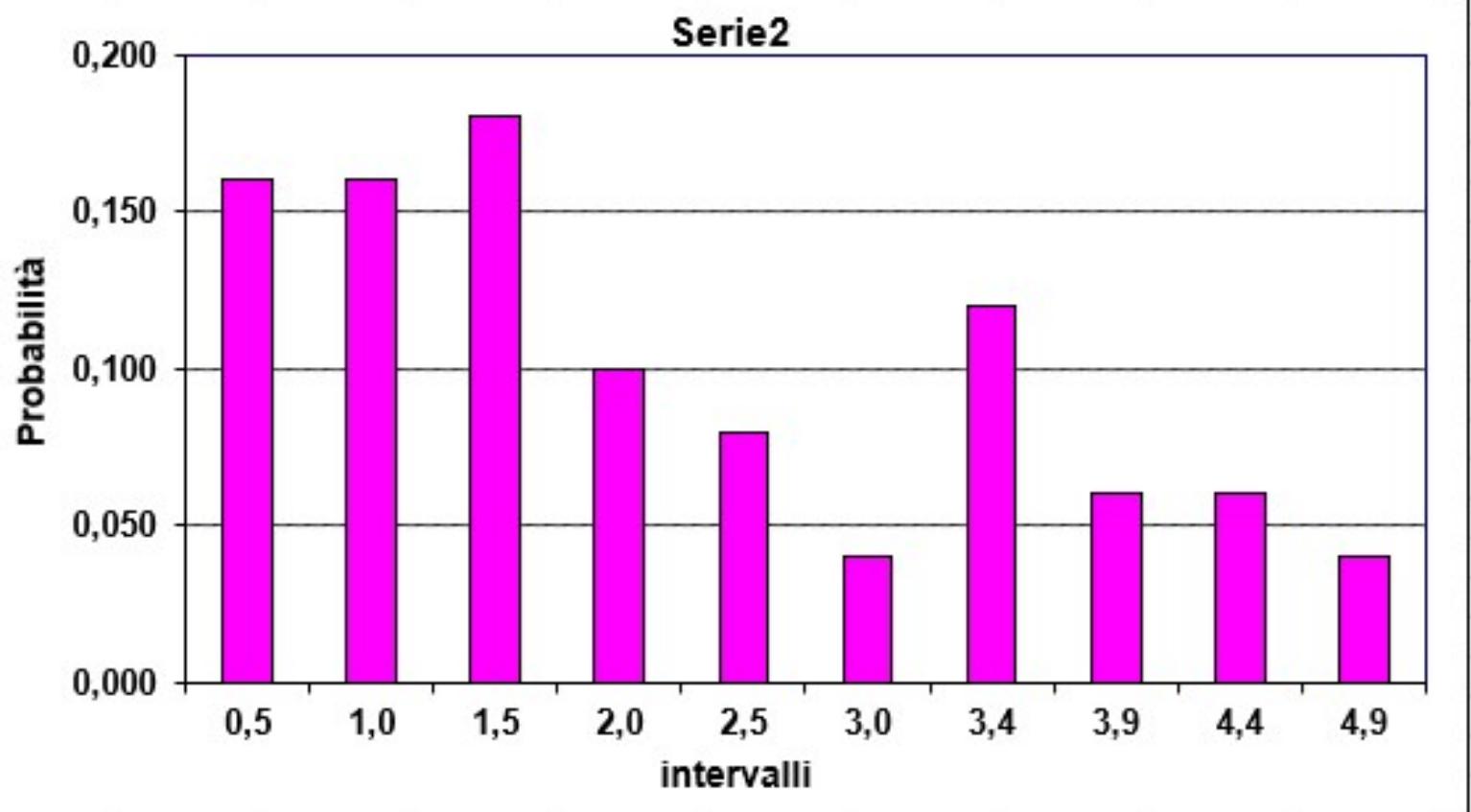
0,187	8	0,160
0,152	18	0,200
0,124	23	0,100
0,100	30	0,140
0,082	32	0,040
0,066	35	0,060
0,054	38	0,060
0,044	39	0,020
0,036	40	0,020
0,029	42	0,040
0,023	43	0,020
0,019	46	0,060
0,015	46	0,000
0,013	48	0,040
0,010	50	0,040

Noi faremo statistica inferenziale.

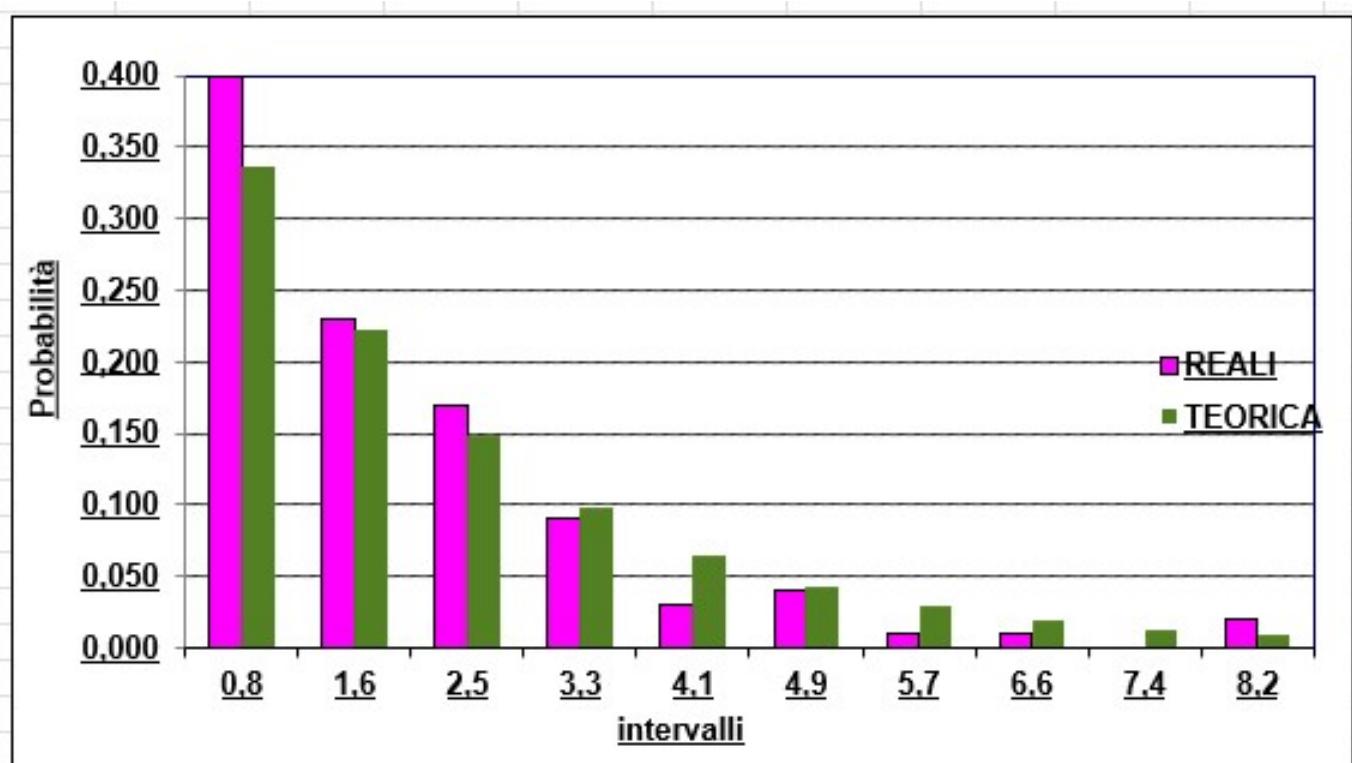
La **statistica descrittiva** consiste nel dire quali parametri ci danno una caratteristica sintetica sulla forma: valore atteso, varianza, momento del secondo ordine, simmetria ecc.

L'**inferenza** (deduzione vedi dizionario) sono i metodi per dire quanto sarà il valore atteso.

Quindi prendo un oggetto/un fenomeno, di cui scelgo quanti e quali attributi e ne provo a dedurre la forma. Quindi in questo caso l'inferenza completa che noi facciamo è la forma della legge. Prendiamo ad esempio il seguente istogramma:



Non assomiglia per niente alla curva della legge esponenziale. La statistica consiste nel ricavare con pochi dati o meno dati possibili una risposta accettando il rischio di sbagliare. Quindi per prima cosa interessa il numero di dati. Poi il numero di classi.

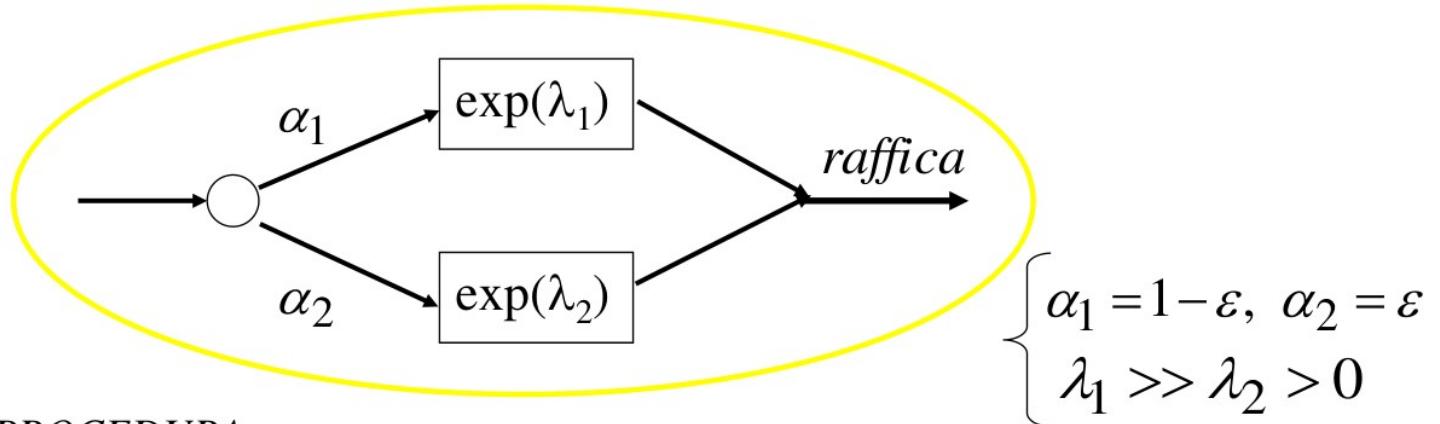


Nel grafico precedente si vede che anche se con media sperimentale 1.62(!), se uso 100 realizzazione quindi il doppio dei dati, per 10 classi la curva della legge esponenziale è abbastanza simile alla teorica. Prendiamo adesso la distribuzione iperesponenziale:

Trasformazione inversa per la distribuzione iperesponenziale

$$F_Y(y) \hat{=} \sum_{i=1}^n \alpha_i (1 - \exp\{-\lambda_i y\}), \quad \lambda_1 \neq \lambda_2 \neq \dots \neq \lambda_n$$

$$\text{con: } \sum_{i=1}^n \alpha_i = 1, \quad \alpha_i > 0 \quad i = 1, \dots, n$$



PROCEDURA:

genera $u \in U\{0,1\}$

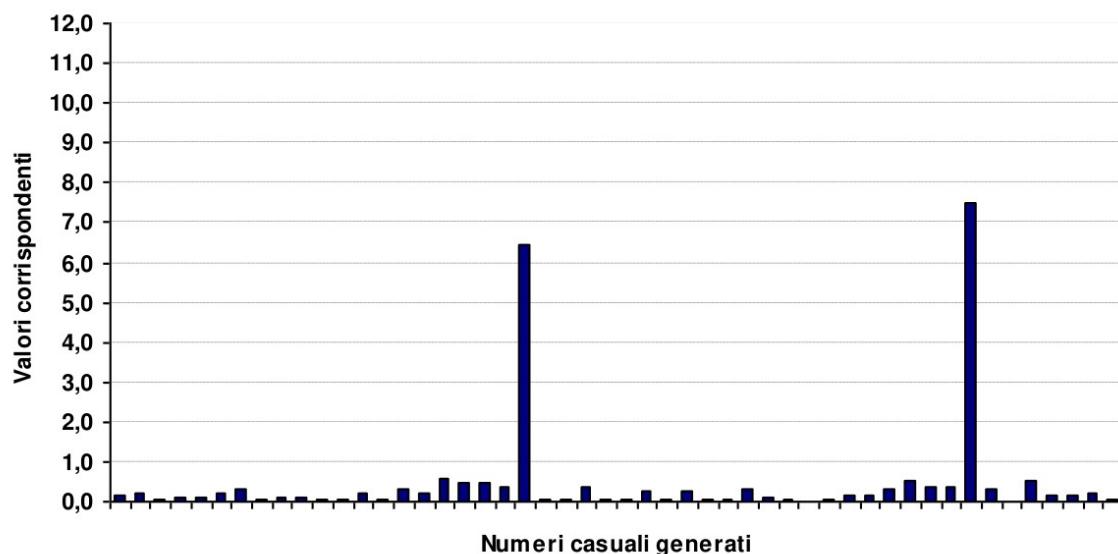
if $u \leq \alpha_1$ *then* *genera*($y \approx EXP(\lambda_1)$)

else *genera*($y \approx EXP(\lambda_2)$)

La trasformazione inversa del iperesponenziale è una combinazione convessa. Tramite l'iperesponenziale si può studiare l'effetto di bursting. In particolare come mostra l'immagine si ha che λ_1 è molto grande, quindi l'esponenziale molto veloce e tempi brevi λ_2 è piccolissimo quindi esponenziale lento e tempi lunghissimi. Quindi generato un numero casuale u tra 0 e 1, si ha che $u \leq \alpha_1$, $\alpha_1 = 0.95$ genera una realizzazione di e^{λ_1} altrimenti di e^{λ_2} . I numeri "passano" uno alla volta ecco perchè si osserva un effetto raffica. Il compito della statistica sarà capire i valori di alpha e lambda, a partire da dei dati. Vediamo di seguito la generazione monte carlo di realizzazioni iperesponenziali

Realizzazioni casuali dalla legge IPEResponenziale

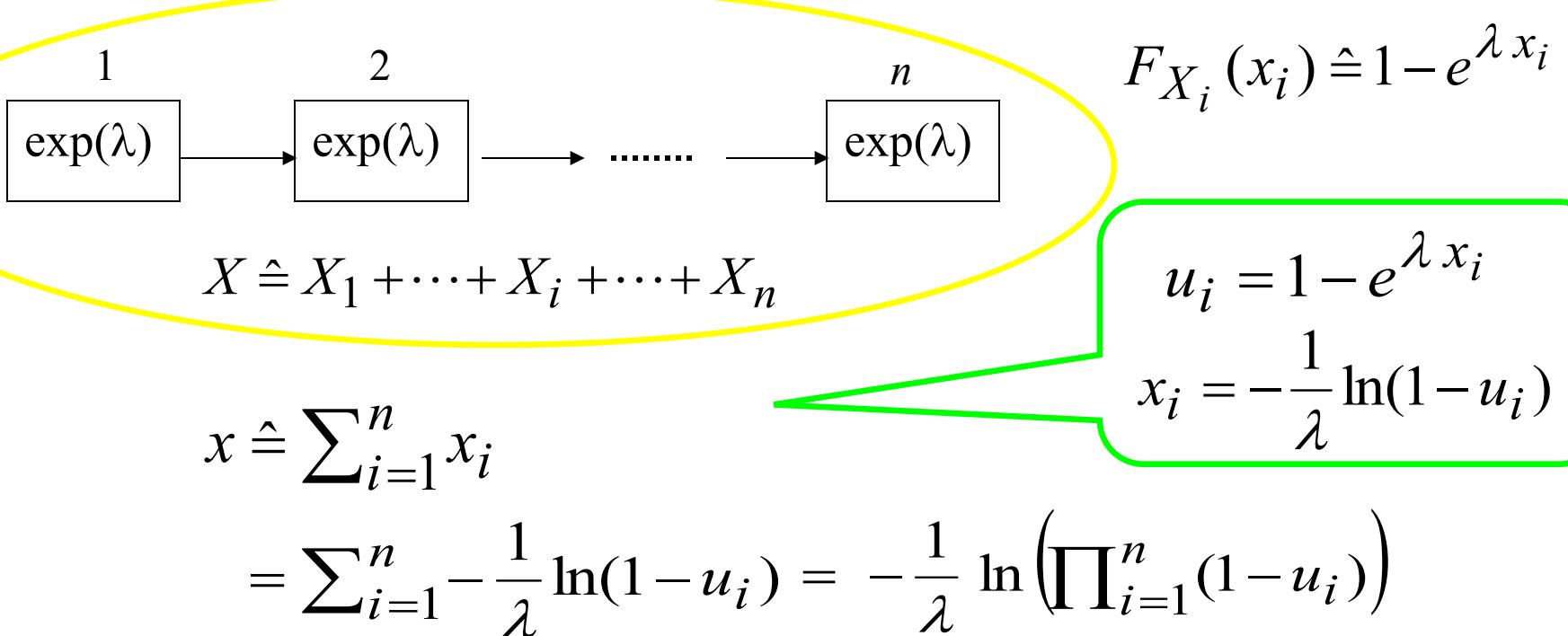
(alfa = 0,95 - lambda_1 = 5 - lambda_2 = 0,5)



Metodo della trasformazione inversa e legge di Erlang

$$F_X(x) \hat{=} 1 - e^{-\lambda x} \sum_{i=0}^{n-1} \frac{(\lambda x)^i}{i!} \quad \lambda > 0$$

Somma di esponenziali identiche e indipendenti



Erlang come vedremo anche dopo non ha assenza di memoria! Questo perchè essendo somma di esponenziali identiche e indipendenti può essenzialmente ritornare la "posizione" all'interno della sua storia.

STATISTICHE E STIMATORI

Una **statistica** è una variabile aleatoria funzione di un numero fissato ($n \geq 1$) di altre variabili aleatorie, ma che **non contiene** alcun parametro incognito.

La realizzazione della statistica è una quantità numerica calcolata a partire da un campione di dati (variabile aleatoria funzione di un numero fissato di altre variabili aleatorie). Proprio perché parte da un campione di dati, non dipende da parametri incogniti

Sia T una statistica per le variabili aleatorie i.i.d. $X_1 \dots X_n$ e sia θ un parametro incognito di queste ultime; allora $T(\mathbf{X})$ è detto **stimatore corretto di θ** se risulta $E[T(\mathbf{X})] = \theta$.

- popolazione: tutti gli studenti di ingegneria
- campione: gruppo di n studenti presi casualmente dalla facoltà di ingegneria

CARATTERISTICHE e valutazione di uno stimatore

($E[T(\mathbf{X})] - \theta$) è detto **errore sistematico (bias)**

è la differenza tra il valore atteso dello stimatore ed il vero valore del parametro. Uno stimatore corretto perciò ha un bias nullo

$MSE \hat{=} E[(T(X_1, \dots, X_n) - \theta)^2]$ è l'**errore quadratico medio**

Misura la variabilità dello stimatore attorno al vero valore del parametro. Un MSE basso indica uno stimatore più preciso

Stimatore consistente se:

capacità di avvicinarsi sempre più al vero valore del parametro che si vuole stimare all'aumentare della dimensione del campione

$\lim_{n \rightarrow \infty} \Pr(|T(X_1, \dots, X_n) - \theta| \geq \varepsilon) = 0$

stimatore
vero valore del parametro
valore molto piccolo
la stima diventa sempre più precisa man mano che si raccolgono più dati
"all'aumentare del campione"
probabilità che la stima T si discosti dal vero valore, diventa sempre più piccola

Valutare tutte le suddette caratteristiche con riferimento ad uno stimatore trovato e usato dagli statistici non è cosa facile, in generale!

Anticipazione di due risultati di analisi probabilistica utili adesso

Dimostreremo in una lezione futura che:

1. il valore atteso di una somma di variabili aleatorie è pari alla somma dei valori attesi di ciascuna, senza bisogno di assumere che quelle variabili siano indipendenti.
2. La varianza di una somma di variabili aleatorie è pari alla somma delle varianze di ciascuna se assumiamo che quelle variabili siano indipendenti.

La statistica “media campionaria”

E' detta media campionaria la var. al seguente :

$$\bar{X}(n) \hat{=} \frac{1}{n} \sum_{i=1}^n X_i$$

X_1 pari alla realizzazione
del primo run
della monte carlo
↓
singole osservazioni
del campione
→ dimensioni
del campione

Se le X_i sono identicamente distribuite e con lo stesso valore atteso, $E[X_i] \hat{=} \mu_i = \mu, i = 1, 2, \dots, n.$

Valore atteso
media campionaria

$$E[\bar{X}(n)] = \frac{1}{n} E\left[\sum_{i=1}^n X_i\right] = \frac{1}{n} \sum_{i=1}^n E[X_i] = \frac{n \cdot \mu}{n} = \mu$$

⇒ $\bar{X}(n)$ e' uno stimatore corretto di μ

se calcolassi la media di tutte le possibili medie campionarie che potrei ottenere da tutti i possibili campioni, otterrei esattamente il valore della media della popolazione.

media della
popolazione

poiché le osservazioni
sono indipendenti e
identicamente distribuite il
valore atteso della somma
è uguale alla somma dei
valori attesi

il valore atteso di ogni
singola osservazione
è pari alla media
della popolazione

$$E[\bar{X}(n)] = \frac{1}{n} (\mu + \mu + \dots + \mu) = \frac{1}{n} \cdot n \cdot \mu$$

Passando alla varianza della media campionaria e assumendo che le X_i siano INDIPENDENTI e con la stessa varianza: $\sigma_i^2 = \sigma^2 \forall i$

indica quanto le medie campionarie di diversi campioni estratti dalla stessa popolazione variano tra loro

$$\begin{aligned} Var[\bar{X}(n)] &= \frac{1}{n^2} E\left[\left(\sum_{i=1}^n X_i - E\left[\sum_{i=1}^n X_i\right]\right)^2\right] = \\ &\quad \text{quadrato della dimensione del campione} \end{aligned}$$

varianza "bassa" allora
stima più precisa

I passaggi algebrici
saranno mostrati in
una futura lezione!

= ...

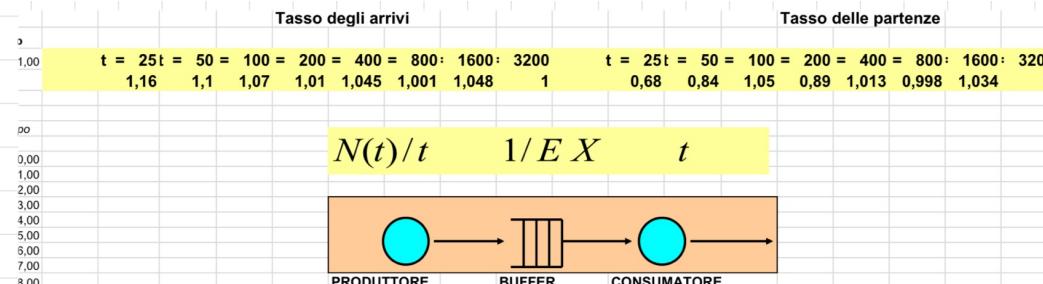
$$= \frac{\sigma^2}{n}$$

⇒ $\bar{X}(n)$ stimatore di bontà crescente con "n"
la varianza della media campionaria

MODELLO PRODUTTORE-CONSUMATORE

MODELLO PRODUTTORE-CONSUMATORE

Utilizzo il metodo Montecarlo per generare realizzazioni e modellare il modello client-server seguente:

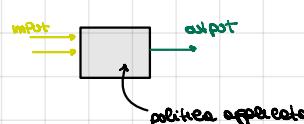


IN PIÙ...

Se gli interarrivi sono esponenziali, anche i tempi di servizio dunque gli input sono di carattere esponenziale. L'output invece non sarà di carattere esponenziale.

Dal punto di vista della simulazione monte carlo, posso riprodurre la storia degli stati di questo sistema nel tempo? Analizzo i due input del sistema: **flusso di arrivo** e la **caratteristica dei servizi**.

L'output è il tempo di risposta/attesa del sistema.
(variabile aleatoria, di cui serve il valore atteso e varianza)



Parto dal generare numeri casuali per generare l'interarrivo. Per passare dal primo al secondo uso la trasformazione inversa e segno l'istante d'arrivo.

* Tramite un altro numero casuale genero un tempo di servizio, dove il primo tempo di servizio ottenuto è proprio quello della prima persona.

Utilizzo il generatore di numeri casuali **PERCHÉ VOGLIO GARANTIRE L'INDIPENDENZA**.

Gli istanti di partenza sono dati dalla somma dell'istante di arrivo più il servizio. Posso ricavarli anche ricorsivamente. Mi concentro sul processo delle partenze che è simmetrico rispetto al processo degli arrivi

Contatore	Numeri Casuali	Tempi di Interarrivo	Istanti di Arrivo	Tasso di Servizio		Numeri Casuali	Tempi di Servizio	Tempi di Soggiorno	
				Valore Atteso in Input	Valore Atteso in Input			Ordine	Tempi di Servizio (1/ μ)
1	0,514899008	0,663784499	0,663784499	1	0,909090909	0,385433153	0,866715914	FIFO	0,909090909
2	0,887865793	0,118934682	0,782711981		0,319521399	1,037210029	2,567710442		
3	0,887994107	0,118790172	0,901509381		0,202910298	1,449992071	4,017702513		
4	0,706357031	0,34763446	1,249143813		0,275110822	1,173255702	5,199058215		
5	0,998920204	0,001080379	1,250224192		0,816787918	0,183978004	5,374936219		
6	0,285359542	1,254005343	2,504229535		0,293072189	1,115760293	6,490696512		
7	0,71201352	0,339658379	2,843887914		0,137924153	1,800955786	8,291652298		
8	0,695067289	0,363746619	3,207634533		0,177181899	1,573253088	9,864905386		
9	0,383131575	0,959376812	4,167011344		0,490920213	0,646794239	10,51169962		
10*	0,759480021	0,275112162	4,442132607		0,442506263	0,741182419	11,25282024		
11*	0,698349439	0,359035673	4,80116828		0,060268553	2,553586197	13,80646824		
12*	0,516391639	0,660889815	5,462058091		0,695429355	0,330205315	14,13667356		
13*	0,084551549	2,470393884	7,932451976		0,672166384	0,361135794	14,49780935		
14*	0,922743416	0,080404072	8,012856048		0,107281114	2,02936605	16,5271754		
15*	0,536793467	0,622141683	8,634997911		0,234827967	1,317183715	17,84435912		
16*	0,683086114	0,381134346	9,016132256		0,522356167	0,590368738	18,43472785		
17*	0,734663742	0,308342379	9,324474635		0,00711789	4,49558384	22,93031324		
18*	0,065606035	2,724087589	12,04856222		0,021809752	3,477634616	26,40794785		
19*	0,178758964	1,721716952	13,77027918		0,135614602	1,816307476	28,22425533		
20*	0,538244808	0,619441788	14,38972096		0,296575003	1,104959209	29,32921454		
21*	0,92428958	0,078145797	14,46786676		0,230974006	1,33227366	30,6614419		
22*	0,571644757	0,559237535	15,0271043		0,897498275	0,988312801	30,75975471		
23*	0,057883943	2,849315255	17,87641955		0,513670597	0,605611892	31,3653666		
24*	0,750335687	0,287234591	18,16365414		0,361302241	0,925491307	32,2908579		
25*	0,496865365	0,699459753	18,86311389		0,767112548	0,241019774	32,53187768		
26*	0,090907199	2,397916082	21,26102998		0,803401428	0,19900709	32,73087839		
27*	0,131174381	2,03122769	23,29225767		0,998592447	0,001280495	32,73215888		
28*	0,851756029	0,160455144	23,45271281		0,302429418	1,087188507	33,81934739		
29*	0,538789346	0,618430607	24,07114342		0,288407013	1,130347779	34,94969517		
30	0,051844256	2,959511139	27,03065456		0,026435746	3,302761973	38,25245714		
31	0,685667236	0,377362848	27,40801741		0,25722867	1,234354386	39,48681153		
32	0,663793344	0,409784407	27,81780181		0,488743856	0,650833399	40,13764493		
33	0,21724197	1,526742431	29,34454424		0,504696894	0,6216333854	40,75927878		

Tempo di Attesa in Coda		Modello analitico di riferimento		passo
Valore Atteso Calcolato	E[W]=p/(μ*(1-p))=	9.090909091		
Media su 3200 Osservazioni				
Tempo di Attesa in Coda	13,01393943	13,01393943		
Tempi di Soggiorno	Tempi di Attesa in Coda	Equazione di Lindley W _i		
0,866715914	0	0	Pearson	
1,784991261	0,747781	0,747781232	Correlazione	
3,11619316	1,666201	1,666201089	-0,152677476	-0,152677476
3,941814402	2,768559	2,7685587	-0,026784313	-0,008168125
4,124712027	3,940734	3,940734023	-0,026082237	0,01868795
3,986466977	2,870707	2,870706684	-0,023201384	-0,023201384
5,447764384	3,646809	3,646808598	0,002299292	0,002299292
6,657270853	5,084018	5,084017765	-0,016934534	-0,016934534
6,344688281	5,697894	5,697894041	-0,027880259	-0,027880259
6,810749438	6,069567	6,069567018	-0,03519187	-0,03519187
9,005299962	6,451714	6,451713765		
8,674615465	8,34441	8,34441015		
6,565357375	6,204222	6,204221581		
8,514319352	6,484953	6,484953303		
9,209361204	7,892177	7,892177489		
9,418595597	8,828227	8,828226859		
13,6058386	9,110253	9,110253218		
14,35938563	10,881751	10,88175101		
14,45397615	12,637669	12,63766868		
14,93949357	13,834534	13,83453437		
16,19357514	14,861348	14,86134778		
15,73265041	15,634338	15,63433761		
13,48894705	12,883335	12,88333515		
14,12720376	13,201712	13,20171246		
13,66876378	13,427744	13,42774401		
11,46984841	11,270848	11,2708477		
9,439901216	9,438621	9,438620721		
10,36663458	9,279446	9,279446072		
10,87855175	9,748204	9,748203972		
11,22180258	7,919041	7,919040612		
12,07879412	10,84444	10,84443974		
12,31984311	11,66901	11,66900972		
41,41473451	10,70210068	10,70210068		

Lindley: relazione ricorsiva tra il tempo di attesa del n-esimo e il tempo del (n-i)-esimo. Ad esempio relazione tra tempo di attesa della 10° persona (ultima) che attende il servizio e della 9°

Calcolo i tempi di attesa mediante l'equazione di Lindley

I tempi di soggiorno sono variabili aleatorie, il primo valore coincide con il servizio. Immagino al sequenza temporale di soggiorno (X₁, X₂...X_n), le X sarebbero le realizzazioni dei tempi di soggiorno a partire dal primo. Esempio: ogni mattina vado alla posta a vedere quanto tempo aspetta la terza persona in fila. Il terzo di ogni mattina rappresenta un campione, dunque soggiorno della terza persona.

per ogni variabile aleatoria viene generata una realizzazione indipendente corrispondente ogni volta che si fa un run

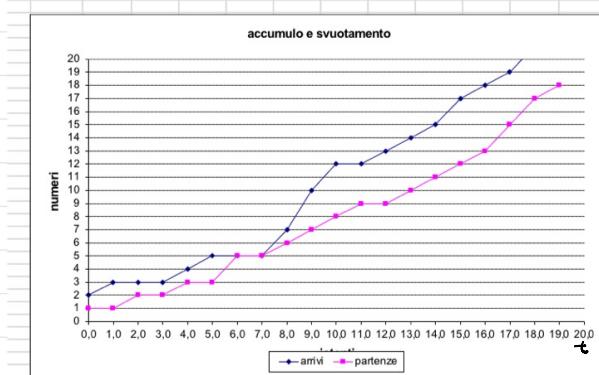
- Potrebbe servire la statistica? La realizzazione di S3 rispetto a quella di S5 è indipendente? O meglio, la sequenza dell'uscita del sistema di osservazioni è indipendente? Gli input per ipotesi sono indipendenti. La realizzazione di S5 è correlata a quella di S3. Le due realizzazioni sono identicamente distribuite? Concetto di stazionarietà (un qualcosa poi non cambia più)

se dovessi studiare questo sistema, spero che per $n \rightarrow \infty$ ci sarà una sola variabile aleatoria, ossia "soggiorno n-esimo", dopo 1000 o più realizzazioni (a lungo termine) diventerà una sola realizzazione. "Ne studio uno ed è come se li studiassi tutti"

- Come stimo la varianza? Se implemento la varianza campionaria sorge il problema della indipendenza e della correlazione (che si vedrà in futuro)

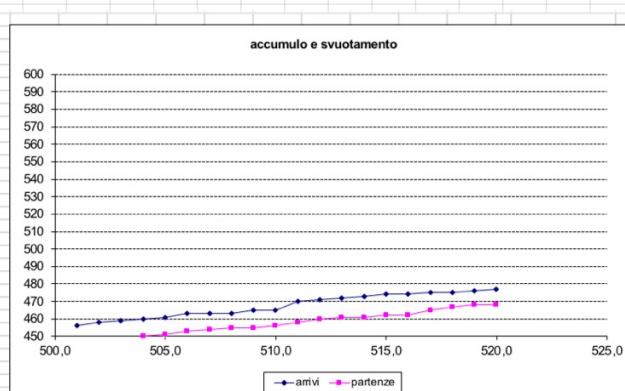
- Mi focalizzo sul tempo di attesa in coda (è una media aritmetica). Se clicco sulla sua casella mi rendo conto che salto da circa 10 a circa 5 con una "pigiamata", quindi che varianza c'è? Come la stimo? Che "fiducia" posso dare a quel valore? Nascono le stime intervallari (intervallo di confidenza). Esse vengono utilizzate per stimare delle prestazioni. Per esempio ad un indice di prestazione (variabile aleatoria) come il tempo di risposta, quale valore atteso minimo "posso chiedere"?

Simulazione Monte Carlo: riproduco la storia degli stati. Di segui le traiettorie che rappresentano arrivi e partenze:



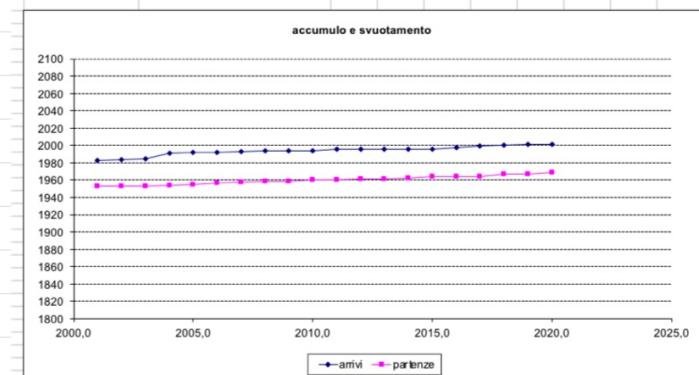
$0 \leq t \leq 20$ (inizio)

Simulo la traiettoria del processo stocastico degli arrivi e quello delle partenze. Inizialmente, non posso sapere se il sistema è congestionato o no



$500 \leq t \leq 525$ (metà)

Osservo il sistema a lungo termine

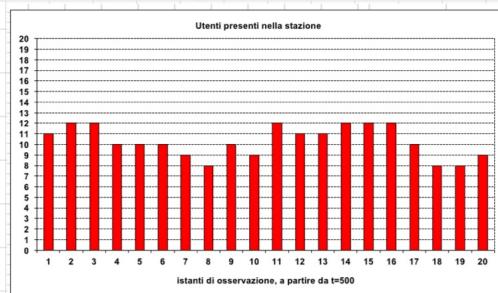
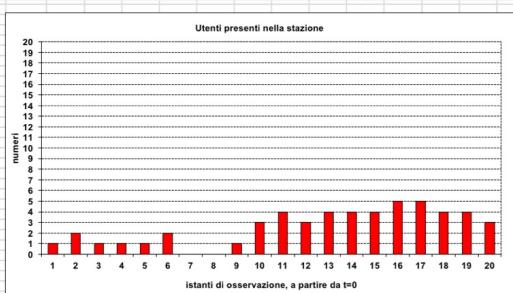


$2000 \leq t \leq 2025$ (verso la fine)

Quando il sistema arriva a regime il numero di arrivi e partenze tende a uguagliarsi

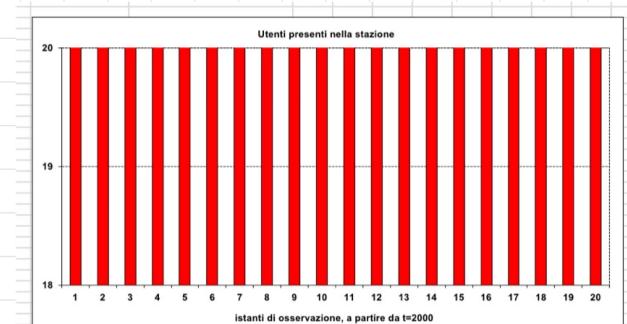
Arrivi (x_0, y_0) – Partenze(x_0, y_0) = numero di persone coinvolte

Tasso Interarrivi	1	Tasso di Servizio	1,1
Valore Atteso in Input		Valore Atteso in Input	
Interarrivi ($1/\lambda$)	1	Ordine	Tempo di Servizio ($1/\mu$)
Media su 3200 Osservazioni		FIFO	0,909090909
Interarrivi	0,958853087	Media su 3200 Osservazioni	
		Tempo di Servizio	0,907979337

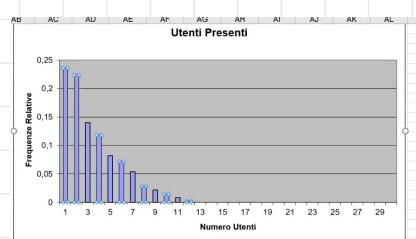


In questo caso, $\text{TASSO DI SERVIZIO} > \text{TASSO INTERARRIVI}$
→ 10% più veloce

Cosa accade se cambio ciò?



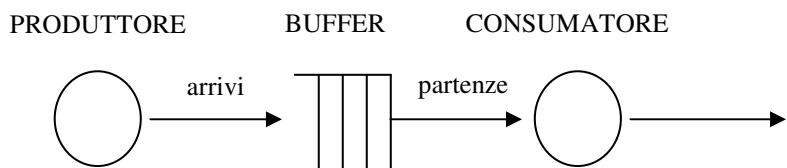
A lungo andare se non "chiudo" il sistema ci sarà solo congestione



Il grafico rappresenta proprio il modello geometrico di occupazione del buffer

Analisi Monte Carlo della coda di accesso ad una risorsa.

In questa sezione si vuole fare riferimento ad un modello gestionale classico composto da due risorse fisiche - un produttore ed un consumatore - separate da un buffer di accumulo degli “oggetti” in transito (clienti nel linguaggio dei modelli, semilavorati nella pratica, ad esempio). Le risorse sono anche dette serventi perché modellano macchine capaci di effettuare un servizio (lavorazione sui semilavorati). I serventi hanno una velocità finita e limitata e la durata di ogni servizio può essere intesa come la realizzazione di una variabile aleatoria afferente ad una distribuzione probabilistica nota. L’aleatorietà del servizio – presso entrambe le risorse - può essere attribuita tanto al servente quanto, più spesso, alla richiesta di servizio effettuata dal cliente. Quale che sia il caso, questa aleatorietà è alla base della formazione della coda d’attesa degli utenti in transito che, vengono accomodati nel buffer intermedio e lì attendono il proprio turno. In particolare la coda può essere limitata o illimitata e qui si adotterà la politica detta di “push”, ovvero il produttore produce/spinge oggetti nel buffer indipendentemente dal consumatore e dallo stato dello stesso buffer. In tal caso, quando i posti in coda risultano tutti occupati, si pone il problema di rigettare e perdere il cliente in arrivo dal produttore, oppure di bloccare il produttore stesso. Si tenga presente che la sola coppia “buffer-risorsa a valle” è anche nota come “stazione di servizio” nella Teoria delle Code classica. Di seguito è raffigurato il sistema produttore - buffer - consumatore:



Osservazione: la coda può crescere all’infinito, se l’intensità degli arrivi è “eccessiva” rispetto alla velocità del servente!

Nel sistema semplificato che vogliamo analizzare il buffer è praticamente inesauribile e gli utenti sono prodotti, “spinti” nel buffer e consumati uno alla volta, sempre secondo l’ordine d’arrivo (disciplina FIFO: First In - First Out). In particolare, vogliamo calcolare

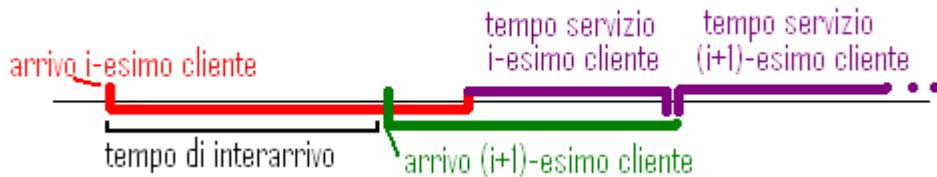
il tempo medio di attesa in coda dei clienti, assumendo che entrambi i serventi siano in una condizione iniziale di ozio e nessuno dei due possa essere soggetto a guasti né ad interruzioni di qualunque sorta, una volta avviato un servizio. Valuteremo anche l'andamento nel tempo del livello di occupazione del buffer e il livello medio su un orizzonte temporale fissato.

Definiamo le seguenti quantità:

W_i = tempo di attesa dell' i -esimo cliente

S_i = tempo di servizio dell' i -esimo cliente

A_i = intervallo di tempo tra l'arrivo dell' i -esimo cliente e quello dell' $(i+1)$ -esimo – in linguaggio “tecnico” *interarrivo* dell' $(i+1)$ -esimo cliente.



Il primo cliente che arriva attende un tempo nullo, dunque per lui: $W_1 = 0$. Allora, il tempo di attesa di ogni cliente che arriva è dato dalla seguente formula ricorsiva, nota come equazione di *Lindley*:

$$W_{i+1} = \max(0, W_i + S_i - A_i)$$

ovvero il tempo di attesa dell' $(i+1)$ -esimo cliente è pari alla somma di tempo d'attesa e tempo di servizio dell' i -esimo cliente meno il tempo di interarrivo dell' $(i+1)$ -esimo cliente.

La formula suddetta restituisce il valore zero quando l'istante di arrivo del cliente $(i+1)$ -esimo risulta successivo all'istante di fine servizio del cliente i -esimo.

Tramite il metodo Monte Carlo generiamo due serie di numeri casuali da cui, applicando le opportune trasformazioni inverse, ricaviamo i valori A_i e S_i .

Immaginiamo che tempi di servizio e di interarrivo seguano entrambi una legge esponenziale:

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

e

$$F_Y(x) = 1 - e^{-\lambda x}$$

dove X rappresenta la variabile aleatoria “tempo di interarrivo” con tasso λ e Y la variabile aleatoria “tempo di servizio” con tasso μ .

I due numeri casuali generati dal metodo Monte Carlo sono rispettivamente i valori di $F_X(x)$ e di $F_Y(x)$. Per ricavare per entrambe la x (ribadiamolo: corrisponde ora alla durata dell’interarrivo, ora alla durata del servizio), ci avvaliamo del metodo della trasformazione inversa, come già fatto nell’esempio precedente.

Posto il tempo di attesa del primo cliente pari a zero, potremo calcolare tutti gli altri tempi d’attesa con la formula di *Lindley*.

Prima di passare ai grafici ottenuti col metodo Montecarlo in Excel, è utile riportare un’osservazione. Accettando, in maniera intuitiva, che il numero medio di arrivi in un intervallo temporale di ampiezza “ $t - 0 \hat{=} t$ ” (ovvero t meno l’istante arbitrariamente fissato come iniziale) sia pari al prodotto $\lambda \cdot t$ (d’altra parte il tasso degli arrivi (λ) è costante!), introduciamo la variabile aleatoria N_S per indicare il numero medio di arrivi durante un servizio del consumatore, a sua volta rappresentato dalla variabile aleatoria S , distribuita ancora secondo una legge esponenziale, ma di parametro μ .

Allora, si può osservare che risulta:

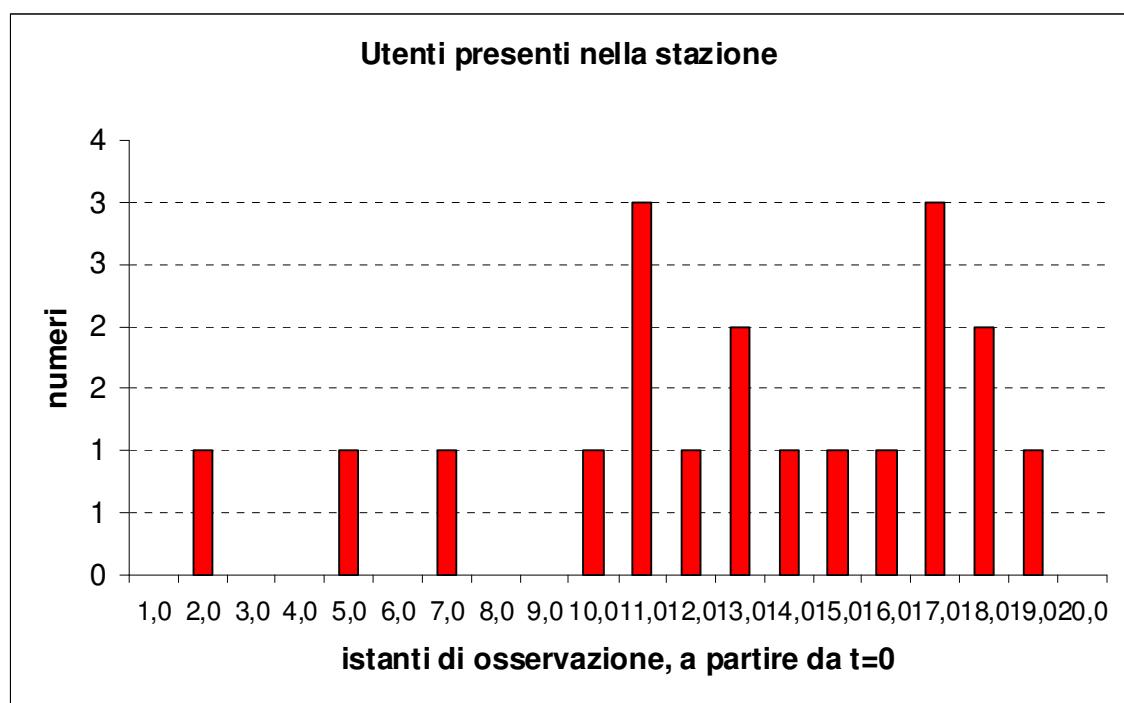
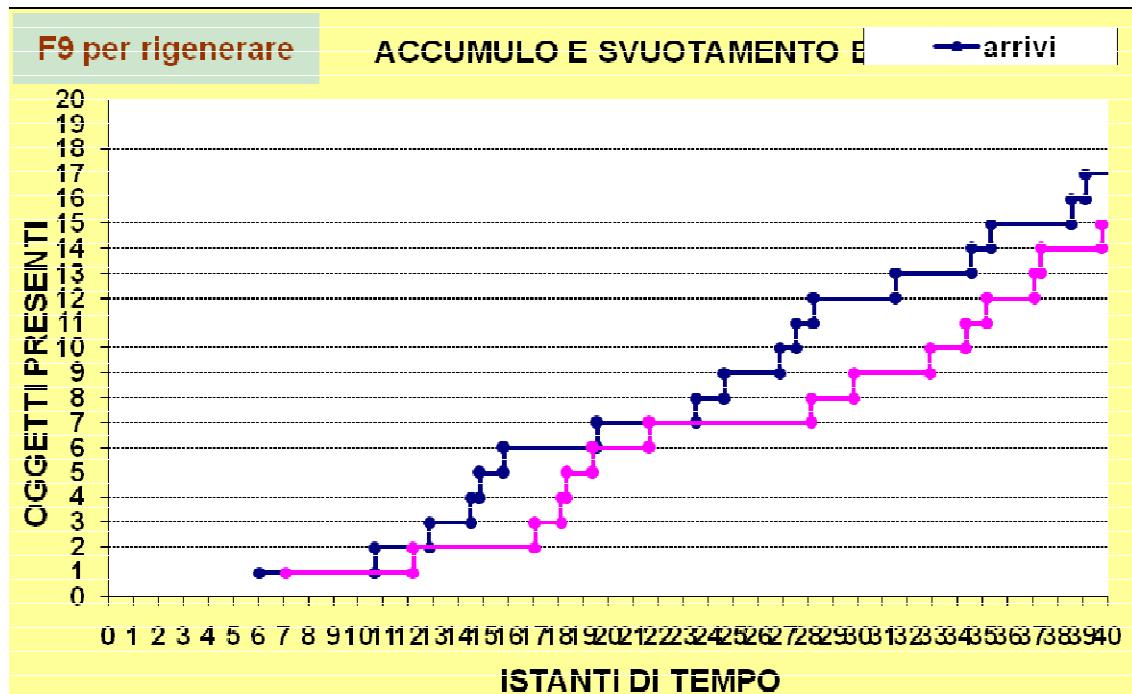
$$E[N_S] = \int_{t=0}^{\infty} E[N_S | S = t] \Pr\{S \in [t, t + dt]\} = \int_{t=0}^{\infty} \lambda t \exp\{-\mu t\} dt = \frac{\lambda}{\mu}$$

E ci si può chiedere: cosa succede quando si verifica la condizione $\lambda / \mu > 1$?

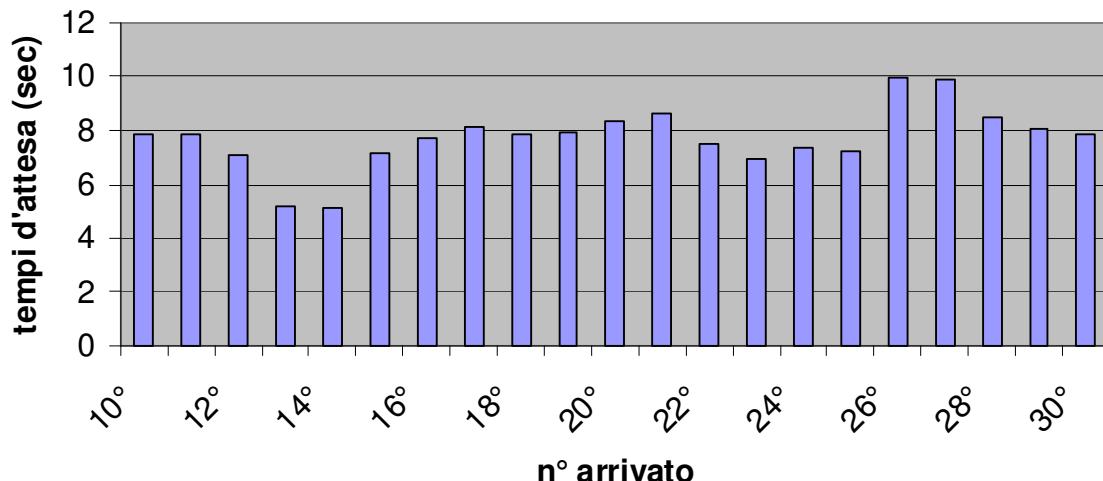
E quando, invece, risulta $\lambda / \mu < 1$?

E varrebbe la pena, nella pratica, di considerare il caso limite/separatore $\lambda / \mu = 1$?

In attesa di dare le risposte corrette, si può passare alla pagina successiva che riporta due grafici costruiti con Excel (vedere fogli Excel Produttore-Consumatore.xls).



Tempi di attesa in coda (dal 10° al 30° utente)



Tempi di attesa in coda

Il TLC per la “media campionaria”

Nell’analisi statistica del modello PRODUTTORE - CONSUMATORE, è assai utile poter assumere che una realizzazione della media campionaria corrisponda alla media aritmetica di una sequenza di realizzazioni indipendenti della stessa variabile aleatoria, X , magari rilevate con osservazioni sperimentali, indipendenti (runs del Metodo Monte Carlo).

Infatti, particolarizzando il teorema limite centrale a questo caso, risulta:

in giallo sarebbe la standardizzazione della media campionaria $\tilde{Z}_n = \frac{\bar{X}(n) - E[\bar{X}(n)]}{\sqrt{VAR[\bar{X}(n)]}} = \frac{\bar{X}(n) - \mu}{\sqrt{\sigma^2 / n}}$ ricordando che il TCL riguarda la somma di variabili aleatorie.

che tende alla normale standard per $n \rightarrow \infty$ e, di conseguenza:

$$\bar{X}(n) = \frac{\sigma}{\sqrt{n}} \tilde{Z}_n + \mu \quad \text{tende ad essere distribuita come una legge normale di media } \mu \text{ e varianza } \sigma^2/n \text{ al crescere di } n$$

deviazione standard

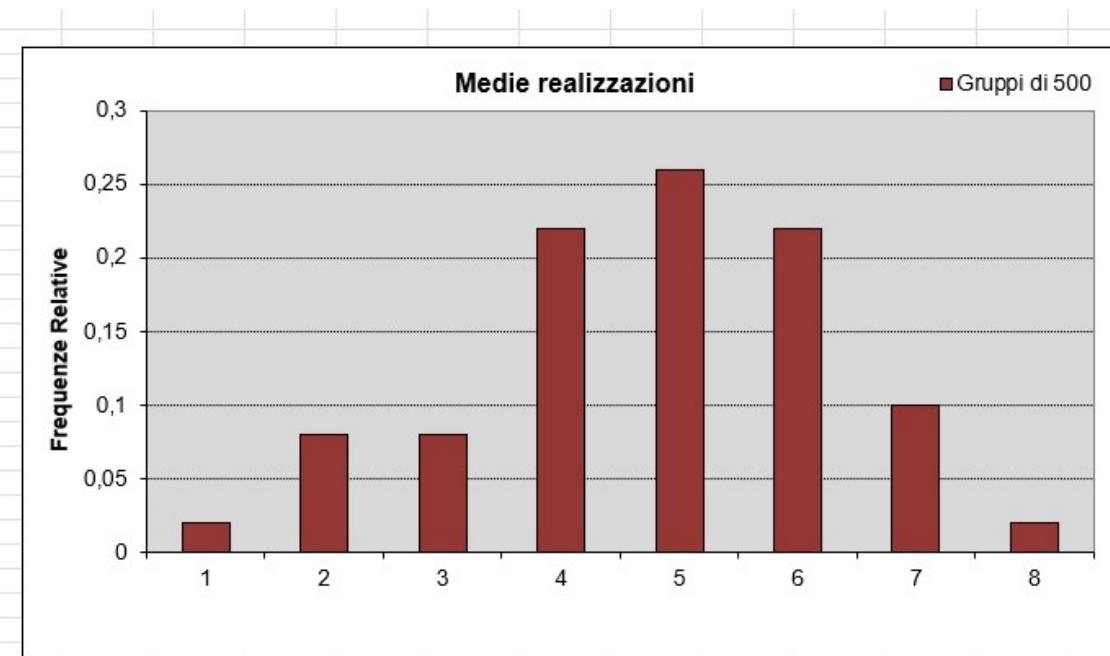
In particolare, si verifica sperimentalmente che la forma della $\bar{X}(n)$ diventa approssimabile ai fini pratici con la forma normale per $n > 30$

Per il modello produttore consumatore

Sperimentalmente aprodo il foglio excel: normalità della media campionaria si verifica il teorema del limite centrale

TESI:
La media campionaria tende a essere distribuita come una gaussiana centrata sul valore medio della legge esponenziale

Attraverso le generazioni Montecarlo, l'obiettivo di tale esperimento è stato quello di cercare di convalidare il teorema del limite centrale. Nella pratica, aumentando le n (dimensione del campione) ho cercato di approssimare fedelmente la "campana di Gauss". Con un numero di medie campionarie pari a 200 e poi ancora più con un numero di campioni pari 500, riesco a verificare sperimentalmente il suddetto teorema, ottenendo nelle varie simulazioni, sempre risultati più o meno attendibili. Ovviamente per campioni sempre maggiori l'attendibilità del teorema risulterà ancora più evidente.



Appunto più sale il gruppo di medie più si tende alla gaussiana. Oppure si poteva salire con il numero di realizzazioni!

La media di medie è detta grande media, il cui valore atteso della grande media è pari alla somma tra il valore atteso delle singole medie.

L'intervallo di confidenza per la media (valore atteso)

Siano X_1, X_2, \dots, X_n variabili aleatorie tutte indipendenti e identicamente distribuite

$\approx N(\mu, \sigma^2)$, allora:

$$\bar{X}(n) \hat{=} \frac{\sum_{i=1}^n X_i}{n} \approx N\left(\mu, \frac{\sigma^2}{n}\right)$$

Ricorda che il valore atteso non è una media di per sé, ma perché la media aritmetica lo stima!

per il “teorema di riproducibilità” della normale.

Rammentando che: (grazie all'indipendenza)

$$Z \hat{=} \frac{\bar{X}(n) - \mu}{\sigma / \sqrt{n}} \approx N(0,1) \quad \leftarrow$$

se inserisco S (deviazione della var. camp.) al posto della deviazione standard σ ottengo un rapporto tra variabili aleatorie e perciò una nuova variabile aleatoria

e che:

$$\Pr_{\text{quantile}}[-z_{\alpha/2} \leq Z \leq z_{\alpha/2}] = 1 - \alpha \quad \leftarrow$$

dove:

$$\alpha/2 = \int_{-\infty}^{z_{\alpha/2}} f_Z(z) dz = \int_{z_{\alpha/2}}^{+\infty} f_Z(z) dz$$

significa:
probabilità che la realizzazione di Z sia compresa tra i quantili è abbastanza vicino a 1

si ottiene: (isolando μ)

$$\Pr\left[\bar{X} - \frac{\sigma}{\sqrt{n}} z_{\alpha/2} \leq \mu \leq \bar{X} + \frac{\sigma}{\sqrt{n}} z_{\alpha/2}\right] = 1 - \alpha$$

In pratica, è stato stabilito che l'intervalle aleatorio:

$$\left[\bar{X} - \frac{\sigma}{\sqrt{n}} z_{\alpha/2}, \bar{X} + \frac{\sigma}{\sqrt{n}} z_{\alpha/2} \right]$$

(IC) Da notare che l'intervalle è simmetrico!

contiene, con probabilità $1 - \alpha$, il valore (incognito) del parametro .

Ogni realizzazione (intervallo numerico) di (IC) può essere considerata una stima di , ma non si può dire che essa contiene e tanto meno che il valore corrispondente a è il centro dell'intervalle: ogni punto dell'intervalle ha la stessa probabilità di essere .

(IC) è detto intervallo di confidenza al $100(1 - \alpha)\%$ (livello di confidenza).

Infine, posto $d \hat{=} z_{\alpha/2} \sigma / \sqrt{n}$, si tenga presente che la numerosità (n) del campione richiesto per stimare con un intervallo di ampiezza $2d$ risulta:

$$n = \frac{(z_{\alpha/2})^2 \sigma^2}{d^2}$$

OSSERVAZIONE: Per ottenere una realizzazione dell'intervallo di confidenza sulla media del processo, occorre una stima della varianza del processo!

Importanza dell'intervallo di confidenza:

- introduce l'idea di stimare un parametro non più attraverso la realizzazione di una variabile aleatoria, bensì attraverso la realizzazione di un intervallo aleatorio che contiene, con la probabilità voluta, il parametro stesso.

L'intervallo però non deve essere né troppo grande né troppo piccolo.

Qualità dell' intervallo:

- fissato il livello di confidenza, $100(1-\alpha)\%$, e con una certa deviazione standard (σ), intervalli migliori (cioè più ristretti) possono essere ottenuti solo aumentando di parecchio la numerosità (n) del campione.

Problemi d'uso:

- La stima per intervallo di un parametro del 1° ordine, quale la "media", richiederebbe la conoscenza di un parametro del 2° ordine, quale la "deviazione standard".
- Sembrerebbe valido solo nell'ipotesi che il parametro-media da stimare sia quello di una legge normale.

Soluzione dei problemi:

- Si può usare la varianza campionaria al posto della varianza vera e, sfruttando il teorema limite centrale, si può dimostrare che:

$$\Pr\left[\bar{X} - \frac{S}{\sqrt{n}} z_{\alpha/2} \leq \bar{X} \leq \bar{X} + \frac{S}{\sqrt{n}} z_{\alpha/2}\right] \cong 1 - \alpha \quad \text{per } n \rightarrow \infty$$

quando manca l'ipotesi di normalità delle variabili i.i.d. X_1, X_2, \dots, X_n .

L'intervallo: $\left[\bar{X} - \frac{S}{\sqrt{n}} z_{\alpha/2}, \bar{X} + \frac{S}{\sqrt{n}} z_{\alpha/2}\right]$ è considerato una buona approssimazione dell'intervallo vero già con $n \geq 30$.

NOTA:

Nella pratica si usa fissare il livello di confidenza al 90% o 95% a ciò corrisponde: $\alpha = 0.10$ $\alpha = 0.05$ $z_{\frac{\alpha}{2}} = 1.645$ $z_{\frac{\alpha}{2}} = 1.960$

La statistica “varianza campionaria” 1(2)

Siano X_1, X_2, \dots, X_n I.I.D. $E[X_i]$ è il valore atteso della popolazione da cui è estratto il campione

con: $E[X_i] \hat{=} \mu, \quad Var[X_i] \hat{=} \sigma^2, \quad i = 1, \dots, n$

DA: $\sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{i=1}^n (X_i - \mu)^2 - n(\bar{X} - \mu)^2 \quad (\bar{X} \hat{=} \frac{1}{n} \sum_{i=1}^n X_i)$

SI OTTIENE: $E \left[\sum_{i=1}^n (X_i - \bar{X})^2 \right] = (n-1)\sigma^2$

I passaggi algebrici sono stati indicati nella dispensa del corso, parte II, “Analisi Statistica”, pag. 24 e 25, ma NON fanno parte del programma di esame!

Si userà come stima della varianza S al posto di sigma

IN CONCLUSIONE: $E \left[\frac{1}{(n-1)} \sum_{k=1}^n (X_k - \bar{X})^2 \right] = \sigma^2$

“n-1” e non “n” perchè per ottenere una stima più accurata della varianza usiamo un “aggiustamento statistico”

se calcolo la varianza campionaria su molti campioni diversi le media di questa varianza campionaria si avvicinerebbe sempre di più alla vera varianza della popolazione

NON SI TRATTA DI GAUSS

Osservazione: La varianza campionaria è la stima puntuale della varianza.

Nella parentesi quadrata c’è uno stimatore corretto del parametro varianza, cioè la statistica “varianza campionaria”!

Formula alternativa usata in Excel per esprimere lo stimatore della varianza

https://www2.isye.gatech.edu/people/faculty/David_Goldsman/

$$\begin{aligned} S^2 &= \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1} \\ &= \frac{\sum_{i=1}^n (X_i^2 - 2\bar{X}X_i + \bar{X}^2)}{n - 1} \\ &= \frac{\sum_{i=1}^n X_i^2 - 2\bar{X} \sum_{i=1}^n X_i + n\bar{X}^2}{n - 1} \\ &= \frac{\sum_{i=1}^n X_i^2 - 2n\bar{X}^2 + n\bar{X}^2}{n - 1} \\ &= \frac{\sum_{i=1}^n X_i^2 - n\bar{X}^2}{n - 1} \end{aligned}$$

La statistica “varianza campionaria” 2(2)

In base al precedente risultato, la variabile aleatoria:

$$S^2 \triangleq \frac{1}{(n-1)} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{\sum_{i=1}^n X_i^2 - n\bar{X}^2}{n-1}$$

è detta “varianza campionaria” e viene usata per “stimare” “puntualmente” il parametro σ^2 grazie alla seguente (che non dimostreremo):

$$Var[S^2] = \frac{2\sigma^4}{n-1} \xrightarrow{n \rightarrow \infty} 0 \quad \text{per } n \rightarrow \infty$$

quadrato della varianza!

Osservazione:

Lo stimatore corretto del parametro varianza migliora al crescere di “n” (perché si riduce la sua varianza)

Però, purtroppo, questo stimatore ha una varianza sua che è pari al quadrato della varianza che deve stimare!

PRODURRE STIME INTERVALLARI

casuale	LAMBDA	ESPOENZIALE	scarti	scarti^2	stima della media e della varianza	n	n-1	alfa	alfa/2	1-alfa/2
0,907059	0,25	B - tempo di soggiorno	A	A	A	100	99	0,05	0,025	0,975
0,071457	MEDIA	9,503177	5,649488	31,91671	stima della media e della varianza					
0,475876	3,853689	0,296555	-3,55713	12,6532	A					
0,620852		2,58411	-1,26958	1,61183	Intervallo di confidenza per la media			$z_{\alpha/2}$	1,96	
0,881526	SOM SCA	3,879311	0,025624	21,88896						
0,242467	2152,032	8,532253	4,678564	21,88896						
0,62759		1,110751	-2,74294	7,523707						
0,94516		3,951039	0,097349	0,009477						
0,210836		11,61336	7,759661	60,21245						
0,315209		0,947125	-2,90565	8,448114						
0,750039		1,154568	-2,33912	5,471486						
0,293832		5,545806	1,692117	2,863259						
0,202667		1,39161	-2,46208	6,061832						
0,146378		0,905933	-2,94776	8,689265						
0,844105		0,633068	-3,22062	10,3724						
0,571085		7,434278	3,580589	12,82061						
0,3715		3,385983	-0,46771	0,218749						
0,997946		1,857675	-1,99601	3,984071						
0,815052		24,75135	20,89767	436,7124						
0,010066		6,750719	2,89703	8,392782						
0,256869		0,040468	4,767497							
0,464483		1,187533	-2,66616	7,10839						
		2,49809	-1,3556	1,83765						

LIVELLO DI "CONFIDENZA" o "FIDUCIA" (DALL'INGLESE "CONFIDENCE")

Il valore atteso del tempo di soggiorno, in precedenza era pari a: 11.055934. Potrebbe cambiare rendendo più complicato il modello se vengono applicate politiche differenti (round-robin, ecc.) In questi casi occorre fare delle simulazioni

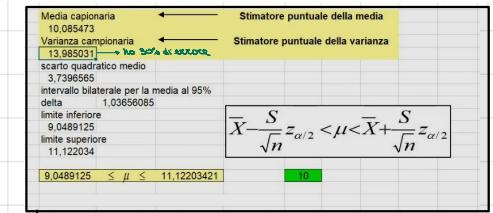
In questo caso specifico, non sono sicuro di trovarmi di fronte ad un errore quindi faccio una stima intervallare.

Se volessi cambiare i dati, l'unico valore che potrei variare è n altrimenti cambierei il problema.

$z_{\alpha/2}$ valori della Gaussiana in corrispondenza del quantile $\frac{\alpha}{2}$
1.96 è l'ascissa entro cui realizzo $\frac{\alpha}{2}$

Non so come si produce una realizzazione da una normale perché non vale il metodo delle trasformazioni inverse. Utilizzo il metodo implementato da Excel

aumento la varianza da 10 a 20 per aggiornare i valori



Gli intervalli di confidenza NON SONO simmetrici ma quello della media lo è. Ha come variabile la gaussiana Z

Quando si valuta il grado di confidenza entra in gioco una statistica a prescindere dalle realizzazioni con cui lavoro. Per il valore atteso è la gaussiana, per la varianza è la chi-square.

Devo dimostrare che con probabilità 0.95 ottengo l'intervallo che contiene il valore VERO. Tale intervallo garantisce che al 95% (livello di confidenza) contiene il valore vero.
Dunque il valore vero è equiprobabile e non si trova solo al centro.

$$X - \frac{S}{\sqrt{n}} z_{\alpha/2} < \mu < X + \frac{S}{\sqrt{n}} z_{\alpha/2} \Rightarrow 9,4987172 \leq \mu \leq 10,6164771$$

SOTTO QUALI IPOTESI L'INTERVALLO È CORRETTO?



Quando si valuta il grado di confidenza entra in gioco una statistica a prescindere dalle realizzazioni con cui lavoro. Per il valore atteso è la gaussiana, per la varianza è la chi-square.

L'ampiezza dell'intervallo è proporzionale a "S". Se ho un'ampiezza "larga" e voglio diminuirla posso lavorare solo su "n" ossia sul carico di dati. Data la radice quadrata il numero di dati deve aumentare sostanzialmente

\bar{X} è la media campionaria

$\frac{S}{\sqrt{n}}$ è la deviazione standard della varianza campionaria

$\bar{X} \pm \frac{S}{\sqrt{n}} z_{\alpha/2}$ è l'intervallo aleatorio

La distribuzione chi-quadrato

è una particolare Gamma ricordalo!

Come si è già detto, la distribuzione gamma, ha dato vita ad una particolare distribuzioni, dette chi-quadrato. Andiamo adesso a caratterizzarne la densità della variabile aleatoria chi-square. Considerando una variabile aleatoria X_γ^2 , detta chi-square la densità della chi-square è pari a:

$$f_X(x) = \frac{1}{\Gamma(\gamma/2)2^{\gamma/2}} x^{\frac{\gamma}{2}-1} e^{-\frac{x}{2}} \text{ con } x \geq 0$$

con $X \equiv X_\gamma^2 \equiv x_\gamma^2$ con γ "gradi di libertà". Su Excel la densità gamma è pari a:

$$f_X(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}} \text{ con } \beta = \frac{1}{\lambda} \quad \text{dunque è una gamma con: } \beta = 2 \alpha = \frac{\gamma}{2} \gamma \in [1, 2 \dots n]$$

x_γ^2 individua il punto della semiretta reale a partire dal quale l'area sottesa della densità è proprio γ , ovvero $Pr[X_\gamma^2 \geq x_\gamma^2] = \gamma$ (N.B. sta definizione nel powerpoint non c'è)

Si ricava che la media è la varianza sono pari a rispettivamente a: $E[X] = \gamma$ $Var[X] = 2 \cdot \gamma$
In particolare i gradi di libertà coincidono con il valore atteso. Si vuole dimostrare il seguente teorema: la variabile aleatoria "chi-quadrato" con 1 grado di libertà corrisponde al quadrato della "normale standard".

$$F_{Z^2}(z) \hat{=} \Pr\{Z^2 \leq z\} = \Pr\{-\sqrt{z} \leq Z \leq \sqrt{z}\} \quad \text{Sarebbe la dimostrazione di Trivedi}$$

$$= F_Z(\sqrt{z}) - F_Z(-\sqrt{z}) = 2 \int_0^{\sqrt{z}} \frac{1}{\sqrt{2\pi}} \cdot \exp(-u^2/2) \cdot du$$

$$= 2 \int_0^{\sqrt{z}} \left(1/\Gamma(1/2)\sqrt{2}\right) \cdot \exp(-u^2/2) \cdot du$$

$$\left(\text{con } u \hat{=} \sqrt{v} \Rightarrow du = v^{-1/2} \cdot dv = v^{1/2-1} \cdot dv \right)$$

$$= \int_0^z \left(1/\Gamma(1/2)2^{1/2}\right) \cdot v^{1/2-1} \cdot \exp(-v/2) \cdot dv \quad \text{Ricordando che:}$$

$$f_{x_\gamma^2}(x) = \left(\frac{1}{\Gamma(\frac{\gamma}{2})}2^{\frac{\gamma}{2}}\right) \cdot x^{\frac{\gamma}{2}-1} \cdot e^{-\frac{x}{2}} \implies F_{x_\gamma^2}(z) = F_{Z^2}(z)$$

Teorema di riproducibilità della legge chi-quadrato (non dimostrato)

Siano: $X_{\gamma_1}^2, X_{\gamma_2}^2 \dots X_{\gamma_n}^2$ indipendenti, con i rispettivi gradi di libertà: $\gamma_1^2, \gamma_2^2 \dots \gamma_n^2$ allora:

$X_{\gamma_1}^2 + X_{\gamma_2}^2 + \dots + X_{\gamma_n}^2$ è ancora una chi-quadrato, con gradi di libertà pari a: $\gamma_1, \gamma_2 \dots \gamma_n$

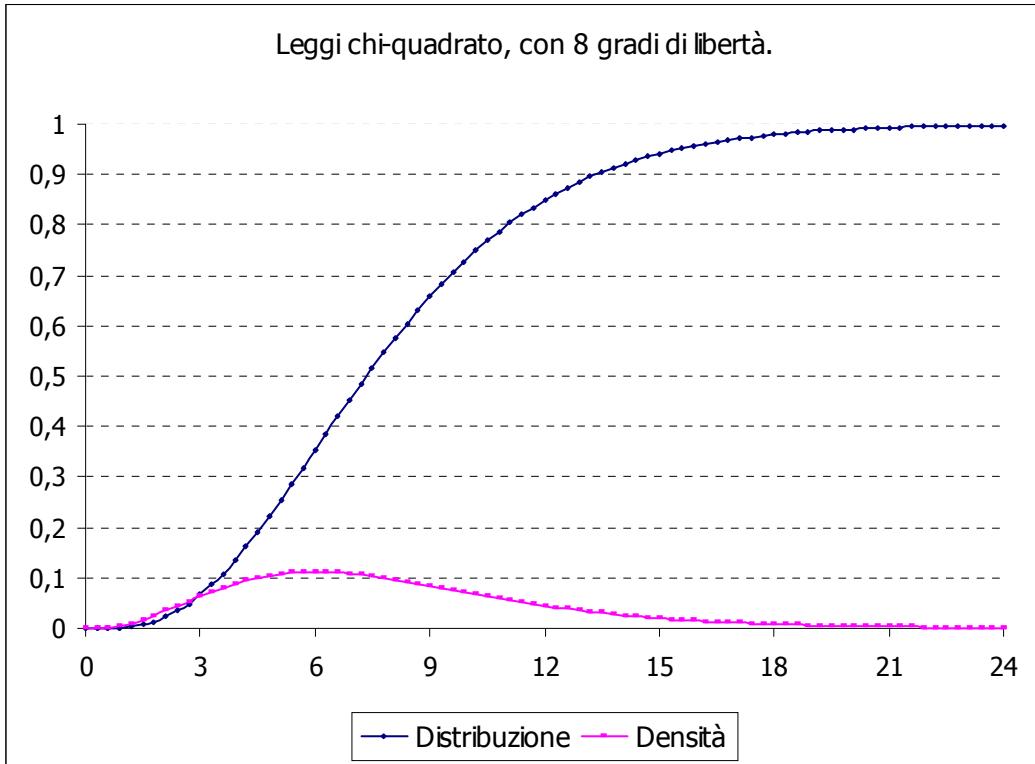
Attraverso le trasformate di Laplace ed al prodotto di convoluzione si prova il teorema.

IMPORTANZA: per "stimare" la varianza e la forma di una distribuzione, a partire da un insieme di realizzazioni sperimentali.

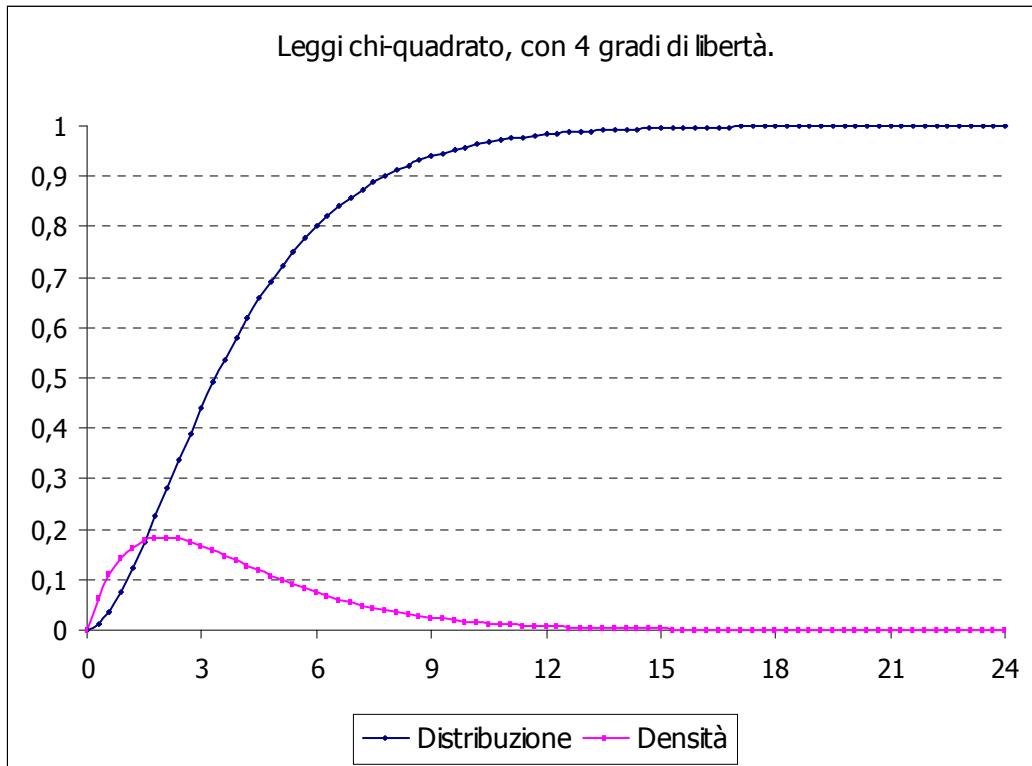
In parole povere il teorema afferma che la somma di chi-quadrato è ancora una chi-quadrato dove i gradi di libertà si sommano. Come detto è importante per "stimare" la varianza e la bontà della forma. (Si vedrà in futuro)(Test sulla forma con chi-square)

Bisognerebbe separare il grafico,
ad esempio il punto di intersezione
non si capisce ed una curva "sovrastra"
l'altra

Rappresentazione della chi-quadrato



da notare anche che non c'è nessun valore negativo. la chi-square è solo positiva!

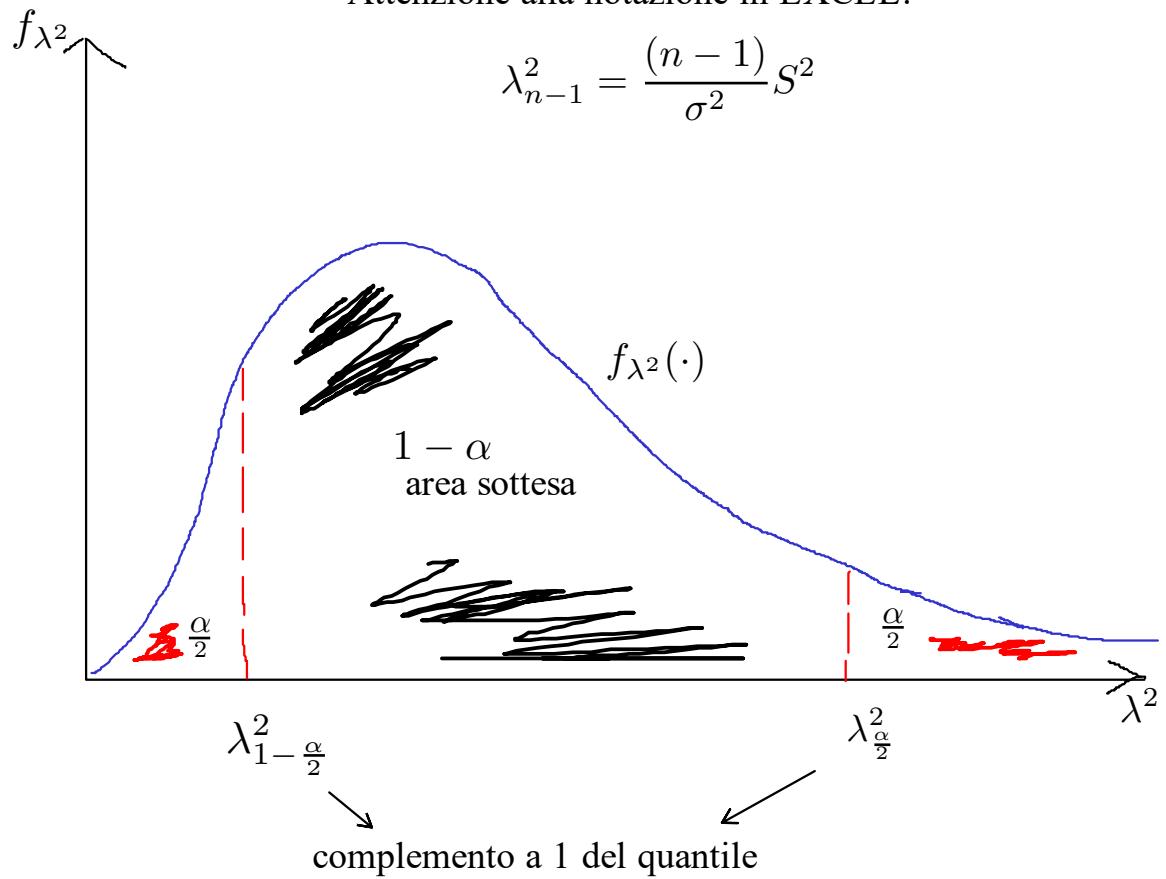


N.B. più aumentano i gradi di libertà più la chi-quadrato tende in forma ad una normale.
Vedi EXCEL

Attenzione alla notazione in EXCEL:

$$\lambda_{n-1}^2 = \frac{(n-1)}{\sigma^2} S^2$$

In rosso si ha l'area pari a $\frac{\alpha}{2}$



complemento a 1 del quantile

$\lambda_{\frac{\alpha}{2}}^2$ sarebbe l'area che si lascia a destra
ovvero il complemento del quantile

L'intervallo di confidenza per la varianza

Ripensando al ragionamento che ha condotto a “scoprire” un intervallo stimatore (detto, poi, intervallo di confidenza) per il parametro σ^2 , si dovrebbe riconoscere che è stata fondamentale la disponibilità di una variabile aleatoria di distribuzione nota che conteneva quel parametro:

$$Z \triangleq \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \approx N(0,1)$$

Dunque, volendo insistere su una strada analoga per individuare un intervallo di confidenza per un altro parametro importante quale la varianza, si deve cercare un'altra variabile aleatoria di distribuzione nota che contenga il parametro varianza e, se possibile, priva di ulteriori parametri che potrebbero creare complicazioni nell'uso pratico dell'intervallo, perché incogniti. Nel linguaggio degli statistici, si direbbe che si sta cercando una “statistica”, cioè una variabile aleatoria capace di produrre una stima.

Col seguente teorema si trova la statistica per la varianza:

Teorema (della chi-quadrato) Stessa ipotesi della media campionaria

Siano X_1, X_2, \dots, X_n indipendenti e identicamente distribuite $\approx N(\mu, \sigma^2)$, allora:

$$\text{statistica} \rightarrow \frac{(n-1)S^2}{\sigma^2} = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X})^2 = \chi^2_{\gamma=n-1} \quad (\text{Tesi}) \quad \begin{matrix} \text{ricordando la varianza campionaria} \\ S^2 = \frac{1}{(n-1)} \sum_{i=1}^n (X_i - \bar{X})^2 \end{matrix}$$

La statistica è: $(n-1)S^2 / \sigma^2$. Infatti, grazie al teorema, si può scrivere:

$$\Pr[\chi^2_{1-\alpha/2} \leq (n-1)S^2 / \sigma^2 \leq \chi^2_{\alpha/2}] = 1 - \alpha$$

ricorda che S è la
stima della varianza

da cui: probabilità che la realizzazione della statistica cada nell'intervallo
isolando sigma

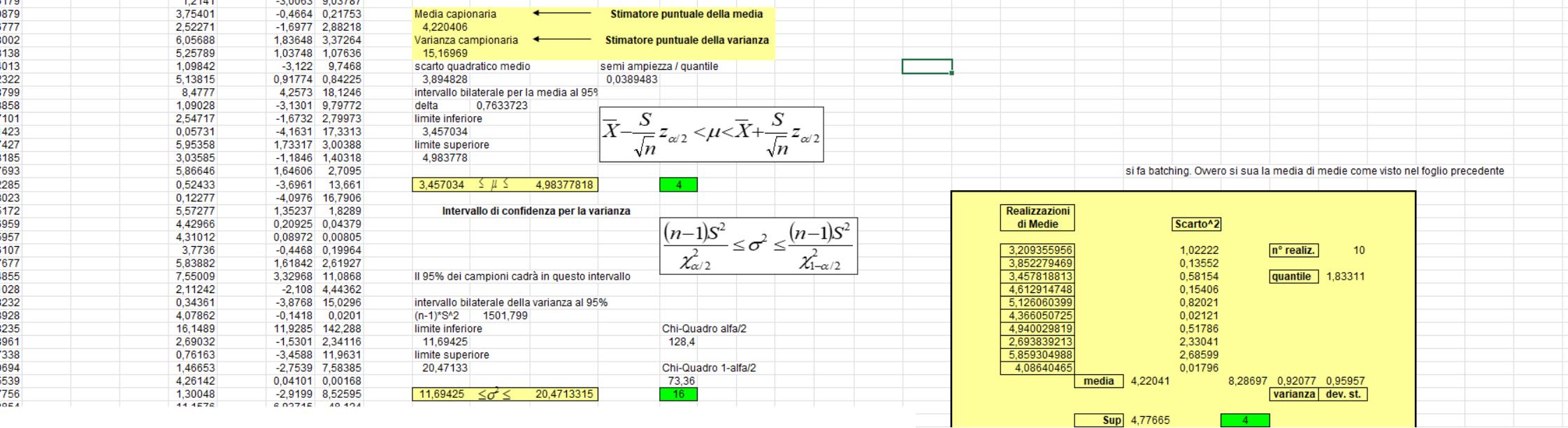
$$\Pr\left[\frac{(n-1)S^2}{\chi^2_{\alpha/2}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{1-\alpha/2}}\right] = 1 - \alpha$$

Quindi,

$$\left[\frac{(n-1)S^2}{\chi^2_{\alpha/2}}, \frac{(n-1)S^2}{\chi^2_{1-\alpha/2}} \right]$$

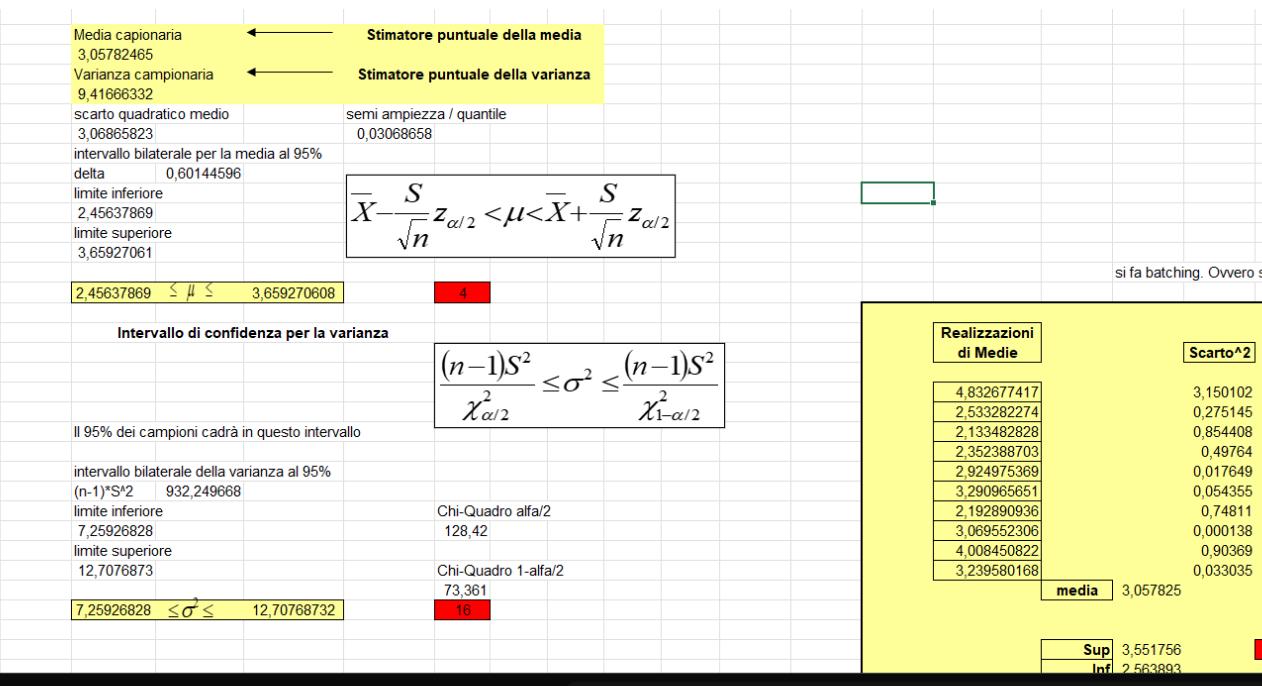
è l'intervallo di confidenza al $100(1-\alpha)\%$ per σ^2 .

Da notare che in particolare gli intervalli di confidenza sulla varianza a differenza di quelli sulla media campionaria, escono "larghi" già in partenza. Si possono stringere quanto si vuole ma resteranno comunque "larghi". Vedi file Excel
INOLTRE NON SONO SIMMETRICI COME NEL CASO DELLA MEDIA CAMPIONARIA
basta guarda i due denominatori dell'intervallo.



si fa batching. Ovvero si usa la media di medie come visto nel foglio precedente

Nello screenshot si nota che gli intervalli sia per media che per varianza contengono il valore vero rispettivo. In particolare l'intervallo per la media è sufficientemente largo, ma non simmetrico(!) dato il caso esponenziale. Mentre il caso della varianza è molto largo. Le stime sono uscite anche relativamente buone. Con un 4.22 a fronte di 4 e 15.19 a fronte di 16.



In questo altro caso invece, si hanno intervalli errati! Per la media si ha una stima di 3.05 a fronte di 4. ed un intervallo "stretto" di : 1.2 e spostato a sinistra del valore reale. Per la varianza si ha una stima di 9.4(!) a fronte di 16. L'intervalllo è largo 5.5 e spostato a destra del valore reale.

STIMATORI CORRETTI, ASINTOTICAMENTE CORRETTI E CONSISTENTI

-) X_1, \dots, X_n v. a. ind. e id. distr. con media μ e var σ^2

$$T(X_1, \dots, X_n) \hat{=} \frac{1}{n} \sum_{i=1}^n X_i \equiv \bar{X}(n) \equiv \bar{X}$$

si puo' omettere di indicare la dipendenza da "n" o "n-1"

$$T(X_1, \dots, X_n) \hat{=} \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \equiv S^2(n-1) \equiv S^2 \text{ stimatore CORRETTO varianza}$$

per n grande il rapporto si semplifica.

quindi asintoticamente corretto

$$T(X_1, \dots, X_n) \hat{=} \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \equiv S^2(n) = \frac{n-1}{n} S^2(n-1)$$

Non corretto sia perché, è stato distorto, ossia "sporcatto" da $\frac{n-1}{n}$

RISULTATI UTILI PER STABILIRE LE PROPRIETA'(o caratteristiche) DEGLI STIMATORI:

$$E[\bar{X}] = \mu, \quad VAR[\bar{X}] = \frac{\sigma^2}{n} \quad E[S^2] = \sigma^2, \quad VAR[S^2] = \frac{2\sigma^4}{n-1}$$

deviazione tra la stima e valore atteso

Lemma di Chebyshev:

$$\Pr \left\{ |T - \theta| \geq n \cdot \sqrt{Var[T]} \right\} \leq \left(1/n^2 \right)^{\text{di quante deviazioni ti allontani dalla deviazione standard?}}$$

$n < 1$ non ha senso perché le prob. SONO 1

Probabilità che la deviazione sia maggiore di n volte della deviazione standard.

Lo stimatore $S^2(n-1)$ è corretto

Dimostrazione

$$T(X_1, \dots, X_n) \hat{=} \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \equiv S^2(n-1) \equiv S^2$$

Mi gioco l'ipotesi che le variabili aleatorie sono indipendenti ed identicamente distribuite. Vale perciò che il valore atteso di una somma è pari alla somma dei valori attesi.

$$\begin{aligned}\mathbb{E}[S^2] &= \mathbb{E}\left[\frac{\sum_{i=1}^n X_i^2 - n\bar{X}^2}{n-1}\right] = \frac{\sum_{i=1}^n \mathbb{E}[X_i^2] - n\mathbb{E}[\bar{X}^2]}{n-1} \\ &= \frac{n}{n-1} \left(\mathbb{E}[X_1^2] - \mathbb{E}[\bar{X}^2] \right) \quad \text{si usa: } VAR[X] = E[X^2] - E[X]^2 \\ &= \frac{n}{n-1} \left(\text{Var}(X_1) + (\mathbb{E}[X_1])^2 - \text{Var}(\bar{X}) - (\mathbb{E}[\bar{X}])^2 \right) \\ &= \frac{n}{n-1} (\sigma^2 - \sigma^2/n) = \sigma^2. \quad \text{Done.}\end{aligned}$$

Risultato fondamentale sull' errore quadratico medio

$MSE \doteq E[(T(X_1, \dots, X_n) - \theta)^2]$ è l'**errore quadratico medio**

ALLORA RISULTA:

$$MSE(T(\mathbf{X})) = E[T^2] - 2\theta E[T] + \theta^2$$

$$= E[T^2] - (E[T])^2 + (E[T])^2 - 2\theta E[T] + \theta^2$$

$$= \text{Var}(T) + \underbrace{(E[T] - \theta)^2}_{\substack{\text{intervallo più} \\ \text{largo o stretto}}}.$$

il bias sposta centro del intervallo ed è indotto dal casualità
Bias vista nell'intervallo di confidenza

Per il momento non fornisco alcun esempio di calcolo di MSE.

Mi limito a definire “stima puntuale” (del parametro θ) qualunque realizzazione della variabile aleatoria stimatore. Pertanto, maggiore è il MSE dello stimatore e più rischioso risulta affidarsi a quella stima.

Quando risulta $E[T(X)] \neq \theta$, si è in presenza del cosiddetto **errore sistematico**, o **bias**, appunto pari alla differenza fra il valore atteso dello stimatore ed il parametro da stimare.

Se Y è la media campionaria, come già precedentemente definito, cioè

$$Y \triangleq \frac{1}{n} \sum_{i=1}^n X_i$$

e se le X_i sono tutte identicamente distribuite e indipendenti, con una media incognita
 $_i = \forall i$, allora:

$$E[Y] = \frac{1}{n} \sum_{i=1}^n E[X_i] =$$

Dunque: $T(X_1, \dots, X_n) \triangleq (1/n) \sum_{i=1}^n X_i$ è uno stimatore corretto per il parametro.

Allo stesso modo, passando alla varianza delle variabili aleatorie, $\sigma_i^2 = \sigma^2 \forall i$ e ricordando che risulta:

$$E\left[\frac{1}{(n-1)} \sum_{i=1}^n (X_i - Y)^2\right] = \sigma^2$$

si ottiene la seguente statistica S^2 , che è uno stimatore corretto per il parametro σ^2 :

$$S^2 \triangleq \frac{1}{(n-1)} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{\sum_{i=1}^n X_i^2 - n\bar{X}^2}{n-1}$$

DEF: Uno stimatore è **asintoticamente corretto**, se è corretto al tendere di n all'infinito.

DEF: L'**efficienza relativa** è il rapporto tra le varianze di due stimatori corretti.

DEF: Uno stimatore è detto **consistente** se converge in probabilità al parametro da stimare, ovvero: $\lim_{n \rightarrow \infty} \Pr(|T(X_1, \dots, X_n) - \theta| \geq \varepsilon) = 0$

La bontà di uno stimatore deve essere valutata anzitutto in termini della sua varianza: più bassa la varianza maggiore la qualità. Infine, si tenga presente che una statistica viene usata come stimatore di un parametro anche quando non risulta essere uno

Stimatori efficienti

conosciamo lo stimatore del valore atteso ovvero la media campionaria. Ma è il migliore stimatore?

Siano $T_1(X)$ e $T_2(X)$ due stimatori differenti dello stesso parametro θ . E siano valide le seguenti:

- 1) $T_1(X)$ e $T_2(X)$ sono entrambi non distorti per θ ;
- 2) $\text{Var}[T_1(X)] \leq \text{Var}[T_2(X)]$, per ogni valore di θ ;
- 3) $\text{Var}[T_1(X)] < \text{Var}[T_2(X)]$, per qualche valore di θ ;

Allora si dirà che $T_1(X)$ è stimatore più efficiente di $T_2(X)$ per θ

ESEMPIO: la media campionaria è lo stimatore più efficiente tra quelli definibili tramite una combinazione lineare delle X_1, \dots, X_n , di coefficienti a_1, \dots, a_n .

Infatti:

$$T(X_1, \dots, X_n) \triangleq \sum_{i=1}^n a_i X_i \Rightarrow \text{Var}\left[\sum_{i=1}^n a_i X_i\right] = \dots = \sum_{i=1}^n a_i^2 \sigma^2$$

essendo indipendenti

Da qui:

$$\begin{cases} \text{MIN : } \sigma^2 \sum_{i=1}^n a_i^2 \\ \text{a}_1, \dots, \text{a}_n \\ \text{sub. to : } \sum_{i=1}^n a_i = 1, \end{cases}$$

risolvendo il problema di ricerca operativa. si dimostra che la media campionaria è lo stimatore più efficiente

$$\rightarrow \boxed{a_i = 1/n, \quad i = 1, \dots, n.}$$

Stimatore consistente: la distribuzione empirica

Sia n_x il numero di valori generati dal modello F

che cadono nell'intervallo $(0,x]$ e sia $\hat{F}_n(x) \triangleq n_x / n$.

Intervalli visti in Monte Carlo

Ebbene, la distribuzione empirica, \hat{F} ,
è uno stimatore consistente della distribuzione vera, F .

Osservando che un meccanismo bernoulliano è alla base del risultato
“numero di realizzazioni che cadono in un intervallo x ” e che la
corrispondente probabilità (di successo) è $p=F(x)$:

La distribuzione
empirica porta
ad uno stimatore consistente.

$$\hat{F}(x) \triangleq \frac{n_x}{n} = \frac{1}{n} \sum_{i=1}^n B_i \text{ binomiale ovvero} \quad \text{con } E[\bar{B}(n)] = np / n = p$$

0,1. cade o non cade nell'intervallo

Grazie al teorema di Bernoulli, caso particolare della LGN:

$$\lim_{n \rightarrow \infty} P(|\bar{B}(n) - E[\bar{B}(n)]| \geq \varepsilon) = 0 \quad \text{appunto DEFINIZIONE stimatore consistente}$$

Stimatore consistente se: $\lim_{n \rightarrow \infty} \Pr(|T(X_1, \dots, X_n) - \theta| \geq \varepsilon) = 0$

Ovvero, esprimendo il risultato in termini di F e di \hat{F}

$$\lim_{n \rightarrow \infty} P(|\hat{F}_n(x) - F(x)| \geq \varepsilon) = 0$$

$\hat{F}_n(x)$ sarebbe una famiglia di funzioni
che converge tutta in forma.
Secondo me sarebbe meglio $\hat{F}_x(n)$

I parametri definiscono/caratterizzano il sistema. L'uscita non è un parametro

Stima di parametri di una distribuzione
con
Metodo dei momenti (equazioni)
e
Metodo della massima verosimiglianza
(problema di ricerca operativa)

Applicazione alle distribuzioni seguenti:

ESPONENZIALE, NORMALE,
IPERESPONENZIALE, di ERLANG,
di BERNOULLI, di POISSON e di WEIBULL

Ricordando che per Weibull i parametri ci fanno ricavare media e varianza

STIMA DEI PARAMETRI

IMPOSTAZIONE FREQUENTISTA

Le realizzazioni che formano il campione contengono tutta la informazione necessaria per fare inferenza sul VERO, UNICO e INCOGNITO valore del parametro di interesse diretto o indiretto perché vogliamo inferire alcune caratteristiche del fenomeno governato da quel parametro.

IMPOSTAZIONE BAYESIANA

Al parametro di interesse, incognito, viene attribuita un modello probabilistico iniziale, a prescindere dalle realizzazioni del campione, e del quale esso sarebbe una realizzazione. Il modello probabilistico iniziale è detto “modello a priori” e sarà aggiornato sulla base delle realizzazioni che formano il campione. Il modello aggiornato è detto “modello a posteriori”.

DALLA FORMULA DI BAYES ALL'INFERENZA BAYESIANA

A è un evento di interesse e B_1, \dots, B_n sono eventi di probabilità nota

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

B_1, \dots, B_n è una partizione di Ω

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

Problema:

Abbiamo un campione di realizzazioni indipendenti (D) di una variabile aleatoria che afferisce certamente ad un modello esistente in letteratura e caratterizzato dal parametro θ : vogliamo individuare questo parametro a partire da quei dati!

$$P(\theta | D) = \frac{P(D | \theta) \cdot P(\theta)}{\sum_{\substack{\text{possibili} \\ \text{valori di } \theta}} P(D | \theta) \cdot P(\theta)}$$

$P(\theta)$ è la prob. "a priori"
 $P(\theta | D)$ è la prob. "a posteriori"

$P(D | \theta)$ è detta
"Verosimiglianza di θ "

Il metodo dei momenti

*Siano X_1, X_2, \dots, X_n indip. e id. distr.
con μ^k momento di ordine "k", $k = 1, 2, \dots$*

Allora, pensando alla legge dei grandi numeri:

$$\sum_{i=1}^n \frac{X_i}{n} \xrightarrow{\text{media campionaria converge al valore atteso come visto tramite monte carlo}} E[X_i] \hat{=} \mu, \quad \sum_{i=1}^n \frac{X_i^2}{n} \xrightarrow{} E[X_i^2] \hat{=} \mu_2, \quad \dots$$

Spesso è utile la seguente :

$$VAR[X_i] = E[X_i^2] - (E[X_i])^2$$

$$\Rightarrow \widetilde{S}^2 \hat{=} \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2 = \frac{\sum_{i=1}^n X_i^2 - n\bar{X}^2}{n} = \frac{n-1}{n} S^2$$

Il metodo dei momenti (MOM)

- Il MOM consiste nell'eguagliare i momenti di ordine $1, 2, \dots k$ ($m_1, m_2, \dots m_k$) con le corrispondenti formule di stima, fino ad ottenere (non sempre facile!) tante equazioni quanti sono i parametri incogniti:

$$\sum_{i=1}^n \frac{x_i^k}{n} = m_k, \quad k = 1, 2, \dots$$

momento di
ordine k momento k-esimo

- Per il modello esponenziale di parametro λ , siamo alla banalità!
perchè è $k=uno!$

$$\frac{1}{n} \sum_{i=1}^n x_i^1 = m_1 (= 1/\lambda) \Rightarrow \hat{\lambda} \equiv m_1 = n / \sum x_i$$

senza saperlo su Excel si stava attuando il metodo dei momenti

M O M per la distribuzione normale:

- **X₁, ..., X_n estratte da N(μ, σ²)**
- **2 parametri da stimare ⇒ sistema in 2 equazioni:**

$$\begin{cases} E(X) = \mu = \frac{1}{n} \sum_{i=1}^n X_i & (\hat{=} \bar{X}) \quad \text{momento primo ordine} \\ E(X^2) = \mu^2 + \sigma^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 & \text{momento del secondo ordine} \end{cases}$$

- Dunque:

$$\hat{\mu} = \bar{X} \quad \text{e} \quad \hat{\sigma}^2 = \frac{n-1}{n} S^2$$

(stessi risultati
della stima MLE)

Es3: MOM per la distribuzione iperesponenziale

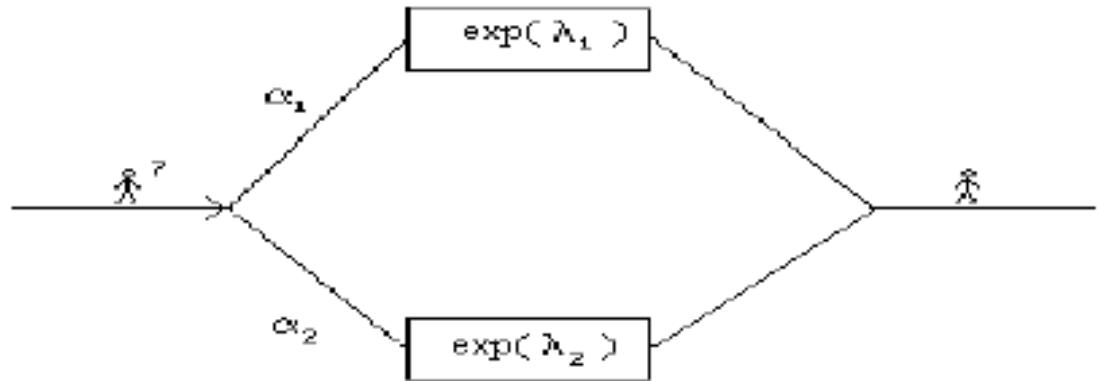
- Distribuzione iperesponenziale

$$F_y(y) = \sum_{i=1}^n \alpha_i (1 - e^{-\lambda_i y}) \quad y \geq 0$$

- 3 parametri da stimare: $\alpha, \lambda_1, \lambda_2$. e non 2 banali

$$\alpha_1 = \alpha$$

$$\alpha_2 = 1 - \alpha$$



- In questo caso ci servirebbero i momenti del 1°, 2° e 3° ordine. Ma abbiamo a disposizione solo la media e la varianza campionaria (un livello di libertà!).
- Perciò si sceglie di impostare un'ulteriore condizione per rendere determinato il sistema:

$$\frac{1}{\lambda_1 + \lambda_2} = \frac{\frac{\alpha}{\lambda_1} + \frac{1-\alpha}{\lambda_2}}{2}$$

\bar{X} media c. $\frac{(n-1)}{n} S^2$ varianza campionaria

- momento del 1° ordine :

$$E[Y] = \sum_{i=1}^n \frac{\alpha_i}{\lambda_i} = \left(\frac{\alpha}{\lambda_1} + \frac{1-\alpha}{\lambda_2} \right)$$

probabilità totale
 cioè valore atteso condizionato
 da α_1 e condizionato da α_2

- Varianza:

$$VAR[Y] = E[Y^2] - E^2[Y]$$

nello svolgimento useremo
 S^2 e non $(n-1)S^2/n$

- momento del 2° ordine :

$$E[Y^2] = \sum_{i=1}^n \frac{2\alpha_i}{\lambda_i^2} = 2 \left(\frac{\alpha}{\lambda_1^2} + \frac{1-\alpha}{\lambda_2^2} \right)$$

anche qui
 probabilità totale

- Stimatore:

- 1° ordine : $M_1 \rightarrow \bar{X}$
- 2° ordine :

$$M_2 \rightarrow S^2 - E^2[Y]$$

- Il nostro sistema di 3 equazioni in 3 incognite diventa perciò:

$$\left\{ \begin{array}{l} \frac{1}{\lambda_1 + \lambda_2} = \frac{\frac{\alpha}{\lambda_1} + \frac{1-\alpha}{\lambda_2}}{2} \quad \text{condizione aggiunta} \\ \\ \bar{X} = \frac{\alpha}{\lambda_1} + \frac{1-\alpha}{\lambda_2} \quad \text{media ottenuta} \\ \\ S^2 = 2 \left(\frac{\alpha}{\lambda_1^2} + \frac{1-\alpha}{\lambda_2^2} \right) - \left(\frac{\alpha}{\lambda_1} + \frac{1-\alpha}{\lambda_2} \right)^2 \quad \text{varianza ottenuta} \end{array} \right.$$

risoluzione che puoi "saltare"

- Osservando che il secondo membro della prima equazione è uguale a quello della seconda: $\lambda_1 = \frac{2}{\bar{X}} - \lambda_2$
- Dalla seconda equazione ricaviamo: $\hat{\alpha} = \frac{\lambda_1(\lambda_2 \bar{X} - 1)}{\lambda_2 - \lambda_1}$
- Sostituendo λ_1 : $\alpha = 1 - \frac{\lambda_2 \bar{X}}{2}$
- Osserviamo che :

$$\frac{\alpha}{\lambda_1} = \frac{1-\alpha}{\lambda_2} = \frac{\bar{X}}{2} \quad \frac{\alpha}{\lambda_1^2} = \frac{\bar{X}^2}{2(2 - \lambda_2 \bar{X})} \quad \frac{1-\alpha}{\lambda_2^2} = \frac{\bar{X}}{2\lambda_2}$$

- La terza equazione diventa:

$$S^2 = 2 \left(\frac{\bar{X}^2}{2(2 - \lambda_2 \bar{X})} + \frac{\bar{X}}{2\lambda_2} \right) - \left(\frac{\bar{X}}{2} + \frac{\bar{X}}{2} \right)^2$$

da cui:

$$\lambda_2(2 - \lambda_2 \bar{X}) = \frac{2\bar{X}}{S^2 + \bar{X}^2}$$

- Manipolando tale equazione giungiamo ad un'equazione di 2° grado in λ_2 ...

$$\lambda_2^2 \bar{X} (S^2 + \bar{X}^2) - 2\lambda_2 (S^2 + \bar{X}^2) + 2\bar{X} = 0$$

si arriva ad un problema di analisi numerica che si risolve

- ... che ammette come soluzioni:

$$\hat{\lambda}_2 = \frac{1}{\bar{X}} \pm \frac{1}{\bar{X}} \sqrt{\frac{(S^2 - \bar{X}^2)}{(S^2 + \bar{X}^2)}}$$

- Da cui: $\hat{\lambda}_1 = \frac{1}{\bar{X}} \mp \frac{1}{\bar{X}} \sqrt{\frac{(S^2 - \bar{X}^2)}{(S^2 + \bar{X}^2)}}$

(si suppone che $S^2 \geq \bar{X}^2$)

- Per completezza riscriviamo il terzo coefficiente:

$$\hat{\alpha} = \frac{\lambda_1(\lambda_2 \bar{X} - 1)}{\lambda_2 - \lambda_1}$$

Metodo dei momenti per la distribuzione iperesponeziale

EXCEL

il metodo di massima verosimiglianza per la iperesponeziale è complicato da attuare

<i>Valori realizzazioni</i>	<i>Stime puntuale dei parametri con il metodo dei momenti</i>	<i>(Scarti)²</i>
0,070992297		0,1053
0,255808902		0,019511
11,42159369		121,5749
0,243130754		0,023214
0,027101342		0,135712
0,316199639		0,006287
0,085576828		0,096048
0,00603031		0,151681
0,481671006		0,007427
0,229857669		0,027435
0,073575389		0,10363
0,019303896		0,141518
0,342766998		0,00278
0,259919485		0,01838
0,182692995		0,045284
0,237436716		0,024982
0,005607306		0,15201
0,264871119		0,017062
0,016183388		0,143875
0,180668483		0,046149
0,190603528		0,041979
0,127738893		0,071692
0,47559937		0,006417
0,119399439		0,076227
0,026289307		0,136311
0,463860674		0,004674
0,100128708		0,08724
0,073025665		0,103985
0,09587458		0,089771
0,079357751		0,099941
0,347303991		0,002322
0,107629016		0,082865
0,107185665		0,083121
0,234467079		0,025929
0,046598669		0,121727
0,108701554		0,082249
0,016925912		0,143312
0,034854334		0,13006
0,454909128		0,00353
0,040687167		0,125887
0,007315512		0,150681
0,595406889		0,039966
0,447704247		0,002726
0,029032395		0,134293
0,04546366		0,12252
0,030887532		0,132937
0,197114629		0,039354
0,29058766		0,011005
0,059033555		0,113204
0,099939163		0,087352
# realizzazioni	50	

Metodi di stima dei parametri (di una forma nota)

La stima di massima verosimiglianza (maximum likelihood estimation – MLE)

È basata sulla cosiddetta funzione di verosimiglianza, incognita nei parametri d'interesse.

I parametri saranno stimati trovando il massimo della funzione stessa.

Stima di massima verosimiglianza

Date le variabili casuali i.i.d. $X_1 \dots X_n$,
con una funzione di densità $f(\mathbf{x})$,
il parametro che si vuole valutare è θ .

massimizzare la probabilità

La stima di massima verosimiglianza (MLE) di θ è il valore di θ che massimizza:

$$L(\theta) = f(X_1) \cdot f(X_2) \dots f(X_n)$$

se inserisco dx1, dx2 ecc. ottengo probabilità. Quindi si mira a massimizzare la probabilità che quelle realizzazioni sono figli di quella forma con quel parametro

Si usa la funzione di verosimiglianza ridotta:

$$l(\theta) = \ln[L(\theta)]$$

Appunto log di una somma tipo iperexp non si usa mentre log di un prodotto è utile!

I massimi restano sempre uguali, trovare quindi il massimo del logaritmo della funzione di massima verosimiglianza è corretto perché il logaritmo è una funzione monotona e molti modelli sono espressi con esponenziali!

Algoritmo della MLE

- L'algoritmo procede nel trovare il massimo della
 $l(\theta)$
ponendo uguale a zero la derivata di tale
funzione nel parametro incognito θ .
in modo classico insomma
- Se i parametri sono **più di uno**, bisogna risolvere
un sistema di equazioni alle derivate parziali
secondo ogni parametro.

Es. 1: la densità esponenziale,

$$L(\theta) \hat{=} \prod_{i=1}^n f(x_i) = \prod_{i=1}^n \theta e^{-\theta x_i} = \theta^n e^{-\theta \sum_i x_i}$$

$$\ln [L(\theta)] = \ln \left[\theta^n e^{-\theta \sum_i x_i} \right] = n \ln \theta - \theta \sum_i x_i$$

$$\frac{d}{d\theta} l(\theta) = \frac{d}{d\theta} [n \ln \theta - \theta \sum_i x_i]$$

non è detto che i due metodi diano la stessa stima

$$= \frac{n}{\theta} - \sum_i x_i = 0 \Rightarrow \hat{\theta} = n / \sum_{i=1}^n x_i$$

Es. 2: la distribuzione normale

$$L(\mu, \sigma^2) = \prod_{i=1}^n f(x_i)$$

$$= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}\frac{(x_i - \mu)^2}{\sigma^2}\right\}$$

$$= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2} \sum_{i=1}^n \frac{(x_i - \mu)^2}{\sigma^2}\right\}$$

- Stima del parametro μ

$$\ln(L(\mu, \sigma^2)) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2$$

$$\frac{\partial}{\partial \mu} \ln(L(\mu, \sigma^2)) = \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) \equiv 0$$

$$\Rightarrow 0 = \sum_{i=1}^n (x_i - \mu) = \sum_{i=1}^n x_i - n\mu$$

$$\Rightarrow \mu = \frac{1}{n} \sum_{i=1}^n x_i$$

- Stima del parametro σ^2

$$\frac{\partial}{\partial \sigma^2} \ln(L(\mu, \sigma^2)) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \hat{\mu})^2 \equiv 0$$

$$\Rightarrow -n\sigma^2 + \sum_{i=1}^n (x_i - \bar{x})^2 = 0$$

la stima qui risulta non distorta

$$\Rightarrow \widehat{\sigma^2} = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n} = \boxed{\frac{n-1}{n} S^2}$$

Es. 3: la ripartizione bernoulliana

Example: Suppose $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \text{Bern}(p)$. Find the MLE for p .

Useful trick for this problem: Since

$$X_i = \begin{cases} 1 & \text{w.p. } p \\ 0 & \text{w.p. } 1 - p \end{cases},$$

we can write the p.m.f. as

probability max function. perchè nel discreto non si usa la densità. noi la potremmo chiamare ripartizione massima di probabilità

$$f(x) = p^x(1-p)^{1-x}, \quad x = 0, 1.$$

La funzione di verosimiglianza per la legge bernoulliana:

$$L(p) = \prod_{i=1}^n f(x_i)$$

$$= \prod_{i=1}^n p^{x_i} (1-p)^{1-x_i}$$

$$= p^{\sum_{i=1}^n x_i} (1-p)^{n-\sum_{i=1}^n x_i}$$

ESPRIME LA PROBABILITA' DI OSSERVARE CONGIUNTAMENTE LE REALIZZAZIONI x_i , $i = 1, \dots, n$

Lo stimatore di massima verosimiglianza:

$$\ln(L(p)) = \sum_{i=1}^n x_i \ln(p) + (n - \sum_{i=1}^n x_i) \ln(1-p)$$

$$\frac{\partial}{\partial p} \ln(L(p)) = \frac{\sum_i x_i}{p} - \frac{n - \sum_i x_i}{1-p} \equiv 0.$$

derivata parziale perchè
ho $\ln(L(p))$

$$(1-p)\left(\sum_{i=1}^n x_i\right) - p\left(n - \sum_{i=1}^n x_i\right) = 0$$

andando ad isolare per ottenere lo stimatore di massima verosimiglianza, ottengo appunto la media campionaria

$$\sum_{i=1}^n x_i - pn = 0 \quad \Rightarrow \hat{p} = \frac{1}{n} \sum_{i=1}^n x_i$$

e ricordando che una binomiale converge ad una Gaussiana riusciamo a fare un intervallo di confidenza

Es. 4: la ripartizione poissoniana

$$L(\lambda) \hat{=} \prod_{i=1}^n \frac{\lambda^{x_i}}{x_i!} e^{-\lambda} = \frac{\lambda^{\sum x_i}}{x_1! x_2! \cdots x_n!} \cdot e^{-n\lambda}$$

$$\ln [L(\lambda)] = -\ln(x_1! x_2! \cdots x_n!) - n\lambda + \left(\sum_i x_i\right) \cdot \ln(\lambda)$$

$$\frac{d}{d\lambda} \ln[L(\lambda)] = -n + \frac{1}{\lambda} \sum_i x_i$$

$$\Rightarrow \hat{\lambda} = \frac{1}{n} \sum_{i=1}^n x_i$$

Es. 5: la distribuzione di Weibull

Weibull di Trivedi e non di EXCEL

$$L(\lambda, \alpha) = \prod_{i=1}^n f(x_i) = \prod_{i=1}^n \lambda \alpha \cdot {x_i}^{\alpha-1} \cdot e^{-\lambda {x_i}^\alpha}$$

$$= \lambda^n \alpha^n \cdot \left(\prod_{i=1}^n x_i \right)^{\alpha-1} \cdot e^{-\lambda \sum_{i=1}^n {x_i}^\alpha}$$

$$\ln L(\lambda, \alpha) = n \ln \lambda + n \ln \alpha + (\alpha - 1) \sum_{i=1}^n \ln x_i - \lambda \sum_{i=1}^n {x_i}^\alpha$$

altro problema di analisi numerica

$$\left\{ \begin{array}{l} \frac{n}{\lambda} - \sum_{i=1}^n {x_i}^\alpha = 0 \\ \frac{n}{\alpha} + \sum_{i=1}^n \ln x_i - \lambda \sum_{i=1}^n {x_i}^\alpha \ln x_i = 0 \end{array} \right.$$

Se cerco di massimizzare i parametri lambda e alpha, ottengo un sistema che non ha un'unica soluzione, cioè non vi sono soluzioni in forma chiusa.

Appunto

- Non esistono soluzioni in forma chiusa per λ ed α .
Comunque si può scrivere λ in funzione di α . per risolvere il problema

formula per stimare
lambda

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^n x_i^\alpha}$$

appunto fissando alpha. si ottiene
lambda

- Sostituendo nella seconda equazione:

problema di analisi numerica

$$\frac{n}{\alpha} + \sum_{i=1}^n \ln x_i - \frac{n \sum_{i=1}^n x_i^\alpha \ln x_i}{\sum_{i=1}^n x_i^\alpha} = 0$$

che può essere risolta con un metodo di “punto fisso”

Trovato il valore di α , lo si userà per ricavare λ .

ovvero calcolo
numerico. Un metodo
è ad esempio l'iterazione

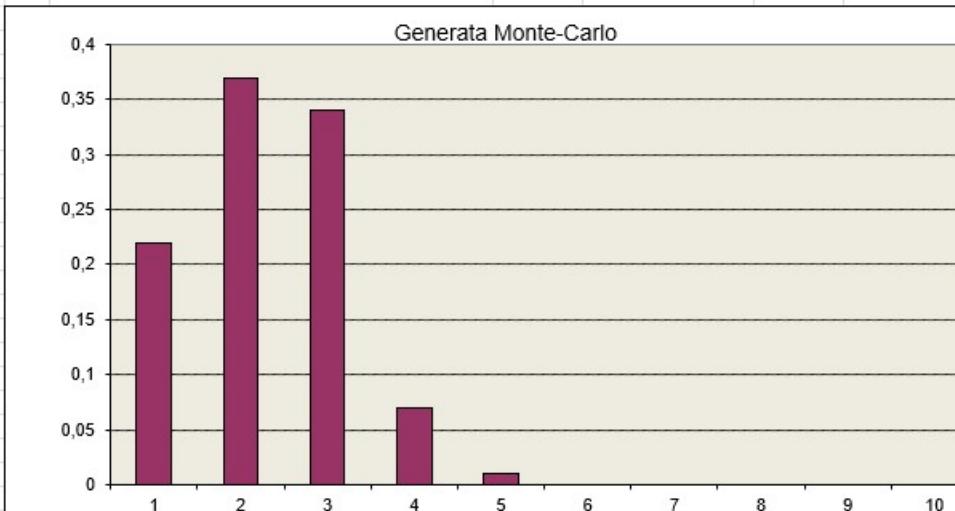
Tramite Excel andiamo a stimare i parametri di una Weibull con una generazione Monte Carlo

Realizz di una Weibull	Real Weib ^alpha	alfa	lambda di WEIB = 0,2500	media camp di real Weib ovvero media aritmetica la sua realizzazione = 1,753074621	num realizzazioni = 100
			assegnato. Per ipotesi stimato	num intervalli = 10	
2,278184076	5,190122686	2	somma x-i ^ α		
2,980528663	8,883551111				
1,374309393	1,888726308				
4,09979484	16,80831773	388,3024			
2,178630612	4,746431344	lambda_cap			
0,513064168	0,26323484	0,2575			
1,272460527	1,619155792				
0,729703423	0,532467086				
0,371302031	0,137865198				
2,035413355	4,142907527				
1,164327462	1,356558439				
3,003493793	9,020974966				
2,568384657	6,596599748				
1,937524553	3,754001395				
1,226054467	1,503209556				
1,551545773	2,407294287				
1,528032867	2,334884444				
0,925785038	0,857077937				
2,9857528	8,914719784				
3,009518122	9,057199324				
1,187021164	1,409019245				
1,496110558	2,238346801				
2,741617091	7,516464273				
3,167702269	10,03433767				
0,559882418	0,313468322				
1,655650865	2,741179787				
2,424995768	5,880604476				
1,027026841	1,054784131				
1,590254249	2,528908578				
1,75606515	3,083764812				
0,913556142	0,834584825				
2,510023938	6,300220168				
1,19593173	1,430252702				
0,327620248	0,107335027				
0,880801391	0,77581109				
2,501043325	6,255217714				

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^n x_i^\alpha}$$

$$E[Weib] = \beta \cdot \Gamma(1 + 1/\alpha)$$

beta
2



La stima risulta abbastanza vicina al valore vero

L'analisi sui parametri di una Weibull porta alla conclusione che:

- Lambda è il parametro di scala, solitamente durante una ricostruzione non "danneggia" il risultato
- Alpha è il parametro di forma, è consigliato scegliere un valore maggiore di 1 per visualizzare bene il metodo di Monte Carlo

La diseguaglianza (o lemma) di Chebyshev

Sia Y la var. al. media campionaria di “n” variabili aleatorie indipendenti e identicamente distribuite, con un valore atteso comune (μ) e una varianza comune (σ^2).

$$\text{Dunque: } Y \hat{=} \frac{1}{n} \sum_{i=1}^n X_i \quad \text{con} \quad E[Y] = \frac{1}{n} \sum_{i=1}^n E[X_i] = \mu \quad \text{e} \quad \text{VAR}[Y] = \frac{\sigma^2}{n} \quad (++)$$

Si è indotti a ritenere che la crescita di “n”, cioè della “numerosità del campione” dovrebbe produrre anche un effetto di crescente concentrazione delle realizzazioni della Y attorno alla sua media (μ), fino a far degenerare la forma a “campana” in una sorta di funzione limite del tutto speciale: con un solo valore all’infinito e nello stesso tempo di area sottesa unitaria!

Sorge, altresì, spontanea la seguente domanda: in che misura è possibile, al crescere di n, escludere il verificarsi di realizzazioni della Y che siano significativamente distanti dalla media?

La risposta alla domanda è fornita dal seguente lemma dovuto a *Chebyshev*, che vale per qualunque variabile aleatoria e sarà dimostrato facendo riferimento ad una Y continua tra $-\infty$ e $+\infty$ e arbitrariamente distribuita, purché di valore atteso $E[Y] \hat{=} \mu$ e varianza $Var[Y]$. La varianza è solitamente indicata come σ^2 , ma non ora per evitare confusione con la (++) .

Lemma di Chebyshev :

$$\Pr\left\{|Y - \mu| \geq k \cdot \sqrt{Var[Y]}\right\} \leq \left(\frac{1}{k^2}\right), \quad k \mid (1/k^2) < 1 \quad (\text{da fissare})$$

$$\text{Ovvero} \quad \Pr\left\{|Y - \mu| \geq h\right\} \leq \left(Var[Y]/h^2\right), \quad k \cdot \sqrt{Var[Y]} \hat{=} h \quad (\#)$$

Si ribadisce che il Lemma è riferito a variabili aleatorie continue e con realizzazioni comprese tra $-\infty$ e $+\infty$ (come la gaussiana, ad esempio, ma non necessariamente!)

La particolarizzazione dello stesso Lemma a variabili aleatorie continue e non negative (come la esponenziale, la Weibull e altre) non è banale, tant’è che è dovuta al matematico siciliano Francesco Paolo Cantelli!

Lemma di Cantelli: $Y \in [0, +\infty) \Rightarrow \Pr\left\{Y \leq \mu + k \sqrt{Var[Y]}\right\} \leq 1 - (1/(1+k^2))$

La qualità pratica del Lemma di Chebyshev, cioè la sua capacità di offrire un limite superiore abbastanza stretto è di solito insoddisfacente, ma la sua importanza teorica è notevole perché essa è alla base del concetto di “convergenza in probabilità” e, da qui, alla base della “legge debole dei grandi numeri” e del “teorema di Bernoulli”. La qualità pratica viene investigata nel file Excel associato a questa lezione. Ad esempio, con riferimento ad una gaussiana di media nulla, la probabilità vera che le realizzazioni siano contenute ad una distanza superiore a 2 deviazioni standard a destra e a sinistra dello zero è 0.046, ma la diseguaglianza di Chebyshev stabilisce solo che ciò accade con probabilità ≤ 0.25 !

PROVA del solo Lemma di Chebyshev:

Partendo dalla definizione di varianza, $\text{Var}[Y] = E[(Y - \mu)^2]$, quindi:

$$\text{Var}[Y] \hat{=} \int_{-\infty}^{+\infty} (y - \mu)^2 f(y) dy$$

vale la seguente diseguaglianza:

$$\text{Var}[Y] \geq \int_{-\infty}^{-\mu-h} (y - \mu)^2 f(y) dy + \int_{+\mu+h}^{+\infty} (y - \mu)^2 f(y) dy$$

perché si sta escludendo dal calcolo dell'integrale la porzione di area sottesa alla $f(y)$ che risulta compresa nell'intervallo di integrazione $[\mu-h, \mu+h]$. E si tratta di una area che è sicuramente ≥ 0 perché $f(y)$ è non negativa e quindi $(y - \mu)^2 f(y)$ è pure ≥ 0 ;

adesso:

$$\text{Var}[Y] \geq h^2 \left(\int_{-\infty}^{-\mu-h} f(y) dy + \int_{+\mu+h}^{+\infty} f(y) dy \right) \quad (\$)$$

perché quando la variabile di integrazione (y) scorre tra $-\infty$ e $(-\mu-h)$ risulta $< \mu$ e, analogamente, quando scorre tra $(\mu+h)$ e $+\infty$ risulta $> \mu$;

infine:

$$\text{Var}[Y] \geq h^2 \left(\Pr \{ |Y - \mu| \geq h \} \right) \quad \text{ovvero}$$

$$\Pr \{ |Y - \mu| \geq h \} \leq \text{Var}[Y]/h^2$$

perché il primo integrale nella (\$) calcola, per definizione, la $\text{Prob}[Y < (-\mu-h)]$ e il secondo calcola la probabilità la $\text{Prob}[Y > (\mu+h)]$. FINE PROVA.

Osservazione conclusiva:

Ritornando alla domanda sulla media campionaria, adesso si può rispondere che, al

crescere della numerosità del campione (n), diventerà sempre più piccola la probabilità

N.B. All'aumentare dei gradi di libertà la chi-quadrato arriva ad assomigliare alla forma di una normale. Vedi EXCEL

di osservare una realizzazione della media campionaria che abbia una distanza apprezzabile dal valore atteso . Precisamente, considerando la Y del Lemma di Chebyshev come una media campionaria con $\text{Var}[Y] = (\sigma^2 / n)$, si può riprendere la seconda forma (#) della disuguaglianza di Chebyshev:

$$\Pr \{ |Y - \vartheta| \geq h \} \leq \left(\frac{\text{Var}[Y]}{h^2} \right) = \left(\frac{(\sigma^2 / n)}{h^2} \right)$$

e fissando $h=\sigma$ (cioè il parametro “h” pari proprio alla deviazione standard delle “n” variabili componenti la media campionaria) si ottiene la seguente indicazione: $\Pr \{ |Y - \vartheta| \geq \sigma \} \leq 1/n$, che fa capire l’importanza della dimensione del campione.

Convergenza in probabilità e legge dei grandi numeri

Un concetto nuovo di convergenza, di tipo probabilistico, è chiaramente affiorato nei ragionamenti appena sviluppati, grazie alla disuguaglianza di Chebyshev, sulla media campionaria. La formalizzazione di esso porta alla definizione di convergenza in probabilità, nell’analisi probabilistica.

Definizione di convergenza in probabilità:

Si dice che una generica sequenza di variabili aleatorie Y_1, Y_2, \dots, Y_n converge in probabilità ad un valore finito ϑ e si scrive:

$$\lim_{n \rightarrow \infty} \Pr \{ |Y_n - \vartheta| = 0 \} = 1$$

quando, per qualsivoglia $\varepsilon > 0$, scelto arbitrariamente piccolo, esistono sempre una quantità $\delta > 0$ e uno specifico valore di n sufficientemente grande (\tilde{n}) tale che risulti verificata la seguente:

$$\Pr \{ |Y_{\tilde{n}} - \vartheta| < \delta \} > 1 - \varepsilon$$

Si osservi che la convergenza in probabilità è riferita alla *singola* quantità aleatoria $|Y_{\tilde{n}} - \vartheta|$ e non può essere estesa alla congiunzione logica (AND) di tutte le seguenti quantità:

$$|Y_{\tilde{n}+1} - \vartheta| \cap |Y_{\tilde{n}+2} - \vartheta| \cap |Y_{\tilde{n}+3} - \vartheta| \cap \dots$$

Dunque, questo tipo di convergenza non può escludere, al crescere di n oltre il valore \tilde{n} il verificarsi di realizzazioni apprezzabilmente distanti dal valore atteso con probabilità

finita. Cosa che è invece esclusa dalla versione “forte” della LGN (che sarà enunciata più avanti).

Legge (debole) dei grandi numeri

Nella formulazione di interesse per questo corso, la legge dei grandi numeri esprime il risultato della convergenza in probabilità della media campionaria Y al valore atteso finito, , di ognuna delle variabili identicamente distribuite, X_1, X_2, \dots, X_n , ma non necessariamente indipendenti, quando la dimensione del campione (n) tende all’infinito. Tale risultato di convergenza è più generale della cosiddetta convergenza “in forma” sul teorema limite centrale, perché è valido senza bisogno che le variabili X_1, X_2, \dots, X_n siano indipendenti.

Per sottolineare l’importanza della dimensione del campione si userà la seguente notazione:

$$Y(n) \hat{=} \frac{1}{n} \sum_{i=1}^n X_i, \quad \text{con} \quad E[Y(n)] = \frac{1}{n} \sum_{i=1}^n E[X_i] =$$

e rinunciando all’ipotesi di indipendenza, si scriverà soltanto:

$$VAR[Y(n)] = \frac{1}{n^2} VAR\left[\sum_{i=1}^n X_i\right] \hat{=} v^2(n)$$

e si farà solo l’ipotesi (meno restrittiva) che risulti : $\lim_{n \rightarrow \infty} v^2(n) = 0$

A questo punto, riscrivendo la diseguaglianza di *Chebyshev* nella forma complementare a quella ricavata prima:

$$\Pr\{|Y(n) - | < k v(n)\} > 1 - (v^2(n)/h^2) = 1 - (1/k^2)$$

e ponendo $\delta = kv(n)$ ed $\varepsilon = (1/k^2)$ si vede subito che l’ipotesi $\lim_{n \rightarrow \infty} v^2(n) = 0$

garantisce il rispetto della definizione di convergenza in probabilità della media campionaria al valore atteso .

Per completezza, viene di seguito viene riportata pure la dimostrazione della LGN debole sotto la ipotesi più restrittiva, che le variabili aleatorie in gioco siano indipendenti.

La diseguaglianza di Chebyshev:

$$\Pr \{ |Y(n) - \mu| \geq h \} \leq \left(\text{Var}[Y(n)] / h^2 \right) \quad \text{con} \quad (h \hat{=} n\sqrt{\text{Var}[Y(n)]})$$

viene riscritta in forma complementare:

$$\Pr \{ |Y(n) - \mu| < h \} \geq 1 - \left(\text{Var}[Y(n)] / h^2 \right)$$

Introducendo l'ipotesi di varianze finite ($\text{Var}[X_k] = \sigma_k^2$, $k=1,2,\dots,n$) e limitate tutte dalla stessa costante, C , risulta (per l'additività della varianza sotto ipotesi di indipendenza!):

$$\text{Var}[Y(n)] = \sum_{k=1}^n \sigma_k^2 < nC$$

sfruttando la precedente, si ottiene:

$$\Pr \{ |Y(n) - \mu| < h \} > 1 - \left(\frac{C}{n \cdot h^2} \right) > 1 - \varepsilon$$

e si arriva alla tesi:

$$\lim_{n \rightarrow \infty} \Pr \{ |Y(n) - \mu| < h \} = 1$$

Un caso speciale della legge debole dei grandi numeri (teorema di Bernoulli)

Si ottiene applicando la diseguaglianza di Chebyshev ad una variabile discreta, B , con legge binomiale di parametri n e p e, dunque, di media: np e varianza: $np(1-p)$.

Si osservi che risulta:

$$B = \sum_{i=1}^n X_i,$$

con X_i indipendenti e tutte bernoulliane di parametro p , ovvero risulta che B/n corrisponde ad una media campionaria sulle n variabili aleatorie che rappresentano gli esiti di n prove di Bernoulli (precisamente è la proporzione di successi su n prove).

Dunque:

$$\Pr \{ |B - np| < k\sqrt{np(1-p)} \} > 1 - \left(1/k^2 \right)$$

o meglio

$$\Pr \left\{ \left| \frac{B}{n} - p \right| < k \sqrt{\frac{p(1-p)}{n}} \right\} > 1 - \frac{1}{k^2}$$

da cui:

$$\Pr \left\{ \left| \frac{B}{n} - p \right| < \delta \right\} > 1 - \varepsilon \quad (+) \quad \text{con} \quad \delta \triangleq \sqrt{\frac{p(1-p)}{n\varepsilon}}, \quad \forall \varepsilon > 0$$

La (+) implica la convergenza in probabilità di B/n a p , mostrando in che misura il “rapporto tra casi favorevoli e casi possibili” possa approssimare la probabilità vera su un numero infinitamente grande di esperimenti indipendenti. A titolo di esempio, lanciando 1,000 volte (n) una moneta ($p=0.5$ per una qualunque delle due facce) e scegliendo di fissare $\delta=0.1$ si trova $\varepsilon = [p(1-p)/(n\delta^2)] = 1/40$; dunque la probabilità di sbagliare la stima del valore vero (0.5) di meno del 10% risulterebbe $>1-\varepsilon=39/40$.

Per altro verso, la (+) potrebbe anche essere riscritta come segue:

$$\Pr \left\{ \left| \frac{B - np}{n} \right| < \delta \right\} > 1 - \varepsilon$$

e vista così rivela che la (+) non può escludere il caso in cui la differenza $|B-np|$ - ovvero la differenza fra il numero di successi realizzati e il numero di successi attesi - diverge con la velocità della radice di “ n ” mentre “ n ” stessa cresce linearmente all’infinito!

Per eliminare questa debolezza della (+), ovvero rendere forte la (+), bisogna ricorrere al concetto di convergenza “forte” e al seguente teorema di Borel, basato appunto sulla convergenza forte:

$$\Pr \left\{ \lim_{n \rightarrow \infty} \left| \frac{B}{n} - p \right| = 0 \right\} = 1, \quad \text{e si dice che} \quad \frac{B}{n} \rightarrow p \quad \text{"con prob. 1", per } n \rightarrow \infty$$

intendendo quanto segue: $\forall \varepsilon > 0$ e quindi $\delta > 0 \quad \exists \tilde{n}, \tilde{n}+1, \tilde{n}+2, \dots$

$$Prob \left[|Y(n) - p| < \delta \cap |Y(\tilde{n}+1) - p| < \delta \cap |Y(\tilde{n}+2) - p| < \delta \cap \dots \right] < 1 - \varepsilon$$

con $\tilde{n} \gg 0$

Questo risultato forte della legge dei grandi numeri garantisce che a partire da un “certo n ” (\tilde{n}) in poi le realizzazioni della variabile aleatoria B/n (proporzione di successi) che risulteranno tanto vicine a p quanto più si vuole corrisponderanno ad un sottospazio di Ω che ha probabilità 1 (convergenza “con prob 1”). In questo senso, il verificarsi di realizzazioni apprezzabilmente distanti da p è un evento che ha probabilità $1-P(\Omega)$, dunque nulla.

Legge dei grandi numeri “debole” e legge “forte”

Per comodità, viene qui ripetuta la formulazione della **LGN debole** d’interesse per questo corso:

IPOTESI_1:

-) X_1, X_2, \dots var. al. indip. e id. distr. con varianze $\sigma_1^2, \sigma_2^2, \dots$

$$\text{tali che } \sigma_1^2 < C, \sigma_2^2 < C, \dots, \sigma_k^2 < C, \dots \text{ con } C < \infty$$

IPOTESI_2:

-) X_1, X_2, \dots var. al. id. distr. con $\frac{1}{n^2} \text{VAR} \left[\sum_{i=1}^n X_i \right] \hat{=} v^2(n)$
sotto ipotesi che $\lim_{n \rightarrow \infty} v^2(n) = 0$

TESI comune:

$$\lim_{n \rightarrow \infty} \text{Prob} [|Y(n) - \mu| = 0] = 1$$

e s’intende così: $\forall \varepsilon > 0$ e quindi $\delta > 0$ $\exists \tilde{n}$ (che dipende da δ , che dipende da ε)

per il quale risulta: $\text{Prob} [|Y(\tilde{n}) - \mu| < \delta] > 1 - \varepsilon$, con

$$\tilde{n} \gg 0$$

E qui di seguito, per comodità di confronto, ecco la **LGN forte**:

IPOTESI_1:

-) X_1, X_2, \dots v. a. indip. e id. distr., con $E[X] = \mu$ comune a tutte.

IPOTESI_2:

-) X_1, X_2, \dots v. a. indip. $\begin{cases} \text{di media } \mu_1, \mu_2, \dots & \text{con } \sum_{k=1}^{\infty} \frac{\sigma_k^2}{k^2} < \infty \\ \text{di varianza } \sigma_1^2, \sigma_2^2, \dots \end{cases}$

TESI comune:

$$\text{Prob} \left[\lim_{n \rightarrow \infty} |Y(n) - \mu| = 0 \right] = 1$$

e s’intende così: $\forall \varepsilon > 0$ e quindi $\delta > 0$ $\exists \tilde{n}, \tilde{n} + 1, \tilde{n} + 2, \dots$

$$\text{Prob} [|Y(n) - \mu| < \delta \cap |Y(\tilde{n} + 1) - \mu| < \delta \cap |Y(\tilde{n} + 2) - \mu| < \delta \cap \dots] < 1 - \varepsilon$$

$$\text{con } \tilde{n} \gg 0.$$

In alternativa al teorema di Borel ma per arrivare sempre allo stesso risultato pratico

Sia "X" la variabile aleatoria bernoulliana (real. 0 e 1) che indica il verificarsi o meno di un evento "A" e sia "n" il numero di esperimenti indipendenti.

$$\begin{aligned} \text{Siccome: } E[X] &= 0 \cdot \Pr(0) + 1 \cdot \Pr(1) \\ \Rightarrow \text{Prob}(A) &\equiv E[X] \end{aligned}$$

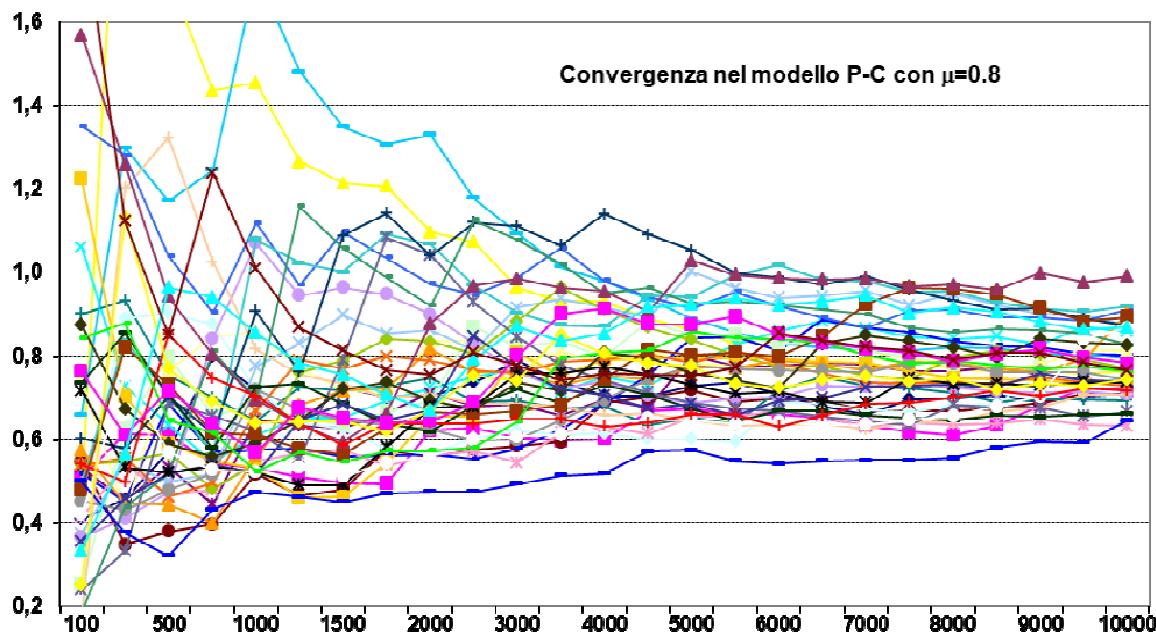
$$\text{allora: } \bar{X}(n) \hat{=} \frac{\sum_{i=1}^n X_i}{n} \xrightarrow{\text{pr 1}} \text{Prob}(A), \quad n \rightarrow \infty$$

Il rapporto fra il numero di volte in cui si verifica l'evento A e il numero sufficientemente grande di esperimenti indipendenti effettuati può essere preso come la Prob(A).

La dimostrazione della LGN forte è oggetto di corsi di dottorato di ricerca.

Ma una illustrazione sperimentale della modalità di convergenza di una media campionaria si può proporre facendo riferimento al modello produttore-consamatore studiato e riprodotto in Excel (simulato) grazie al metodo Monte Carlo. La media campionaria di interesse è quella relativa alla sequenza di variabili aleatorie W_1, W_2, \dots, W_n che rappresentano, rispettivamente, il tempo di attesa del primo utente arrivato e servito, quello del secondo arrivato e servito e così via fino a $n=10,000$. Ripetendo 40 simulazioni in Excel, ognuna da 10,000 osservazioni, è possibile ottenere 400,000 numeri; questi possono essere organizzati in una matrice avente un numero di colonne pari alle 40 simulazioni e un numero di righe pari a 10,000. E s'intende che i 40 numeri che è possibile leggere su una riga comunque scelta sono proprio 40 realizzazioni (indipendenti) della variabile aleatoria W_k , se k è l'indice della riga scelta. Adesso, passando a ragionare su una colonna comunque scelta fra le 40 a disposizione, ad esempio la j -ma, si immagini di calcolare una sequenza di medie aritmetiche cosiddette "a cipolla" e siano esse $y_j(k)$. La media aritmetica a cipolla si può ottenere sommando i primi " k " numeri (ad esempio i primi 100) e poi i primi 500 e poi 1000: in

pratica ogni media contiene la precedente (come le foglie di una cipolla) e la espande fino a $k=10,000$. Si osservi che la scelta di passare dai primi 100 a 500 e poi 1,000 è fatta solo per arrivare a 10,000 senza avere troppi punti da rappresentare nella fase iniziale. A tale scopo si osservi il grafico riportato sotto, dove ogni media aritmetica a cipolla ha un colore diverso e tutte tendono a convergere al valore atteso vero (0.8 nel caso raffigurato). E tendono a convergere a dispetto del fatto che le medie aritmetiche a cipolla non siano realizzazioni indipendenti!



Certamente le medie a cipolla convergono in forma debole: a 10,000 nessuna dista più di $|0.2|$ da 0.8; ma allungando le simulazioni fino a 30,000 e oltre resta confermato che le traiettorie colorate si mantengono tutte dentro una distanza ancor più piccola e ancor meglio centrata attorno 0.8. Infine, si osservi che quelle raffigurate sono appunto 40 traiettorie del processo stocastico definito dalla sequenza di var. al. “Media camp. a cipolla su 100,..., Media camp. a cipolla su 10,000,”. Allora si può pure capire che con il termine “convergenza con probabilità 1 sull’insieme delle traiettorie possibili” si deve intendere che la modalità forte di convergenza attribuisce probabilità nulla ad una traiettoria che possa scappare via verso il basso o verso l’alto, lontano da 0.8, per $k > 10,000$.

La legge forte dei grandi numeri e il metodo Monte Carlo

La legge forte dei grandi numeri esprime un risultato di convergenza matematica (quasi ovunque) della media aritmetica delle realizzazioni (indipendenti?) di una variabile aleatoria, X , al valore atteso della stessa variabile.

Formalmente:

$$Pr \left\{ \lim_{n \rightarrow \infty} \frac{1}{n} (x_1 + x_2 + \dots + x_n) = E[X] \right\} = 1$$

quando, fissato arbitrariamente $\varepsilon > 0$ e comunque piccolo, esiste sempre un valore \tilde{n} tale che:

$$\forall n > \tilde{n} \quad \text{risulti: } \left| \frac{x_1 + \dots + x_n}{n} - E[X] \right| < \varepsilon, \quad \text{"con prob. 1".}$$

Sul piano teorico, rappresentando con X (a valori 0 e 1) l'occorrenza o meno di un evento (A) e con "n" il numero di esperimenti indipendenti, è facile riconoscere che $P(A) = E[X]$ e quindi il rapporto tra il numero di volte ($x_1 + \dots + x_n$) in cui si osserva l'evento A ed il totale (n) degli esperimenti tende a stimare la $P(A)$ "con certezza".

Sul piano pratico, si vuole illustrare il metodo Monte Carlo per il calcolo approssimato del seguente integrale:

$$I \triangleq \int_a^b g(x) dx$$

A tal fine, s'introduce la variabile aleatoria "uniforme", $U[a,b]$, cioè definita da una densità uniforme nell'intervallo [a,b]:

$$f_U(u) \triangleq \begin{cases} (b-a)^{-1} & a \leq u \leq b \\ 0 & \text{altrimenti} \end{cases}$$

e si considera il seguente valore atteso, riferito non ad u ma a $g(u)$:

$$E[g(u)] \triangleq \int_a^b g(u) f_U(u) du$$

Osservando che risulta: $I = E[g(u)](b-a)$ e sfruttando la legge forte dei grandi numeri, si può stimare $E[g(u)]$ e quindi I con la seguente:

$$\frac{\sum_{i=1}^n g(u_i)}{n}$$

dove u_1, u_2, \dots, u_n è un campione sufficientemente grande di realizzazioni indipendenti della U .

Il metodo Monte Carlo rimanda al **problema di generare il campione** (u_1, u_2, \dots, u_n) di punti in corrispondenza dei quali calcolare i valori ($g(u_1), g(u_2), \dots, g(u_n)$) assunti dalla funzione integranda di I .

A tal fine si consideri il seguente ragionamento.

Sia $X\{0,\infty\}$ una generica variabile aleatoria a valori reali non negativi e con funzione di distribuzione $F_X(x)$ arbitraria, purché invertibile, e sia $U\{0,1\}$ una seconda variabile a valori reali compresi tra 0 e 1, con distribuzione uniforme. Si considerino adesso i valori reali ottenibili con la seguente operazione:

$$x \doteq F_X^{-1}(u), \text{ dove } u \text{ è una realizzazione della } U\{0,1\}$$

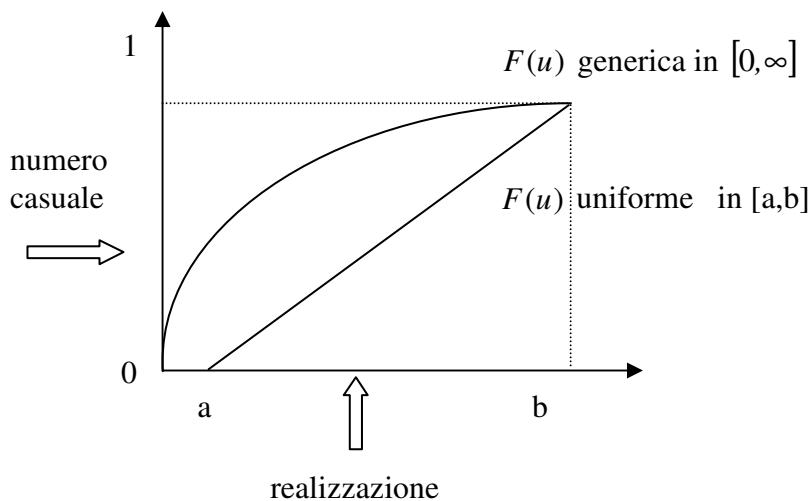
Allora risulta:

$$\Pr\{X \leq x\} \doteq F_X(x) = F_X(F_X^{-1}(u)) = u = \Pr\{U \leq u\}$$

e quindi anche la x ottenuta come $F_X^{-1}(u)$ è una realizzazione della $X\{0,\infty\}$.

Quello appena illustrato è il cosiddetto metodo Monte Carlo per la generazione di realizzazioni (indipendenti) di una variabile aleatoria, a partire da numeri (indipendenti) completamente casuali, compresi tra 0 e 1.

E' noto come metodo della "trasformazione inversa" ed è sintetizzato graficamente qui di seguito:



Il metodo della trasformazione inversa è fondato sulla disponibilità di un generatore di numeri casuali che determini il numero casuale compreso tra zero ed uno, con il quale posizionarsi sull'ordinata per poi leggere l'ascissa corrispondente.

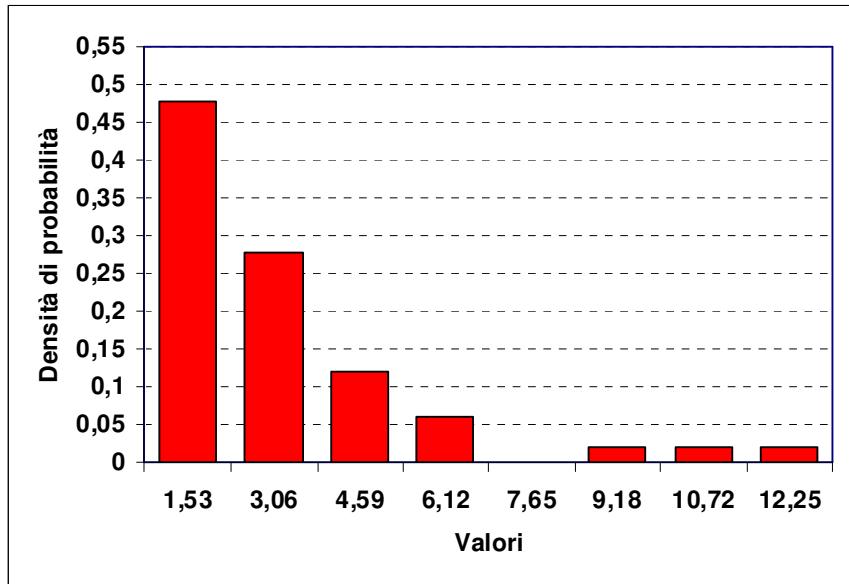
La stima dei parametri (di input)

La stima dei parametri è una fase dell’analisi statistica che gioca un ruolo basilare nella modellazione dei dati che saranno d’ingresso rispetto al modello da studiare. Ad esempio, ritornando al modello produttore-buffer-consumatore, è immediato riconoscere che, per formulare il problema della stima della funzione di distribuzione del tempo d’attesa degli oggetti nel buffer, occorre anzitutto definire o scegliere il modello di distribuzione che meglio si adatta a rappresentare il tempo di produzione (interarrivo) e il tempo di consumo (servizio).

Il primo passo della stima dei parametri è quello di costruire un istogramma – v. appendice – a partire da osservazioni reali, indipendenti, ed individuare in prima approssimazione una famiglia di funzioni (ad esempio la famiglia di Erlang o di Weibull) alla quale potrebbe appartenere quella in esame e per la quale si dovrà, allora, stimare il o i parametri che precisano la funzione (parametri “n” e “l” per la Erlang, parametri “a” e “l” per la Weibull).

In realtà, fase ancora precedente a questa è la verifica che quei dati siano effettivamente estratti da una stessa distribuzione (e.g. test di *Kruskal-Wallis*) e siano indipendenti (e.g. test di *von Neumann*). Non si dimentichi che l’indipendenza delle osservazioni è un’ipotesi fondamentale per l’applicazione di tanti risultati già studiati; in particolare, per la statistica detta “varianza campionaria” e poi per poter applicare il teorema limite centrale alla statistica “media campionaria”.

Prima, però, di qualunque confronto o test, grafico e non, è necessario stimare i parametri caratterizzanti di ognuna delle distribuzioni che si sono selezionate, poiché, nel passo precedente, se ne è semplicemente determinato il “tipo”. Ad esempio osservando l’istogramma seguente, costruito su dati sperimentali si sarebbe indotti a pensare di avere a che fare con una legge esponenziale.



Di essa però non conosciamo il tasso λ , senza il quale non possiamo effettuare, in prima approssimazione, alcun confronto con la reale densità esponenziale.

La stima dei parametri, come il tasso λ cui facevamo riferimento, è una procedura molto delicata e importante che affronteremo ora nel dettaglio.

Ultimi passi per la scelta della distribuzione statistica che meglio modella i nostri dati sono i test di bontà di adattamento (*Goodness of fit texts* – test della chi-quadrato e di *Kolmogorov-Smirnov*), di cui si parlerà più avanti e che si serviranno dei parametri ricavati in questa fase.

Due metodi saranno presi in considerazione:

- la *stima a massima verosimiglianza* (*o maximum likelihood estimation – “MLE”*);
- il *metodo dei momenti* – “MOM”.

Il primo consiste nel definire una funzione, detta “funzione di verosimiglianza”, e stimare i parametri trovandone il massimo (come si vedrà, infatti, la funzione risulterà incognita nei parametri).

Il secondo prevede, invece, di eguagliare i momenti di ordine k della distribuzione ipotizzata, con i rispettivi stimatori di momenti di ordine k dai dati, risolvendo il sistema incognito nei parametri.

Come si noterà, entrambi fanno ancora riferimento ai dati poiché si basano sulle statistiche da essi ricavate (media campionaria, varianza campionaria, ecc).

Il metodo dei momenti

Un generico momento di ordine k di una variabile aleatoria X , cioè il valore atteso della variabile elevata al momento k è così definito:

$$E[X^k] = \begin{cases} \sum_x x^k f(x) & \text{se } X \text{ è discreta} \\ \int_{\mathbb{R}} x^k f(x) dx & \text{se } X \text{ è continua} \end{cases}$$

Consideriamo delle variabili casuali i.i.d. X_1, \dots, X_n estratte da una stessa funzione densità $f(x)$.

Il metodo dei momenti (MOM) afferma che lo stimatore di ordine k è:

$$E[X^k] \rightarrow \sum_{i=1}^n \frac{X_i^k}{n}.$$

che è funzione dei parametri. Il MOM consiste nell'eguagliare a zero i momenti di ordine $1, 2, \dots, n$ con le corrispondenti stime di momenti fino ad ottenere tante equazioni quanti sono i parametri incogniti e poi risolverle contemporaneamente per ottenere le stime richieste.

A volte il MOM dà stimatori su cui è facile lavorare, altre volte può fornire stimatori distorti.

Di seguito alcuni esempi di stimatori MOM:

■ Per $= E[X_i]$ (ordine $k=1$) lo stimatore è $\bar{X} = \sum_{i=1}^n \frac{X_i}{n}$

■ Per $E[X_i^2]$ (ordine $k=2$) lo stimatore è $\sum_{i=1}^n \frac{X_i^2}{n}$

■ Per $VAR[X_i]$ lo stimatore è $\frac{n-1}{n} S^2$, poiché $VAR[X_i] = E[X_i^2] - (E[X_i])^2$ da cui:

$$\frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2 = \frac{\sum_{i=1}^n X_i^2 - n\bar{X}^2}{n} = \frac{n-1}{n} S^2$$

NB: S^2 è distorto, ma è possibile usare anche solo S^2 .

Esempi (metodo MLE)

Legge esponenziale: solo un parametro da stimare. Dunque:

$$E[X] = \frac{1}{\lambda} = \bar{X} \quad \text{da cui} \rightarrow \hat{\lambda} = \frac{1}{\bar{X}}.$$

Legge di Poisson: supponiamo che X_1, \dots, X_n siano i.i.d da una distribuzione di Poisson di parametro λ . Poiché $\lambda = E[X_i]$, un stimatore di tipo MOM per λ è \bar{X} . Ma dato che $\lambda = Var(X_i)$ allora un altro stimatore MOM per λ è $\frac{n-1}{n}S^2$.

Legge normale: supponiamo di avere una distribuzione normale di media μ e varianza σ^2 . Poiché ci sono due parametri da stimare avremo un sistema in 2 equazioni:

$$\begin{cases} E(X) = \mu = \bar{X} \\ E(X^2) = \mu^2 + \sigma^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 \end{cases}$$

Perciò gli stimatori MOM sono:

$$\hat{\mu} = \bar{X} \quad \text{e} \quad \hat{\sigma}^2 = \frac{n-1}{n} S^2.$$

Legge iperesponenziale: nella distribuzione iperesponenziale

$$F_y(y) = \sum_{i=1}^n \alpha_i (1 - e^{-\lambda_i y}) \quad \text{per} \quad y \geq 0$$

Ci sono tre parametri da stimare: α , λ_1 e λ_2 e valgono: $\alpha_1 = \alpha$ e $\alpha_2 = 1 - \alpha$. In questo caso ci servirebbero i momenti del 1°, 2° e 3° ordine. Ma abbiamo a disposizione solo la media e la varianza campionaria (ci rimane un livello di libertà!). Perciò si sceglie di impostare un'ulteriore condizione per rendere determinato il sistema:

$$\frac{1}{\lambda_1 + \lambda_2} = \frac{\frac{\alpha}{\lambda_1} + \frac{1-\alpha}{\lambda_2}}{2}$$

Consideriamo i seguenti risultati:

- | | |
|---|--|
| a. Momento del I ordine | $E[Y] = \sum_{i=1}^n \frac{\alpha_i}{\lambda_i} = \left(\frac{\alpha}{\lambda_1} + \frac{1-\alpha}{\lambda_2} \right)$ |
| b. Momento del II ordine | $E[Y^2] = \sum_{i=1}^n \frac{2\alpha_i}{\lambda_i^2} = 2 \left(\frac{\alpha}{\lambda_1^2} + \frac{1-\alpha}{\lambda_2^2} \right)$ |
| c. Varianza | $VAR[Y] = E[Y^2] - E^2[Y]$ |
| d. Media campionaria | \bar{X} |
| e. Varianza campionaria | $\frac{n-1}{n} S^2$ |
| f. Stimatore: | |
| i. 1° ordine: $M_1 \rightarrow \bar{X}$; | |
| ii. 2° ordine: $M_2 \rightarrow S^2 - E^2[Y]$. | |

Da notare che si è usata la sola S^2 e non il suo stimatore distorto. Il nostro sistema di tre equazioni in tre incognite diventa perciò:

$$\begin{cases} \frac{1}{\lambda_1 + \lambda_2} = \frac{\frac{\alpha}{\lambda_1} + \frac{1-\alpha}{\lambda_2}}{2} \\ \bar{X} = \frac{\alpha}{\lambda_1} + \frac{1-\alpha}{\lambda_2} \\ S^2 = 2 \left(\frac{\alpha}{\lambda_1^2} + \frac{1-\alpha}{\lambda_2^2} \right) - \left(\frac{\alpha}{\lambda_1} + \frac{1-\alpha}{\lambda_2} \right)^2 \end{cases}$$

Osservando che il secondo membro della prima equazione è uguale a quello della seconda:

$$\lambda_1 = \frac{2}{\bar{X}} - \lambda_2$$

Dalla seconda equazione ricaviamo:

$$\hat{\alpha} = \frac{\lambda_1(\lambda_2 \bar{X} - 1)}{\lambda_2 - \lambda_1}$$

Sostituendo λ_1 :

$$\alpha = 1 - \frac{\lambda_2 \bar{X}}{2}$$

Osserviamo che:

$$\frac{\alpha}{\lambda_1} = \frac{1-\alpha}{\lambda_2} = \frac{\bar{X}}{2} \quad \frac{1-\alpha}{\lambda_2^2} = \frac{\bar{X}}{2\lambda_2} \quad \frac{\alpha}{\lambda_1^2} = \frac{\bar{X}^2}{2(2-\lambda_2 \bar{X})}$$

La terza equazione diventa:

$$S^2 = 2 \left(\frac{\bar{X}^2}{2(2-\lambda_2 \bar{X})} + \frac{\bar{X}}{2\lambda_2} \right) - \left(\frac{\bar{X}}{2} + \frac{\bar{X}}{2} \right)^2$$

da cui:

$$\lambda_2(2-\lambda_2 \bar{X}) = \frac{2\bar{X}}{S^2 + \bar{X}^2}$$

Manipolando l'ultimo risultato giungiamo ad un'equazione di 2° grado in λ_2 :

$$\lambda_2^2 \bar{X} (S^2 + \bar{X}^2) - 2\lambda_2 (S^2 + \bar{X}^2) + 2\bar{X} = 0$$

che ammette come soluzioni:

Da cui:

$$\hat{\lambda}_1 = \frac{1}{\bar{X}} + \frac{1}{\bar{X}} \sqrt{\frac{(S^2 - \bar{X}^2)}{(S^2 + \bar{X}^2)}}$$

$$\hat{\lambda}_2 = \frac{1}{\bar{X}} - \frac{1}{\bar{X}} \sqrt{\frac{(S^2 - \bar{X}^2)}{(S^2 + \bar{X}^2)}}$$

(si suppone che $S^2 \geq \bar{X}^2$).

Per completezza riscriviamo il terzo coefficiente:

$$\hat{\alpha} = \frac{\lambda_1(\lambda_2 \bar{X} - 1)}{\lambda_2 - \lambda_1}$$

Esempio dal foglio Excel:

<i>α1</i>	<i>λ1</i>	<i>λ2</i>
0,9	5	0,5

0,038035	1,182715	0,220085	0,005408	0,09198
0,409635	0,159324	0,128586	0,04661	0,283353
0,181822	0,10365	3,393875	0,407684	0,207476
0,137333	0,042202	0,170603	0,039083	3,563589
0,306768	0,138842	0,044659	0,147523	0,849025
0,293062	0,49593	0,189722	0,06778	0,115149
0,137179	0,130769	0,277224	0,046152	0,00011
0,672394	0,424989	0,152111	0,003551	0,183039

0,275342	0,043464	0,331043	0,101144	0,143709
0,07655	0,143325	4,654475	1,515698	0,396286

$$\hat{\lambda}_1 = 3,827974$$

$$\hat{\lambda}_2 = 0,48794$$

$$\hat{\alpha} = 0,886944$$

Legge beta: per una distribuzione beta di parametri a e b , la corrispondente densità è:

$$f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1} (1-x)^{b-1}, 0 < x < 1.$$

Avremo che:

$$E[X] = \frac{a}{a+b} \quad \text{e} \quad Var(X) = \frac{ab}{(a+b)^2(a+b+1)}$$

Il metodo della massima verosimiglianza.

Consideriamo delle variabili casuali i.i.d. X_1, \dots, X_n con una funzione di densità $f(x)$.

Definiamo nel modo seguente la **funzione di verosimiglianza**:

$$L(\theta) \equiv f(X_1) \cdot f(X_2) \cdots f(X_n) = \prod_{i=1}^n f(X_i)$$

Poiché conosciamo le X_i , tale funzione ha come unica incognita il parametro (o i parametri) θ che si vuole valutare.

Lo **stimatore a massima verosimiglianza** (MLE) di θ è il valore $\hat{\theta}$ che massimizza la funzione appena definita, ponendo uguali a zero la sua derivata secondo θ . Se i parametri sono più di uno, bisogna risolvere un sistema di equazioni alle derivate parziali rispetto ad ogni parametro. Nella pratica si suole massimizzare, in luogo della $L(\theta)$, il logaritmo di essa, ovvero

$$l(\hat{\theta}) = \ln[L(\theta)]$$

che è nota come **funzione di verosimiglianza ridotta**. Il tutto funziona poiché il logaritmo di una funzione ne segue in qualche modo l'andamento e, data la natura delle distribuzioni più comuni (per esempio l'esponenziale), la funzione di verosimiglianza ridotta rende i calcoli più semplici.

Per essere sicuri che $\hat{\theta}$ sia un massimo, piuttosto che un minimo o un punto di flesso, la derivata di $l(\theta)$, valutata in $\hat{\theta}$ deve essere negativa.

Infine si tenga presente il seguente risultato: se $\hat{\theta}$ è un MLE per un parametro θ e $h(\cdot)$ è una funzione "one-to-one", allora $h(\hat{\theta})$ è il MLE di $h(\theta)$.

Esempi (metodo MLE)

Legge esponenziale: a partire dalla densità esponenziale

$$f_X(X) = \lambda e^{-\lambda X}$$

λ è il parametro che vogliamo stimare. La funzione di verosimiglianza ridotta è pari a:

$$\ell(\lambda) = \ln L(\lambda) = \ln \left(\prod_{i=1}^n \lambda e^{-\lambda X_i} \right) = \ln \left(\lambda^n e^{-\lambda \sum_{i=1}^n X_i} \right) = n \ln \lambda - \lambda \sum_{i=1}^n X_i$$

Le quantità X_i sono le realizzazioni della legge esponenziale, cioè i dati che abbiamo.

Consideriamo il logaritmo della funzione appena trovata:

$$\frac{d\ell}{d\lambda} = \frac{n}{\lambda} - \sum_{i=1}^n X_i = 0$$

da cui si ottiene il valore del parametro cercato:

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^n X_i} = \frac{1}{\bar{X}}$$

dove \bar{X} è esattamente la media campionaria.

Legge normale: X_1, \dots, X_n sono estratte da una normale con media μ e varianza σ^2 .

Questi ultimi sono i parametri da stimare. La funzione a massima verosimiglianza sarà:

$$\begin{aligned} L(\mu, \sigma^2) &= \prod_{i=1}^n f(x_i) \\ &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(x_i - \mu)^2}{\sigma^2}} \\ &= \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{1}{2}\sum_{i=1}^n \frac{(x_i - \mu)^2}{\sigma^2}} \end{aligned}$$

Calcoliamone il logaritmo:

$$\ln(L(\mu, \sigma^2)) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2$$

E quindi la derivata in μ e in σ^2 :

$$\frac{\partial}{\partial \mu} (L(\mu, \sigma^2)) = \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) \equiv 0$$

da cui:

$$\hat{\mu} = \bar{X}$$

(con \bar{X} media campionaria) e:

$$\frac{\partial}{\partial \sigma^2} \ln(L(\sigma^2)) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \bar{x})^2 \equiv 0$$

$$\Rightarrow -n\sigma^2 + \sum_{i=1}^n (x_i - \bar{x})^2 = 0$$

da cui:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}$$

che già sappiamo essere pari a $\frac{n-1}{n} S^2$.

Legge di Weibull: $\{t_i, i=1, \dots, n\}$ sono le osservazioni (tempi di vita degli n componenti) e r sono i fallimenti osservati (t_r = tempo al guasto).

Per semplicità chiamiamo $x_i = \min\{t_i, t_r\}$. Ecco distribuzione e densità di Weibull:

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

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

Dobbiamo stimare 2 parametri (λ e α). La funzione a massima verosimiglianza è:

$$\begin{aligned} L(\lambda, \alpha) &= \prod_{i=1}^r f(t_i | \lambda, \alpha) \cdot \prod_{i=r+1}^n R(t_r | \lambda, \alpha) \\ &= \prod_{i=1}^r \lambda \alpha x_i^{\alpha-1} e^{-\lambda x_i^\alpha} \cdot \prod_{i=r+1}^n e^{-\lambda x_i^\alpha} \\ &= \lambda^r \alpha^r \left(\prod_{i=1}^r x_i \right)^{\alpha-1} e^{-\lambda \sum_{i=1}^n x_i^\alpha} \end{aligned}$$

Il suo logaritmo:

$$\ln L(\lambda, \alpha) = r \ln \lambda + r \ln \alpha + (\alpha - 1) \sum_{i=1}^r \ln x_i - \lambda \sum_{i=1}^n x_i^\alpha$$

Deriviamo nei parametri ed eguagliamo a 0. Otteniamo:

$$\frac{r}{\lambda} - \sum_{i=1}^n x_i^\alpha = 0 \quad \text{e} \quad \frac{r}{\alpha} - \sum_{i=1}^r \ln x_i - \lambda \sum_{i=1}^n x_i^\alpha \ln x_i = 0$$

Non esistono soluzioni in forma chiusa per λ e α . Comunque si può scrivere λ in funzione di α :

$$\hat{\lambda} = \frac{r}{\sum_{i=1}^n x_i^\alpha}$$

Sostituendo nella seconda equazione si ottiene:

$$\frac{r}{\alpha} + \sum_{i=1}^r \ln x_i - \frac{r \sum_{i=1}^n x_i^\alpha \ln x_i}{\sum_{i=1}^n x_i^\alpha} = 0$$

che si risolve iterativamente. Trovato il valore di α , si potrà usare per risolvere λ .

La distribuzione Student-t

Nei casi in cui non è possibile disporre di un campione di numerosità adeguata, ma si vuole, comunque, costruire un intervallo di confidenza (esatto) per la media si può ricorrere alla distribuzione Student-t. Essa permette di usare la varianza campionaria al posto della varianza vera (non conosciuta), però a patto che le osservazioni sperimentali a disposizione siano quelle di una legge normale.

A partire da una variabile aleatoria normale standard, Z , ed una seconda indipendente dalla prima, X_γ^2 , cioè una chi-quadrato con γ gradi di libertà, si definisce la seguente variabile aleatoria:

$$T_\gamma \hat{=} \frac{Z}{\sqrt{X_\gamma^2 / \gamma}}$$

alla quale è associata (si dimostra) la seguente funzione densità:

$$f_T(t) = \frac{\Gamma((\gamma+1)/2)}{\Gamma(\gamma/2)\sqrt{\pi\gamma}} \left(1 + \frac{t^2}{\gamma}\right)^{-(\gamma+1)/2}, \quad -\infty < t < +\infty$$

detta, appunto, densità della Student-t con γ gradi di libertà.

La funzione generatrice dei momenti non esiste, però media e varianza si calcolano facilmente, risultando:

$$E[T] = 0 \quad \text{e} \quad \text{VAR}[T] = \gamma/(\gamma-2), \quad \{\rightarrow 1 \text{ per } \gamma \rightarrow \infty\}$$

L'utilità della T_γ nella costruzione degli intervalli di confidenza per la media campionaria poggia sui seguenti risultati:

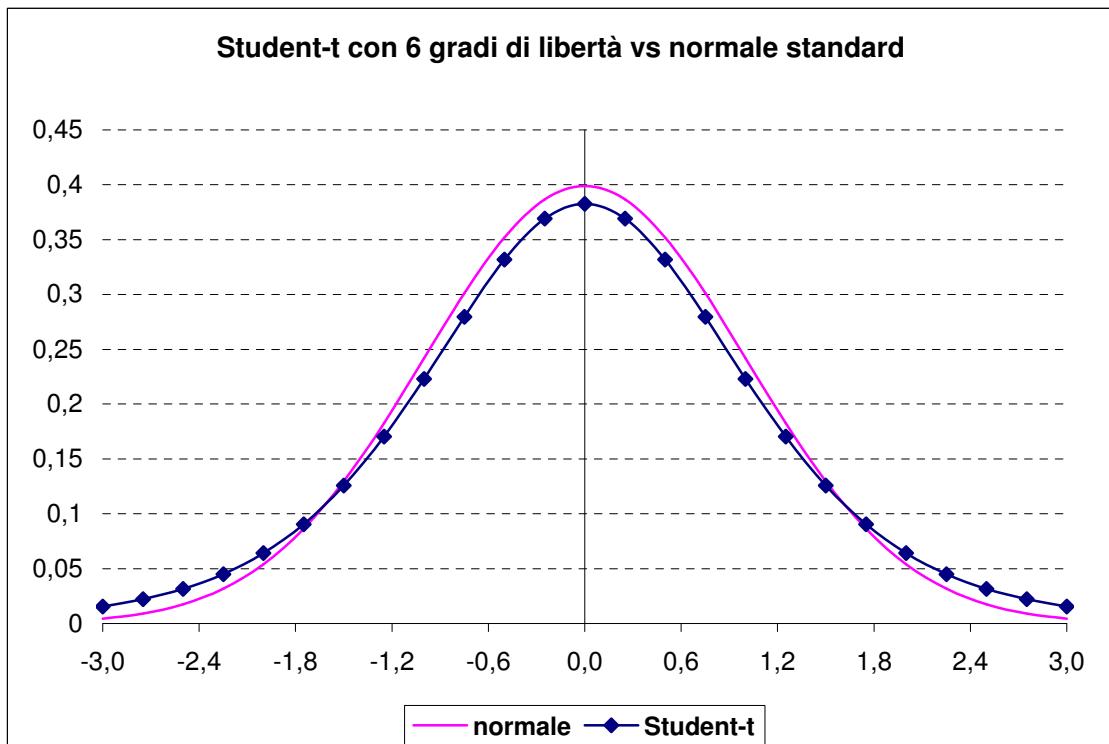
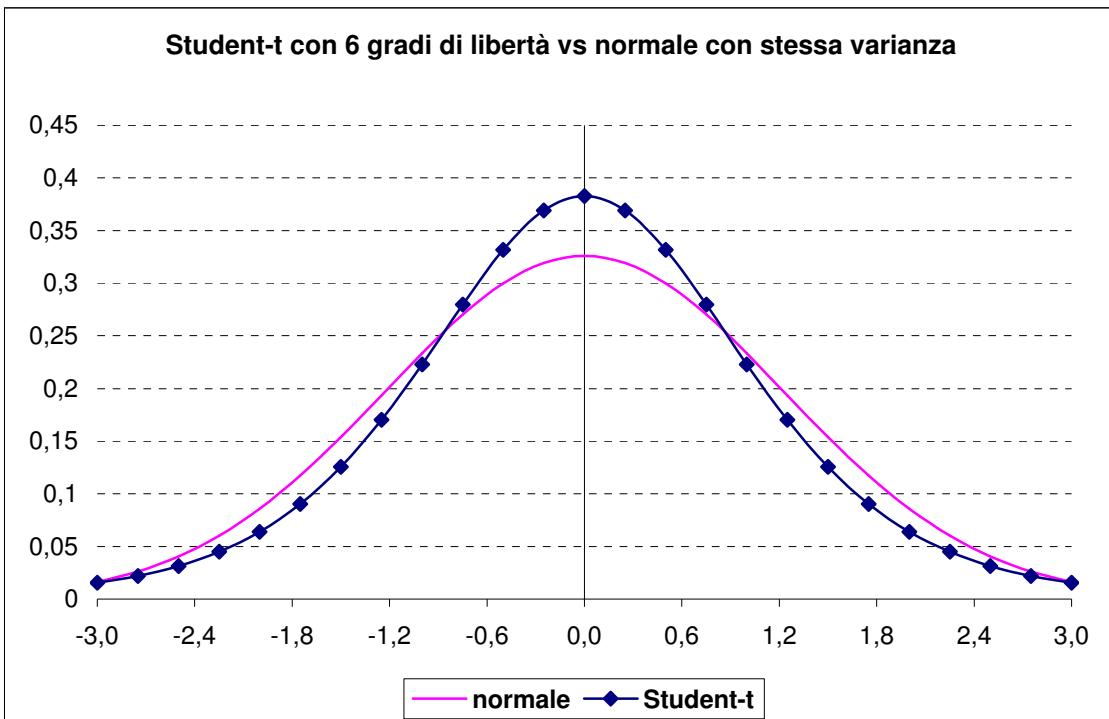
$$(1) \quad \frac{\bar{X} - S/\sqrt{n}}{\sigma/\sqrt{n}} = Z \quad \text{e} \quad (2) \quad \frac{(n-1)S^2}{\sigma^2} = X_{\gamma=n-1}^2$$

Infatti:

$$\frac{\bar{X} - S/\sqrt{n}}{\sigma/\sqrt{n}} = T_{\gamma=n-1} \quad \text{e perciò l'intervallo:} \quad \left[\bar{X} - \frac{S}{\sqrt{n}} t_{\alpha/2}, \bar{X} + \frac{S}{\sqrt{n}} t_{\alpha/2} \right]$$

è un intervallo di confidenza per la media di un campione relativamente piccolo, $n < 30$, di realizzazioni indipendenti estratte da una legge normale di varianza non nota.

Confronti fra la legge Student-T e la normale



Test d'ipotesi: concetti di base e test sulla media

Facendo sempre riferimento al modello produttore-consumatore, non dovrebbe essere difficile immaginare che fra i vari obiettivi dell'analisi statistica del sistema ci possa essere almeno uno dei due seguenti:

1. “stabilire se il tempo medio di soggiorno dei clienti nel buffer d'attesa è superiore, inferiore o approssimativamente uguale ad un valore fissato”
2. “stabilire se la forma della distribuzione della variabile aleatoria che rappresenta il tempo di attesa dei clienti nel buffer è assimilabile ad una forma fissata”

Ancora, si può immaginare di avere a disposizione due sistemi che funzionano con due diverse politiche di gestione del buffer (inserimento di nuovi prodotti da parte del produttore ed estrazione degli stessi da parte del consumatore) e si potrebbe voler stabilire qual'è il sistema dove si registrano tempi medi di permanenza nel buffer minori. In tal caso, si dovrebbe ragionare sulla differenza di due medie.

In tutti i casi appena enunciati, la Statistica suggerisce di ragionare formulando un'opportuna ipotesi e poi di tentando di confutarla con un metodo adeguato. Il metodo si chiama Test d'ipotesi. L'ipotesi formulata è detta ipotesi *nulla* ed è quella che ha un fondamento di convinzione scientifica: per questo è anche detta ipotesi del ricercatore o, più modestamente, ipotesi di lavoro. Ad esempio, nel modello produttore-consumatore con interratrivi esponenziali di parametro $\lambda = 0.8 \text{ u.t.}^{-1}$, tempi di consumo esponenziali di parametro $\mu = 1 \text{ u.t.}^{-1}$ e gestione FIFO del buffer l'ipotesi del ricercatore è che la media dei tempi d'attesa, w_0 , calcolata solo su quelli che aspettano, risulti: $w_0 = 1 / (1 - \rho) = 5 \text{ u.t.}$, dove $\rho \doteq \lambda / \mu = 0.8$. Allora, se si pensa che il sistema reale sotto osservazione si comporti effettivamente secondo quanto previsto dal modello con arrivi e consumi esponenziali, si proverà a confutare l'ipotesi (nulla) che la media dei tempi d'attesa nel sistema reale (w_{re}) sia, appunto, pari a 0.8.

Il punto è quello di stabilire come si possa confutare quell'ipotesi, ovvero di progettare un test.

Prima di entrare nel dettaglio della progettazione del test, si osservi che l'ipotesi nulla rimarrà in campo fino a quando non si riuscirà a rigettarla e, viceversa, se con un certo test (progettato) si riuscirà a rigettare l'ipotesi nulla, allora si potrà concludere che

risulta accettata la cosiddetta ipotesi *alternativa* che, di fatto, è la semplice negazione della prima. Nell'esempio del produttore-consumatore, l'ipotesi alternativa consiste nella dichiarazione seguente: la media dei tempi d'attesa nel sistema reale è diversa da 0.8.

Sul piano formale le ipotesi nulla e alternativa, con le notazioni del modello produttore-consumatore, si formulano, rispettivamente, alla seguente maniera:

$$H_0: w_{re} = w_0$$

$$H_1: w_{re} \neq w_0$$

Passando, finalmente, alla progettazione di uno specifico test per la media dei tempi d'attesa, si osserva, in via preliminare, che la media del sistema reale non può che essere stimata attraverso la media di un campione di osservazioni (indipendenti e di numerosità sufficientemente grande). Infatti, partendo da qui, si può ricorrere al fatto (sperimentale) che la media $\bar{W}(k)$, costruita su un campione di k tempi d'attesa rilevati, tende ad essere distribuita come una normale, di media pari a w_{re} , per $k \rightarrow \infty$. A partire da questo, si conviene di “recuperare” dal sistema reale tanti (n) campioni di dimensione k , in modo da avere tante medie campionarie $\bar{W}_1(k)$, $\bar{W}_2(k)$, ..., $\bar{W}_n(k)$. La cosiddetta grande media, cioè la media delle medie campionarie (è anch'essa una campionaria, ma costruita con n osservazioni “tendenti alla normale”!), indicata con $\bar{\bar{W}}(n, k)$ continua ad essere corretta per stimare la w_{re} , ma, in più, per essa vale anche il seguente risultato:

$$\frac{\bar{\bar{W}}(n, k) - w_0}{S / \sqrt{n}} \approx T_{n-1}, \quad k \rightarrow \infty$$

dove S indica la solita radice quadrata della varianza campionaria, riferita alla grande media e T_{n-1} indica la statistica distribuita secondo la legge di Student con $n-1$ gradi di libertà.

Attorno al risultato appena stabilito può essere progettato un test perché si osserva che, se l'ipotesi nulla è vera, allora la probabilità che $\bar{\bar{W}}(n, k)$ risulti abbastanza vicina a w_0 corrisponde alla probabilità che il valore della statistica T_{n-1} sia compreso in un certo intervallo di valori. Formalmente:

$$\Pr(-t_{n-1;1-\alpha/2} \leq \frac{\bar{W}(n,k) - w_0}{S/\sqrt{n}} \leq t_{n-1;1-\alpha/2}) = 1 - \alpha$$

dove i valori $-t_{n-1;1-\alpha/2}$ e $t_{n-1;1-\alpha/2}$ sono i quantili della legge di Student che risultano dopo aver fissato la probabilità “ $1 - \alpha$ ” ad un valore adeguato, quale potrebbe essere il valore 0.9, oppure 0.95 o, al massimo, 0.99. Nel linguaggio della Teoria dei Test, si dice che “ α ” è il *livello di significatività* del test e l’intervallo $[-t_{n-1;1-\alpha/2}; t_{n-1;1-\alpha/2}]$ è detto *regione di accettazione*, nel senso che, se risulta:

$$\frac{\bar{W}(n,k) - w_0}{S/\sqrt{n}} \in [-t_{n-1;1-\alpha/2}; t_{n-1;1-\alpha/2}]$$

cioè se il valore della statistica appartiene alla regione di accettazione, allora rimane stabilito l’esito (positivo per il ricercatore!) del test nell’evidenza sperimentale che *non consente di rigettare l’ipotesi nulla*. Accanto alla regione di accettazione, rimane pure definita la cosiddetta *regione di rifiuto* che, nel caso in questione, è rappresentata dall’unione dei due intervalli:

$$(-\infty, -t_{n-1;1-\alpha/2}] \text{ e } [t_{n-1;1-\alpha/2}, +\infty).$$

Se il valore della statistica appartiene alla regione di rifiuto, allora rimane stabilito l’esito (negativo per il ricercatore!) del test nell’evidenza sperimentale che *consente di rigettare l’ipotesi nulla*.

Si tenga presente che, nella pratica, l’esito di non rigettare l’ipotesi nulla equivale ad accettarla e ritenerla vera fino a futura prova contraria, cioè fino a che un eventuale altro test stabilisca il contrario. E il contrario consisterebbe nel fatto che il valore di un’altra statistica, diversa da quella di Student ma ugualmente “valida”, vada a cadere nella cosiddetta regione di rifiuto (o regione critica). Per capire in che senso la statistica di Student è stata valida per progettare il test appena definito e quindi per avere le linee guida per cercarne una seconda (se esiste!) occorre riflettere sul fatto che grazie al risultato

$$\frac{\bar{W}(n,k) - w_0}{S/\sqrt{n}} \approx T_{n-1}, \quad k \rightarrow \infty$$

siamo stati in grado di incorporare la differenza di nostro interesse, $\bar{W}(n,k) - w_0$, in una variabile aleatoria avente una distribuzione nota che ci ha permesso di tradurre la piccolezza di quella differenza nell’appartenenza della realizzazione (costruita con i

campioni di osservazioni reali) della variabile aleatoria ad un determinato intervallo (regione di accettazione). Il ragionamento è stato dunque un ragionamento da prova di necessità: se l'ipotesi nulla è vera allora i valori delle possibili realizzazioni (e ogni volta che si esegue il test se ne costruisce una e una sola) della statistica individuata devono appartenere alla regione di accettazione e, viceversa, se la realizzazione costruita in un'esecuzione del test risulta appartenere non alla regione di accettazione ma a quella di rifiuto, allora si conclude che l'ipotesi nulla debba essere rigettata. L'esplicitazione di questo ragionamento ne illustra immediatamente i limiti ed è facile riconoscere che questi limiti si traducono in due possibilità di errore proprie del test. Le possibilità di errore risiedono nella seguente doppia eventualità: la prima è che un'ipotesi (nulla) vera possa apparire falsa solo per un caso e ciò conduce ad un errato rigetto (che mortifica il ricercatore che, invece, aveva visto giusto!); la seconda eventualità è che un'ipotesi (nulla) falsa possa apparire vera solo per un caso e ciò conduce ad una errata accettazione (o, meglio, al mancato rigetto).

In sintesi, si può riconoscere che esistono due tipi di errori:

- 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*).

Il primo tipo di errore è considerato peggiore del secondo.

In ogni caso:

$$\Pr(\text{errore di I specie}) = \Pr(\text{rigettare } H_0 \mid H_0 \text{ è vera}) = \alpha$$

$$\Pr(\text{errore di II specie}) = \Pr(\text{accettare } H_0 \mid H_0 \text{ è falsa}) = \beta.$$

È interessante osservare che l'occorrenza di un errore di 1^a specie dipende anche dalla scelta, soggettiva, di fissare α ad un valore relativamente alto. Supponendo di avere calcolato il valore della statistica di test corrispondente alle osservazioni reali, si può pensare di determinare il meno alto livello di significatività, α , che, se adottato, porterebbe a perdere l'accettazione. Nella letteratura in lingua inglese esso è detto “*p-value*”. Traducendolo come “*valore p*” di un test, esso può essere definito come il più piccolo livello di significatività, a partire dal quale verrebbe rigettata l'ipotesi nulla.

In questo senso, riprendendo l'intervallo di rigetto corrispondente ad un generico α :

$$\left| \frac{\overline{\overline{W}}(n, k) - w_0}{S / \sqrt{n}} \right| > t_{n-1; 1-\alpha/2},$$

ponendo per comodità

$$\frac{\overline{\overline{W}}(n, k) - w_0}{S / \sqrt{n}} \doteq t_{n-1; 1-p/2},$$

e ragionando sui corrispondenti valori della distribuzione di Student, $F_T(\cdot)$:

$$F_T(|t_{n-1; 1-p/2}|) > F_T(t_{n-1; 1-\alpha/2}) = 1-\alpha/2$$

si ha

$$F_T(|t_{n-1; 1-p/2}|) > 1-\alpha/2 \text{ e da qui: } \alpha > 2 \cdot (1 - F_T(|t_{n-1; 1-p/2}|)) \doteq p.$$

Ancora a proposito della possibilità di accettare/rigettare un'ipotesi nulla, si può fare un'osservazione sulla problematicità della dimensione (n) del campione: poiché la differenza di nostro interesse, $\overline{\overline{W}}(n, k) - w_0$, viene moltiplicata per \sqrt{n} , nel calcolo della statistica, allora per $n \rightarrow \infty$ si potrebbe finire per rigettare sempre l'ipotesi nulla.

Infine, è importante la seguente definizione, con la quale si vuole misurare la potenza di un test.

DEF: Il **potere** di un test d'ipotesi è la probabilità di rifiutare una ipotesi falsa, ossia:

$$1-\beta$$

Un caso particolare: normalità delle osservazioni e varianza nota.

La possibilità di avere a che fare con un campione di osservazioni reali estratte da una legge normale e, per di più, con varianza nota è considerata in questa sede come un caso particolare. Per completezza, dunque, sarà delineato il relativo test sulla media, cogliendo l'occasione di aggiungere qualche ulteriore dettaglio sulla teoria dei test.

Anzitutto è il caso di precisare che, quando ci si riferisce alla media di uno qualunque dei parametri o degli indici di prestazione di un modello d'interesse, diverso da quello produttore-consumatore, allora si è soliti scrivere così:

$$H_0: \mu = \mu_0 \quad (\text{ipotesi nulla})$$

$$H_1: \mu \neq \mu_0 \quad (\text{ipotesi alternativa})$$

Tale test è detto bilaterale e così lo si distingue dagli altri due possibili, che sono il test monolaterale_1 e il test monolaterale_2, rispettivamente formulati come segue:

$$H_0: \leq 0 \quad H_0: \geq 0$$

$$H_1: > 0 \quad H_1: < 0$$

Con le ipotesi fatte, la statistica del test sulla media diventa la normale standard:

$$\frac{\bar{X} - 0}{\sigma/\sqrt{n}} \approx Z,$$

con \bar{X} media campionaria e σ^2 quale varianza nota.

Da qui:

$$\Pr(-z_{1-\alpha/2} \leq \frac{\bar{X} - 0}{\sigma/\sqrt{n}} \leq z_{1-\alpha/2}) = 1 - \alpha$$

e perciò

$$\left| \frac{\bar{X} - 0}{\sigma/\sqrt{n}} \right| > z_{1-\alpha/2}$$

è la regione di rifiuto (o regione critica) e

$$\left| \frac{\bar{X} - 0}{\sigma/\sqrt{n}} \right| \leq z_{1-\alpha/2}$$

quella di accettazione.

Nel caso di test monolaterale_1, è immediato riconoscere che la regione d'accettazione diventa $(-\infty, z_{1-\alpha}]$ e quindi si rifiuta l'ipotesi se risulta:

$$\frac{\bar{X} - 0}{\sigma/\sqrt{n}} > z_{1-\alpha}.$$

Viceversa, per il test monolaterale_2, l'intervallo di accettazione è $[-z_{1-\alpha}, \infty)$ e l'ipotesi viene rifiutata se risulta:

$$\frac{\bar{X} - 0}{\sigma/\sqrt{n}} < -z_{1-\alpha}.$$

Test sulla differenza di due medie

Come già accennato, il problema di stabilire se c'è differenza (e di che segno) fra le medie (μ_x e μ_y) di due distribuzioni normali, è particolarmente significativo se si immagina che le due leggi normali possano essere quelle secondo le quali è distribuito l'indice di prestazione di un sistema che può essere gestito con due politiche diverse. Col test sulle medie (\bar{X} e \bar{Y}) di due campioni di osservazioni (x_1, x_2, \dots, x_{n_x} e y_1, y_2, \dots, y_{n_y}) reali, indipendenti e relative alle due diverse politiche si cercherà di stabilire se esse siano equivalenti, e, in caso negativo, quale politica sia preferibile.

Qui si comincerà trattando il caso particolare in cui le due rispettive varianze siano note.

Volendo stabilire se le medie sono uguali o diverse, le ipotesi saranno:

$$H_0: \mu_x = \mu_y$$

$$H_1: \mu_x \neq \mu_y$$

La statistica del test è la seguente:

$$Z_0 = \frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}}$$

Se l'ipotesi nulla è vera, allora la statistica è distribuita come una normale standard, per cui si rifiuta l'ipotesi se risulta: $|Z_0| > z_{1-\alpha/2}$.

Passando, per completezza, ai risultati per i due test monolaterali:

Per il test_1:

$$H_0: \mu_x \leq \mu_0$$

$$H_1: \mu_x > \mu_0$$

si rifiuta l'ipotesi nulla se risulta:

$$Z_0 = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}} > z_{1-\alpha}.$$

Per il test_2:

$$H_0 : \mu \geq \mu_0$$

$$H_1 : \mu < \mu_0$$

si rifiuta l'ipotesi nulla se risulta:

$$Z_0 = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}} < -z_{1-\alpha}$$

- **Il caso più generale, con varianze incognite**

Lavorando con varianze campionarie (S_\bullet^2) si può fare ricorso a statistiche di test (\tilde{T}) distribuite in accordo a leggi “ T_γ di Student” con opportuni gradi di libertà (γ) e possono essere distinti tre casi.

CASO 1: “pooled t-test”

Si suppone che le varianze delle due distribuzioni siano uguali, seppur incognite. In una situazione reale, questo caso può essere applicato qualora si possa sostenere l'ipotesi che due diverse politiche di gestione di un sistema non influiscano sulla varianza dell'indice di prestazione d'interesse.

Con $S_p^2 \triangleq \frac{(n_x - 1)S_x^2 + (n_y - 1)S_y^2}{n_x + n_y - 2}$, la statistica del test è: $\tilde{T} = \frac{\bar{X} - \bar{Y}}{S_p \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} \approx T_\gamma$

con $\gamma = n_x + n_y - 2$ gradi di libertà.

CASO 2: “approximate t-test”

Si suppone che le varianze delle due distribuzioni siano diverse.

In tal caso la statistica del test è:

$\tilde{T} = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_x^2}{n_x} + \frac{S_y^2}{n_y}}} \approx T_\gamma$, con $\gamma = \frac{\left(S_x^2/n_x + S_y^2/n_y \right)^2}{\left(S_x^2/n_x \right)^2/(n_x+1) + \left(S_y^2/n_y \right)^2/(n_y+1)} - 2$ gradi di libertà.

CASO 3: “paired t-test”

Questo caso ha senso quando le osservazioni con lo stesso indice sono dipendenti, ma si può assumere che le differenze $d_i \hat{=} x_i - y_i$, $i = 1, \dots, n$ siano realizzazioni di una sequenza (D_1, D_2, \dots, D_n) di variabili aleatorie indipendenti e identicamente distribuite, come una normale di media $D \hat{=} \bar{x} - \bar{y}$ che, quindi, dovrebbe risultare pari a zero sotto l’ipotesi nulla, $\bar{x} = \bar{y}$.

La statistica del test è: $\tilde{T} = \frac{\bar{D} - 0}{\sqrt{S_d^2 / n}} \approx T_\gamma$, con $\gamma = n - 1$ gradi di libertà.

Per tutti e tre i casi delineati, valgono le seguenti regole:

- test bilaterale, si rifiuta l’ipotesi nulla se risulta: $|\tilde{T}| > t_{\gamma; 1-\alpha/2}$.
- test monolaterale_1, invece: $\tilde{T} > t_{\gamma; 1-\alpha}$ e per il test monolaterale_2:
$$\tilde{T} < -t_{\gamma; 1-\alpha}.$$

Il test della chi-quadrato per la bontà della forma

A partire da un campione di osservazioni reali (x_1, x_2, \dots, x_n) indipendenti, nel continuo o nel discreto, con questo test si cerca di stabilire se esse possano essere considerate realizzazioni indipendenti di una variabile aleatoria che abbia una funzione densità nota, ovvero scelta fra quelle disponibili dall'analisi probabilistica. La densità ipotizzata sarà detta $\hat{f}(x)$ nel seguito, e non $f(x)$ come al solito, per evidenziare il fatto che la sua espressione analitica contiene parametri che devono essere stimati in via preliminare al test. Ad esempio, se si trattasse della ben nota densità esponenziale, $f(x) = \lambda \cdot \exp(-\lambda x)$, di parametro λ , allora occorrerebbe stimare questo parametro e poi, detta $\hat{\lambda}$ la stima di λ , si avrebbe la densità esponenziale ipotizzata: $f(x) = \hat{\lambda} \cdot \exp(-\hat{\lambda} x)$.

Sul piano formale, l'ipotesi nulla del test può essere espressa come segue:

$$H_0 : x_1, \dots, x_n \leftarrow \hat{f}(x) = f(x; \hat{\theta})$$

come per chiedersi: "può, la densità ipotizzata, essere quella che caratterizza la variabile aleatoria d'interesse e quindi produrre quel campione di realizzazioni?"

Il primo passo dell'esecuzione del test consiste nel raggruppare i valori x_1, x_2, \dots, x_n in un certo numero di intervalli (k) adiacenti. Si tenga presente sin da ora che scegliere questa numerosità è cosa non banale, anzi è il punto debole del metodo perché i risultati sono abbastanza sensibili a quella scelta. Comunque, per il momento, è il caso di andare avanti definendo il seguente:

N_j = numero di osservazioni reali, x_i , $i=1, \dots, n$ raggruppati nel j -esimo dei k intervalli

Il secondo passo consiste nel calcolare la proporzione p_j di valori x_i , $i=1, \dots, n$ che dovrebbero essere riscontrati nel j -esimo dei k intervalli, qualora la distribuzione ipotizzata fosse quella vera.

Nel continuo si ha:

$$p_j = \int_{a_{j-1}}^{a_j} \hat{f}(x) dx.$$

Nel discreto, invece:

$$p_j = \sum_{a_{j-1} \leq x_i < a_j} \hat{p}(x_i)$$

La statistica del test è la seguente:

$$\tilde{T}(n) \triangleq \sum_{j=1}^k \frac{(N_j - np_j)^2}{np_j} \equiv \sum_{j=1}^k \frac{(ValoriOsservati - valoriAttesi)^2}{valoriAttesi}$$

perché si può dimostrare che risulta:

$$\tilde{T}(n) \rightarrow \chi^2_{n-1} \quad per \quad n \rightarrow \infty$$

Per eseguire il test, occorre ipotizzare un livello di sicurezza α e quindi confrontare il valore assunto dalla statistica del test con il valore $\chi^2_{k-1,1-\alpha}$. Si rigetta l'ipotesi se $\chi^2 > \chi^2_{k-1,1-\alpha}$ (come illustrato nella figura in seguito).

La difficoltà che si incontra usando questo metodo non risiede tanto e solo nel poter disporre di un campione sufficientemente grande di osservazioni reali, quanto e più nella scelta soggettiva dell'ampiezza degli intervalli e quindi della loro numerosità. La prassi ricorrente, riportata come raccomandazione nella letteratura specializzata, è quella di rispettare le seguenti condizioni:

$$\begin{aligned} k &\geq 3 \\ n \cdot p_j &\geq 5 \quad per \quad ogni \quad j \end{aligned}$$

applicando il test anche con un numero di osservazioni limitate a qualche decina.

Quanto alla scelta degli estremi di ciascun intervallo, $(a_0, a_1), \dots, (a_{j-1}, a_j), \dots, (a_{n-1}, a_n)$, di solito viene seguito un approccio detto equiprobabile, ossia tale che le $p_j \quad j=1, \dots, n$ risultino tutte uguali. Se si fissano le probabilità p_j tali che $p_1 = p_2 = \dots = p_k = p = 1/k$, (con k intervalli), si ottiene $\hat{F}(a_j) = j/k$. Questo è giustificato dalla seguente uguaglianza:

$$\begin{aligned} p &= \int_{a_{j-1}}^{a_j} \hat{f}(x) dx = \hat{F}(a_j) - \hat{F}(a_{j-1}) \\ \Rightarrow 1/k &= j/k - (j-1)/k \end{aligned}$$

da cui si ricava

$$a_j = \hat{F}^{-1}(j/k).$$

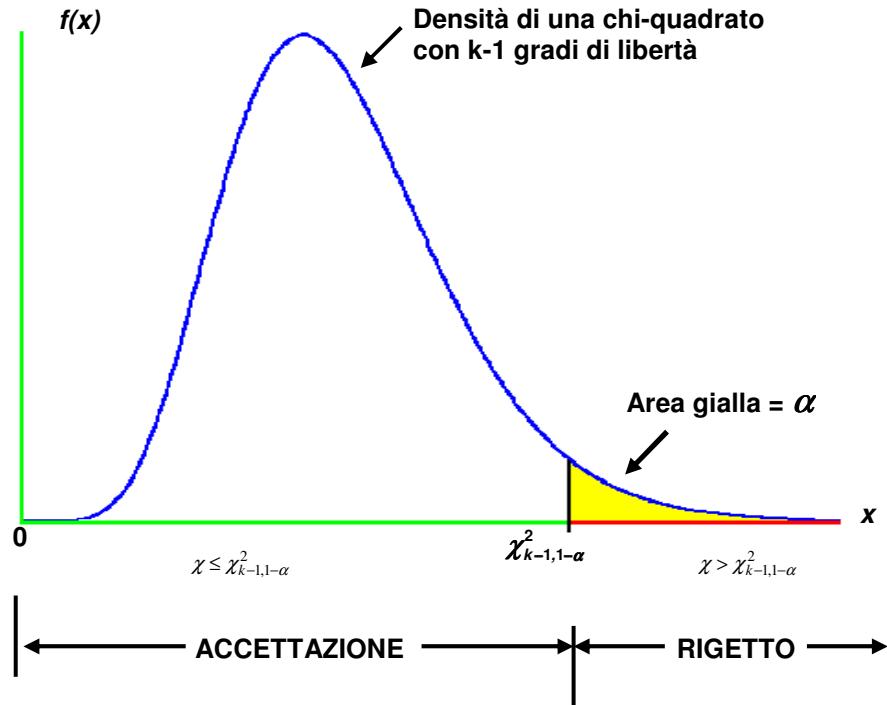
Una difficoltà potrebbe sorgere nel calcolare l'inversa della funzione \hat{F} .

A tal proposito, qui di seguito è descritto il procedimento per una \hat{F} normale.

Fissando la probabilità $p_j = 1/k \Rightarrow \hat{F}(a_j) = j/k$, allora, se la normale è standard il valore di a_j si legge direttamente dalla tabella in corrispondenza del valore di area pari a j/k . Viceversa, se la normale non è standard, si ha: $a_j = \bar{X}(n) + z_j S(n)$, dove z_j è il valore di ascissa per cui si ha la probabilità j/k per la distribuzione normale standard, mentre $S(n)$ e $\bar{X}(n)$ sono gli stimatori, rispettivamente, della deviazione standard e della media. Questo perché $X' \approx N(\mu, \sigma)$ e $X \approx N(0,1)$ sono legate dalla seguente relazione $X' = \bar{X} + \sigma X$. A questo punto si calcola N_j per $j = 1, \dots, n$ e si prosegue.

Alternativamente si possono fissare, anziché le probabilità uguali, gli estremi di integrazioni uguali, ossia si sceglie una certa ampiezza per i k intervalli in cui si è diviso il dominio della distribuzione, uguale per tutti.

Illustrazione del test (della) chi-quadrato



Valori critici della chi-quadrato					
gdl	$\alpha = 0,05$	$\alpha = 0,01$	gdl	$\alpha = 0,05$	$\alpha = 0,01$
1	3,84	6,63	16	26,30	32,00
2	5,99	9,21	17	27,59	33,41
3	7,81	11,34	18	28,87	34,81
4	9,49	13,28	19	30,14	36,19
5	11,07	15,09	20	31,41	37,57
6	12,59	16,81	21	32,67	38,93
7	14,07	18,48	22	33,92	40,29
8	15,51	20,09	23	35,17	41,64
9	16,92	21,67	24	36,42	42,98
10	18,31	23,21	25	37,65	44,31
11	19,68	24,72	26	38,89	45,64
12	21,03	26,22	27	40,11	46,96
13	22,36	27,69	28	41,34	48,28
14	23,68	29,14	29	42,56	49,59
15	25,00	30,58	∞	43,77	50,89

Esempio di esecuzione del test per una legge esponenziale

Facendo riferimento ad una legge esponenziale di parametro fissato, $\lambda = 0.5$, sono stati riprodotti col metodo Monte Carlo, su foglio excel, numerosi campioni di 100 realizzazioni indipendenti. Quindi si è voluto eseguire il test della chi-quadrato su ciascuno di quei campioni per verificare l'occorrenza o meno dell'errore di prima specie, ovvero l'aspetto (in termini di istogramma) dei campioni che portano il test ad accettare l'ipotesi nulla (che in questo caso è ovviamente vera!) e, di contro, l'aspetto di campioni che (pur capitando raramente) portano il test alla conclusione falsa della non appartenenza alle leggi esponenziali in questione, con rigetto dell'ipotesi nulla.

Essendo la legge esponenziale facilmente invertibile, è stato usato il metodo degli intervalli equiprobabili con $p_j = 0.1$, che porta a fissare $k (= p_j)$ in 10.

Allora, da:

$$a_j = F^{-1}(j/k) = -\frac{1}{\lambda} \ln(1 - j/k) = -2 \ln(1 - j/10), \quad j = 0, \dots, 9$$

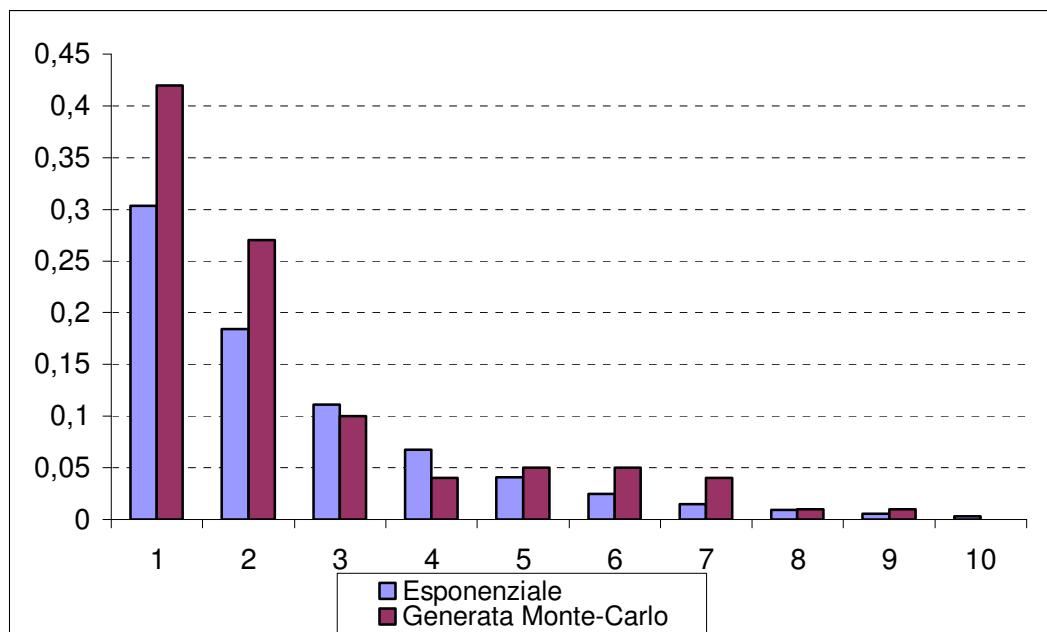
si ricavano i seguenti valori:

$a_0 = 0$	$a_5 = 1.3862$
$a_1 = 0.2107$	$a_6 = 1.8325$
$a_2 = 0.4462$	$a_7 = 2.4079$
$a_3 = 0.7133$	$a_8 = 3.2188$
$a_4 = 1.0216$	$a_9 = 4.6051$

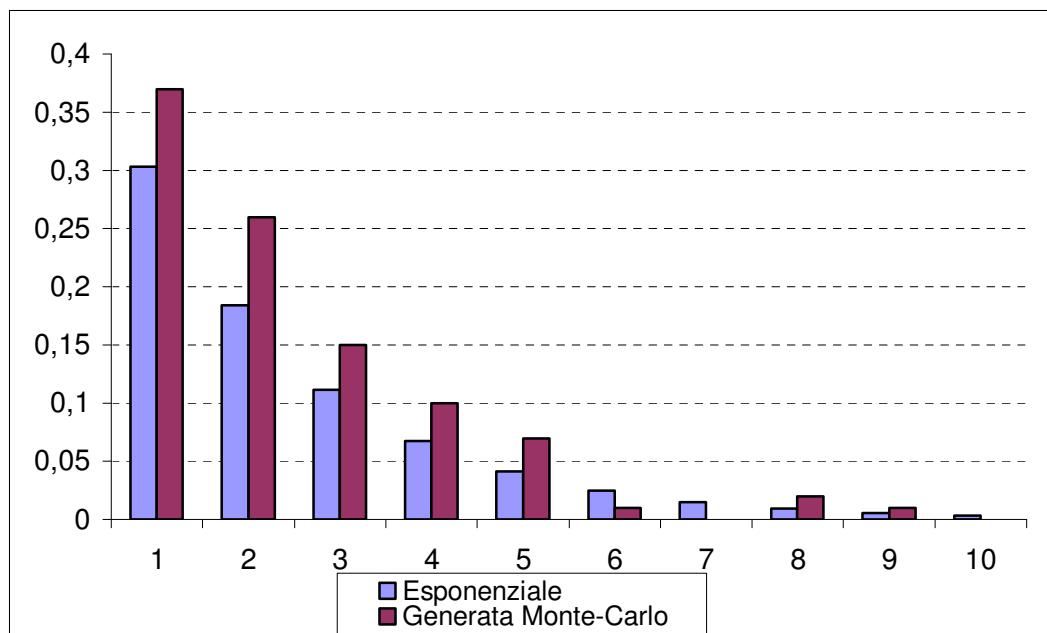
Il livello di significatività del test è stato fissato in $\alpha = 0.05$ e, con 10 intervalli, la statistica del test è una chi-quadrato con 9 gradi di libertà, che offre un valore critico pari a 16.92 ($= \chi^2_{0.95; 9}$).

I due grafici seguenti rappresentano due casi riferiti a due campioni che hanno portato ad esiti opposti del test. Nel primo è stato ottenuto un valore della statistica pari a 6.4 (<16.92) e quindi il campione è stato riconosciuto come generabile dalla legge ipotizzata, mentre nel secondo caso il valore della statistica è risultato pari a 22 (>16.92) e quindi si è verificato l'errore di prima specie.

Caso 1: accettazione dell'ipotesi nulla



Caso 2: rigetto dell'ipotesi nulla



Valore atteso condizionato e curva di regressione

Sia X una prima variabile aleatoria vista quale condizionante e Y una seconda, condizionata dalla prima. Ad esempio, in un modello produttore consumatore, potrebbe essere X = “tempo fra due consumi consecutivi” e Y = “durata della giacenza nel buffer”.

Riprendendo la formula del valore atteso condizionato:

$$E[Y | X = x] \hat{=} \int_{y=0}^{\infty} y \cdot f(y | x) \cdot dy$$

si osserva che essa definisce una funzione sullo spazio delle realizzazioni (continue e non negative) della X :

$$E[Y | X = x], \quad 0 \leq x < \infty.$$

Tale funzione, che descrive l'andamento del valore atteso della Y al variare della x , è detta curva di regressione e si potrebbe ricavare, in linea di principio, a partire dalla conoscenza della densità condizionata

$$f_{Y|X}(y | x)$$

o della congiunta

$$f_{Y,X}(y, x).$$

In pratica, è più spesso compito della statistica stimare i parametri della curva di regressione, a partire da osservazioni sperimentali della coppia (X, Y) .

La retta di regressione

La retta di regressione è quella particolare curva di regressione che si ottiene ponendo:

$$E[Y | X = x] \doteq a \cdot x + b, \quad 0 \leq x < \infty$$

ed è comunemente usata nell'analisi (statistica) della dipendenza della variabile aleatoria Y dalla X , dopo aver stimato i coefficienti reali a e b a partire dalle realizzazioni sperimentali della coppia X, Y .

Si può dimostrare che, indipendentemente dalla forma completa delle funzioni di distribuzione della X e della Y , i coefficienti a e b potrebbero essere ricavati dalle seguenti formule:

$$a = \rho \frac{\sqrt{VAR[Y]}}{\sqrt{VAR[X]}} = \frac{COV(X, Y)}{VAR[X]}; \quad b = E[Y] - \frac{COV(X, Y)}{VAR[X]} E[X];$$

dove ρ è il coefficiente di correlazione di Pearson.

Punto di partenza della dimostrazione è la formula del valore atteso condizionato:

$$\int_{y=0}^{\infty} y \cdot f(y | x) \cdot dy \doteq E[Y | X = x]$$

riscritta usando la densità congiunta e la densità marginale:

$$\frac{1}{f_X^{(m)}(x)} \int_{y=0}^{\infty} y \cdot f_{Y,X}(x, y) \cdot dy = E[Y | X = x]$$

Con essa si può scrivere:

$$\int_0^{\infty} y f_{Y,X}(x, y) dy = (ax + b) f_X^{(m)}(x)$$

e integrando in x ambo i membri

$$\int_{x=0}^{\infty} \int_0^{\infty} y f_{Y,X}(x, y) dy \cdot dx = \int_{x=0}^{\infty} (ax + b) f_X^{(m)}(x) \cdot dx$$

si ottiene:

$$E[Y] = a \cdot E[X] + b. \quad (\text{r1})$$

OSSERVAZIONE: la relazione di linearità tra i valori attesi di Y e X è cosa ben diversa dalla relazione di linearità tra le variabili aleatorie Y e X e non implicante quest'ultima!

Ora, riprendendo la: $E[Y | X = x] \doteq a \cdot x + b$ e integrando ambo i membri, dopo aver moltiplicato entrambi per “ $x \cdot f_X(x)$ ” risulta

$$\begin{aligned} & \int_0^\infty x \cdot f_X(x) \cdot E[Y | x] \cdot dx = \int_0^\infty x \cdot f_X(x) \cdot (ax + b) \cdot dx \\ &= \int_0^\infty x \cdot f_X(x) \cdot \left\{ \int_0^\infty y \cdot f_{Y|X}(y | x) \cdot dy \right\} \cdot dx = b \cdot E[X] + a \cdot E[X^2] \\ &= \int_0^\infty \int_0^\infty xy \cdot f_{X,Y}(x, y) \cdot dxdy = b \cdot E[X] + a \cdot E[X^2] \end{aligned}$$

ovvero:

$$E[X \cdot Y] = b \cdot E[X] + a \cdot E[X^2] \quad (\text{r2})$$

ma, per altra via:

$$\begin{aligned} E[X \cdot Y] &= COV[X, Y] + E[X]E[Y] \\ &= \rho \sqrt{VAR[X]} \sqrt{VAR[Y]} + E[X]E[Y] \end{aligned}$$

Allora, usando la (r2) e ricordando che $E[X^2] = Var[X] + E[X]^2$:

$$\rho \sqrt{VAR[X]} \sqrt{VAR[Y]} - E[X]E[Y] = b \cdot E[X] + a \cdot \{Var[X] + E[X]^2\}$$

A questo punto, invocando la (r1), si riconosce che $b = E[Y] - a \cdot E[X]$ e si ricava:

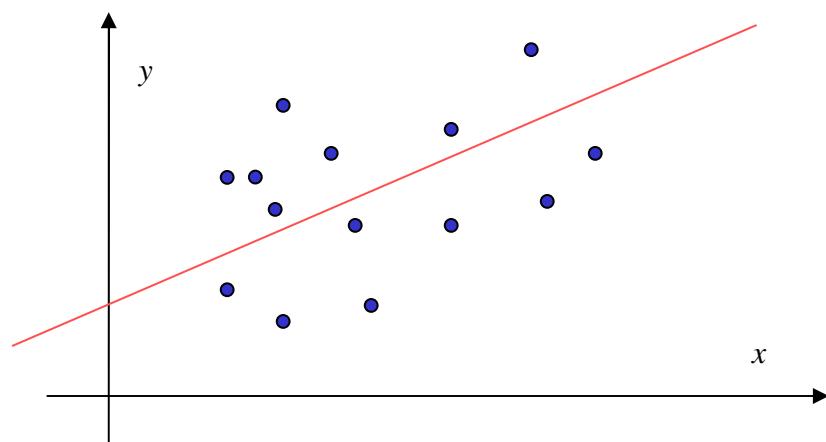
$$a = \rho \frac{\sqrt{VAR[Y]}}{\sqrt{VAR[X]}} \quad \text{e quindi} \quad b = E[Y] - \rho \frac{\sqrt{VAR[Y]}}{\sqrt{VAR[X]}} E[X]$$

In definitiva, è stata dimostrata la seguente:

$$E[Y | X = x] \doteq E[Y] + \rho \frac{\sqrt{VAR[Y]}}{\sqrt{VAR[X]}} (x - E[X]), \quad 0 \leq x < \infty$$

Stima dei parametri della retta di regressione

Qui si farà vedere come possano essere determinati i parametri (a e b) della retta di regressione, a partire da un insieme di coppie $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ di realizzazioni congiunte delle variabili aleatorie X e Y . L'idea è che la retta di regressione debba essere proprio quella che, nel piano euclideo, passa “il più possibile vicino” ai punti corrispondenti alle coppie di realizzazioni $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$. La figura seguente illustra l'idea:



Per formalizzare il concetto di “il più possibile vicino”, si conviene di cercare i parametri a e b tali che risulti minima la somma dei quadrati degli scostamenti dei valori y_1, y_2, \dots, y_n osservati rispetto ai valori sulla retta stessa:

$$\sum_{i=1}^n [y_i - (ax_i + b)]^2.$$

Considerando, dunque, la precedente somma (S) come funzione dei due parametri a e b , si deriva prima rispetto all'uno e poi rispetto all'altro:

$$\begin{cases} \frac{\partial S(a, b)}{\partial b} = -2 \cdot \sum_{i=1}^n (y_i - b - ax_i) = 0 \\ \frac{\partial S(a, b)}{\partial a} = -2 \cdot \sum_{i=1}^n (y_i - b - ax_i) \cdot x_i = 0 \end{cases}$$

da cui:

$$\begin{cases} n \cdot b + a \cdot \sum_{i=1}^n x_i - \sum_{i=1}^n y_i = 0 \\ b \cdot \sum_{i=1}^n x_i + a \cdot \sum_{i=1}^n x_i^2 - \sum_{i=1}^n x_i y_i = 0 \end{cases}.$$

Risolvendo rispetto ad a e b si ricavano le stime ai minimi quadrati (denotate come “ \hat{a} ” e “ \hat{b} ”:

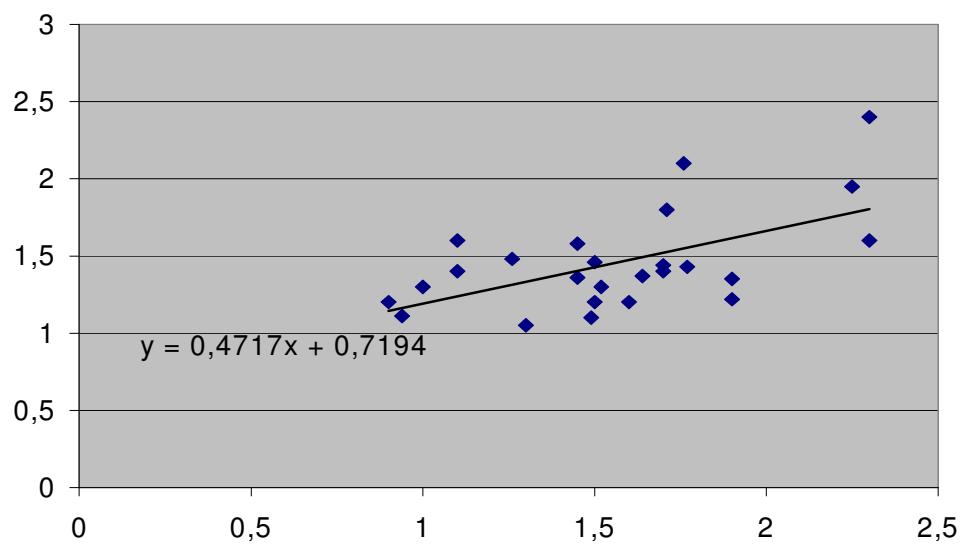
$$\hat{a} = \frac{n \cdot \sum_{i=1}^n x_i \cdot y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \cdot \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2} \quad \text{e} \quad \hat{b} = \frac{\sum_{i=1}^n y_i}{n} - \hat{a} \cdot \frac{\sum_{i=1}^n x_i}{n}, \quad (\text{r3})$$

che risultano in accordo con le formule già ricavate:

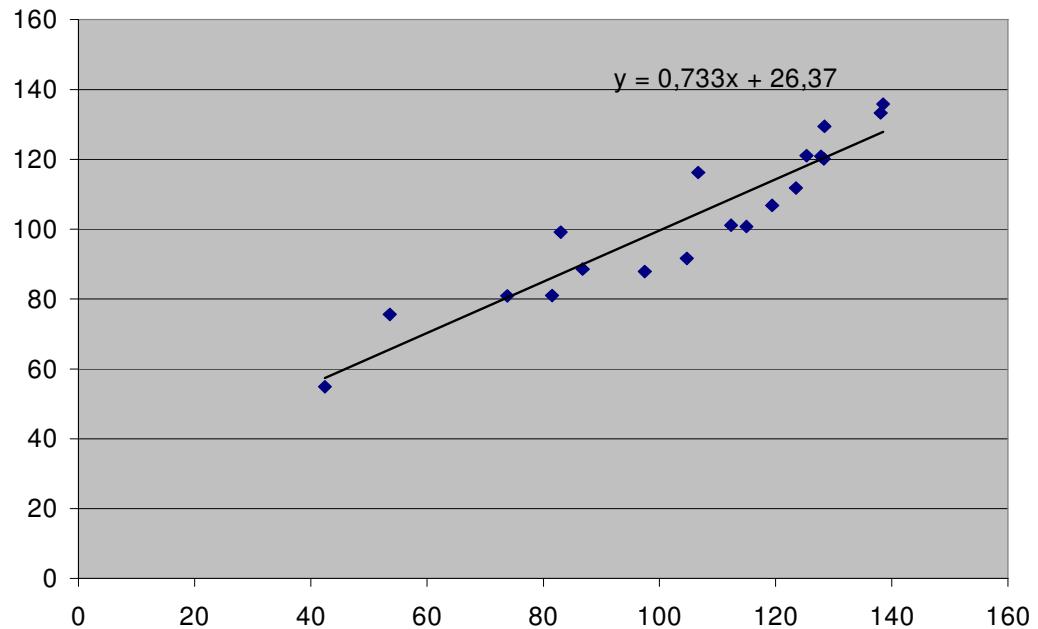
$$a = \frac{COV(X, Y)}{VAR[X]} \quad \text{e} \quad b = E[Y] - a \cdot E[X].$$

Esempi di rette di regressione

Due esempi numerici di costruzione di rette di regressione sono illustrati nelle prossime due figure e sono tratti da un caso reale di analisi della correlazione presso un terminale marittimo per container, dove erano state individuate due sottoaree separate di stoccaggio dei container sul piazzale che potevano ospitare gruppi di container che arrivavano con la stessa nave e con una seconda nave, ancora comune, erano destinati a ripartire. In tal caso, sia i tempi di giacenza nelle due sottoaree sia i rispettivi livelli di occupazione dovevano risultare dipendenti, come confermato dalle rette di regressione ricavate con i dati della pagina seguente.



Retta di regressione per i tempi di giacenza (settimane)



Retta di regressione per i livelli di occupazione (unità)

Dati per la costruzione delle rette di regressione

Rilevazioni congiunte

Tempi di giacenza di singoli container

(settimane)

$$y = 0,4717x + 0,7194$$

Livelli di occupazione delle due aree

(unità)

$$y = 0,733x + 26,37$$

0,9	1,2
1,9	1,35
1,5	1,46
1,45	1,36
1,1	1,6
1,6	1,2
1,45	1,58
0,94	1,11
1,3	1,05
1,1	1,4
1,9	1,22
1,26	1,48
2,3	1,6
1	1,3
1,52	1,3
1,7	1,44
1,5	1,2
1,49	1,1
2,3	2,4
1,7	1,4
1,71	1,8
1,64	1,37
1,76	2,1
2,25	1,95
1,77	1,43

42,4	54,9
73,8	80,9
83	99,2
106,7	116,2
128,4	129,5
138,5	135,8
138,1	133,3
125,3	121,1
128,3	120,1
127,8	120,9
123,5	111,8
119,4	106,8
115	100,8
112,3	101,1
104,7	91,6
86,8	88,6
97,5	87,9
81,5	81
53,6	75,6

Perfetta regressione lineare

Con riferimento al campione di realizzazioni congiunte $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ delle variabili aleatorie X e Y , si può definire perfetta quella retta di regressione con parametri a e b per i quali risulta $y_i = a \cdot x_i + b$, $i = 1, 2, \dots, n$. Ovviamente si tratta di un caso limite, perché è assai raro che gli n punti sperimentali della Y giacciono tutti su un'unica retta.

Più in generale, indicando con δ_i , $i = 1, \dots, n$ le deviazioni dei valori osservati y_i , $i = 1, 2, \dots, n$ rispetto ai corrispondenti valori sulla retta di regressione e ponendo $\delta_i \hat{=} |y_i - (a \cdot x_i + b)|$, $i = 1, 2, \dots, n$, è opportuno interpretare quelle deviazioni come realizzazioni (indipendenti) di una variabile aleatoria Δ . Così si può riconoscere che il criterio dei minimi quadrati per la determinazione della retta di regressione corrisponde a minimizzare la stima del momento del secondo ordine della Δ , $E^2[\Delta]$.

Infatti:

$$\hat{E}^2[\Delta] = \left(\sum_{i=1}^n \delta_i^2 \right) / n$$

Più precisamente, ricordando che:

- se Y e X sono legate da una perfetta relazione lineare allora risulta (r1, a pag.99) che $E[Y] = a \cdot E[X] + b$ e quindi $E[\Delta] = E[Y] - (a \cdot E[X] + b) = 0$,
- $VAR[\Delta] = E^2[\Delta] - (E[\Delta])^2$

si deduce, infine, che il criterio dei minimi quadrati corrisponde a determinare la retta cui corrisponde una variabile aleatoria Δ di media nulla e di varianza minima.

Questa retta è individuata dalle stime, \hat{a} e \hat{b} , (r3 a pag. 102), dei seguenti parametri:

$$a = \rho \frac{\sqrt{VAR[Y]}}{\sqrt{VAR[X]}} \quad (+) \quad b = E[Y] - \rho \frac{\sqrt{VAR[Y]}}{\sqrt{VAR[X]}} E[X]$$

Per capire come si possa stabilire l'esistenza di una perfetta regressione lineare a partire dalla stima ($\hat{\rho}$) del coefficiente di correlazione e, in ultima analisi, dal campione di realizzazioni congiunte $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$, è determinante la seguente implicazione:

$$|\rho| = 1 \Rightarrow Y = aX + b$$

Prova

Da $\Delta \hat{=} Y - (aX + b)$ risulta:

$$VAR[\Delta] = VAR[Y] + a^2 VAR[X] - 2a \cdot COV[X, Y]$$

e usando la (+):

$$VAR[\Delta] = VAR[Y] + \rho^2 VAR[Y] - 2\rho^2 \cdot VAR[Y]$$

ovvero:

$$VAR[\Delta] = VAR[Y] \cdot (1 - \rho^2). \quad (\$)$$

Da qui:

$$|\rho| = 1 \Rightarrow \rho^2 = 1 \Rightarrow VAR[\Delta] = 0$$

e, quindi, si avrà:

$$y_i = a \cdot x_i + b, \quad i = 1, 2, \dots, n,$$

ovvero tutte le realizzazioni della Y saranno proprio sulla retta.

OSSERVAZIONE CONCLUSIVA:

la relazione (\$) è interessante e meriterebbe di essere approfondita, perché essa mette in luce che il termine $(1-\rho^2)$ collega la varianza degli scarti alla varianza della Y.