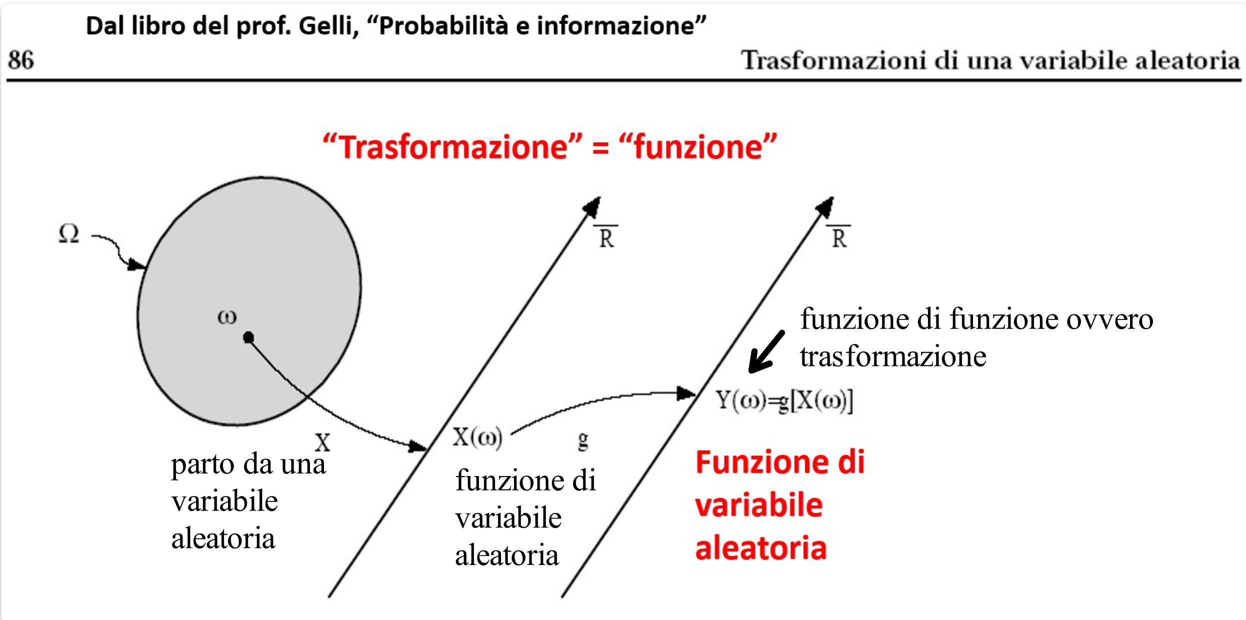
Funzioni di variabili aleatorie

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



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

S. Miller and D. Childers, “Probability and Random Processes:…”, Academic Press 2012

4.6 Transformations of Random Variables

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

probabilità che Y sia minore o uguale ad un valore y

probabilità che X sia minore o uguale all'inverso della trasformazione applicata a y

Note that this can also be written as

differenziabile ?

Si ma non detto esplicitamente

$$F_X(x) = F_Y(g(x)). \tag{4.21}$$

"Si cambia la forma ma non la sostanza"

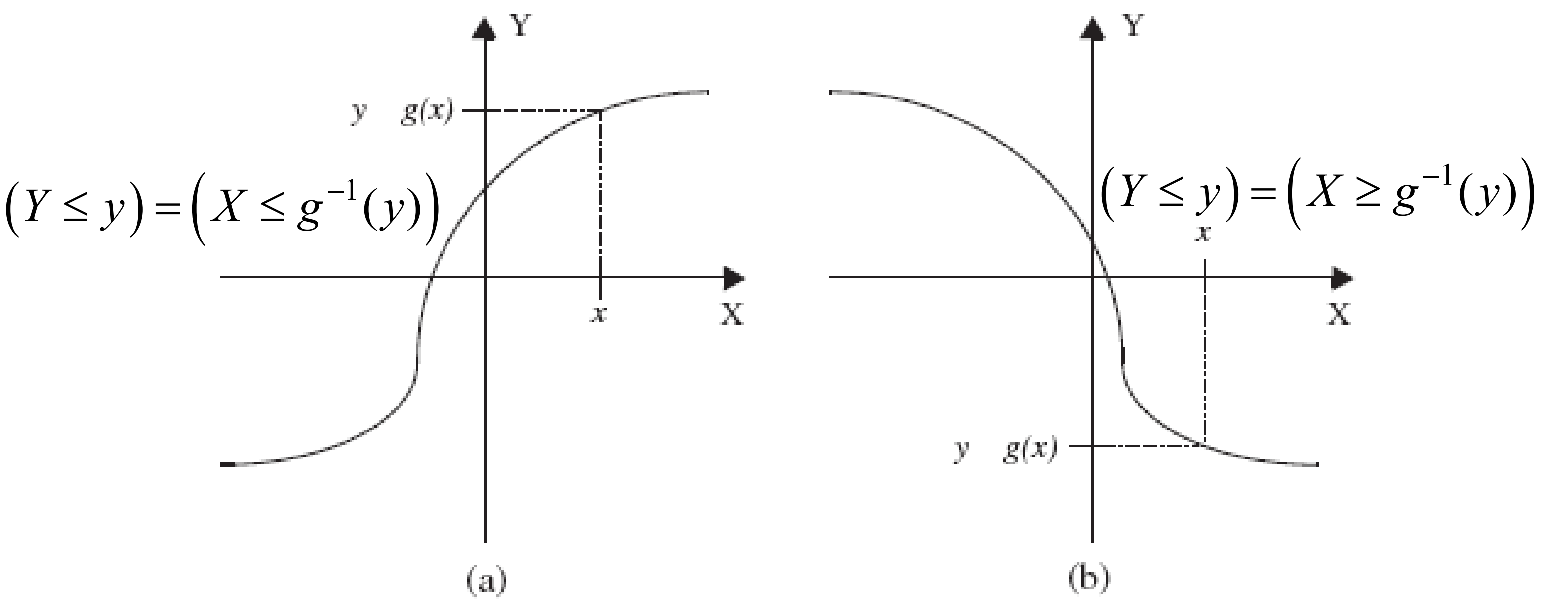
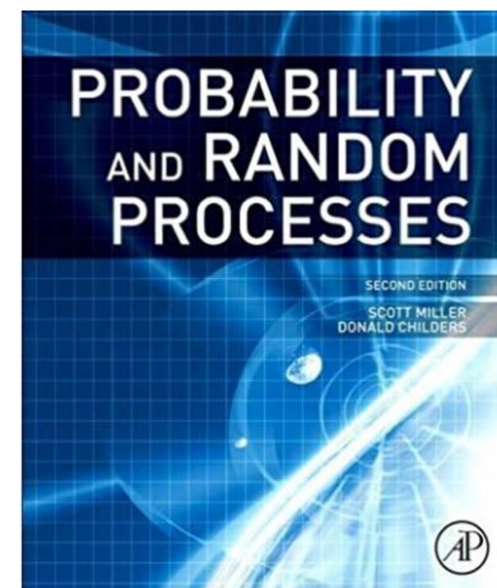


Figure 4.1 A monotonic increasing function (a) and a monotonic decreasing function (b).



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

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

densità di probabilità della variabile trasformata Y

$$f_Y(y) = f_X(g^{-1}(y))\frac{dg^{-1}(y)}{dy} = f_X(x)\frac{dx}{dy}\bigg|_{x=g^{-1}(y)}, \tag{4.22}$$

while differentiating Equation 4.21 with respect to x gives

Da questa slide si capisce che la funzione è differenziabile!

$$f_X(x) = f_Y(g(x))\frac{dy}{dx} \Rightarrow f_Y(y) = \frac{f_X(x)}{\frac{dy}{dx}}\Bigg|_{x=g^{-1}(y)} \tag{4.23}$$

come ottenere la densità di Y a partire dalla densità di X

Per funzioni monotone decrescenti si parte dalla distribuzione

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

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

usa la complementare

$$F_Y(y) = \Pr(Y \le y) = \Pr(X \ge g^{-1}(y)) = 1 - F_X(g^{-1}(y)). \tag{4.24}$$

Differentiating with respect to y gives

$$f_Y(y) = -f_X(x)\frac{dx}{dy}\bigg|_{x=g^{-1}(y)} \tag{4.25}$$

segno negativo perché la funzione è decrescente!

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

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

Dal libro del prof. Gelli, “Probabilità e informazione”

▶ *Esempio* 4.7. Consideriamo nuovamente la trasformazione lineare

$$Y = aX + b\,,$$

ricordando che X è la variabile originale e Y 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

$$x = \frac{y-b}{a}\,,$$

ed inoltre risulta

ricavata tramite differenziabilità

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

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:

sostituisco ed isolo X

$$F_Y(y) = P(Y \leq y) = P(aX + b \leq y) = P\left(X \leq \frac{y-b}{a}\right) = F_X\left(\frac{y-b}{a}\right).$$

usando la 4.22 per funzioni monotone crescenti

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

$$F_Y(y) = P(Y \leq y) = P(aX + b \leq y) = P\left(X \geq \frac{y-b}{a}\right)$$

$$= 1 - P\left(X < \frac{y-b}{a}\right) = 1 - F_X\left(\frac{y-b}{a}\right).$$

la probabilità che Y sia <= 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

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

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 $E[Y] \hat{=} \int_{y=0}^{\infty} y \cdot f_Y(y)dy, \quad$ con $Y \hat{=} a \cdot X, \quad a > 0$ (esempio)

quindi b=0

$$\boxed{E[aX] \hat{=} \int_{ax=0}^{\infty} ax \cdot f(ax)d(ax) = \int_{x:\,ax=0}^{\infty} ax \cdot (1/a) f(y/a) \cdot d(ax) = \int_{x:\,ax=0}^{\infty} ax \cdot (1/a) f(ax/a) \cdot a dx}$$

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

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

$$\boxed{= a \cdot E[X]}$$

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

Generalizzazione, con g(x) "idonea":

per idonea si intende monotona non decrescente e differenziabile

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

Probability and Statistics with Reliability, Queuing and Computer Science Applications:

Second edition

by K.S. Trivedi

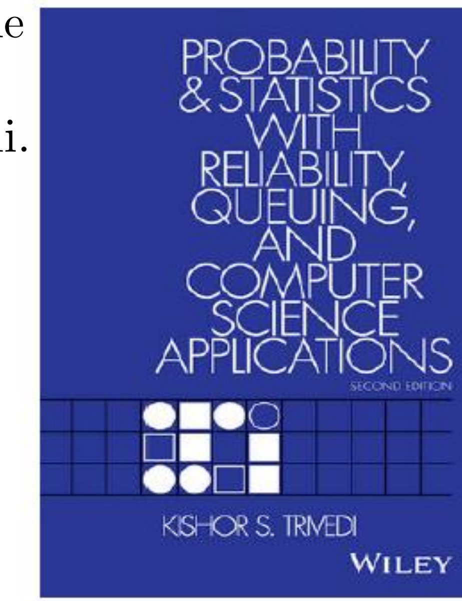
Publisher-John Wiley & Sons

https://trivedi.pratt.duke.edu/people/current-members

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

Theorem: pdf for a transformed RV

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



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

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

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

Prova:

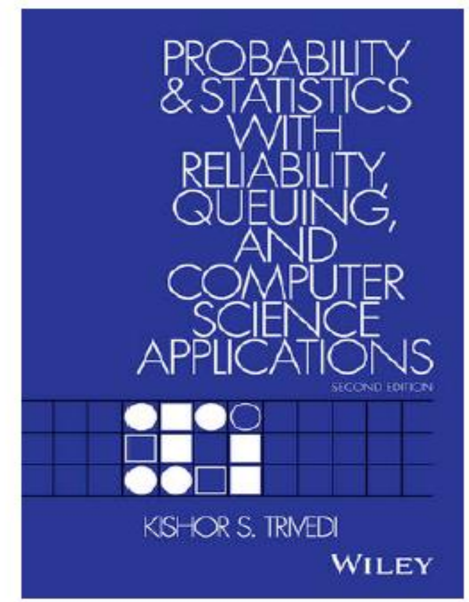
Derivando ed utilizzando la regola di derivazione delle funzioni composte si ottiene che

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

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 $F_Y(y) = P(Y \le y)$

$F_Y(y) = 0$, per $y \le 0$ $P(Y \le y)$ è zero!

Essendo $Y = X^2 \implies X^2 \le y$ ovvero $X^2 \le$ di una quantità <= di zero si risolve appunto solo per X=0

invece nel caso $y > 0$:

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

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

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

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

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

- A partire dal secondo integrale di Eulero si definisce la funzione di densità Gamma

portando fuori $\Gamma(\alpha)$ perchè costante ottengo $\frac{1}{\Gamma(\alpha)} \cdot \Gamma(\alpha) = 1$

$\hat{f}(t)$

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

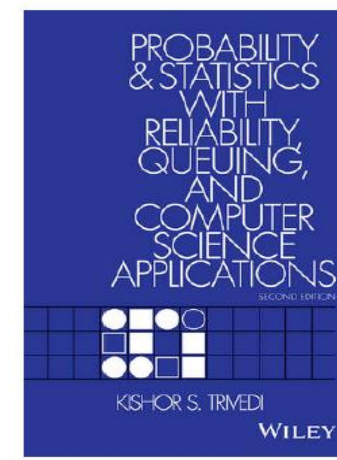
$$\int_0^{\infty} \frac{\hat{f}(t)}{\Gamma(\alpha)} dt = 1$$

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

dove

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

- Show the recurrence for the gamma function: $\Gamma(\alpha) = (\alpha\text{-}1)\ \Gamma(\alpha\text{-}1)$; and show that $\Gamma(1/2)=\sqrt{\pi}$
- Because $\Gamma(1)=1$, it follows that for an integer *r*, $\Gamma(r)=(r-1)\ \Gamma(r-1)=\ldots=(r-1)!$
- 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

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!

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

Example 3.9

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

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

- Using result from Example 3.8:

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

or,

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

Passaggi algebrici

$$f(y) = \frac{1}{\Gamma(\frac{1}{2})} \cdot \frac{1}{2} \cdot (\frac{y}{2})^{-\frac{1}{2}} \cdot e^{-\frac{y}{2}}$$

$$= \frac{1}{\Gamma(\frac{1}{2})} \cdot \frac{1}{\sqrt{2} \cdot \sqrt{2}} \cdot \frac{\sqrt{2}}{\sqrt{y}} \cdot e^{-\frac{y}{2}}$$

$$= \frac{1}{\sqrt{\pi}} \cdot \frac{1}{\sqrt{2}} \cdot \frac{1}{\sqrt{y}} \cdot e^{-\frac{y}{2}}$$

$$\Gamma(\frac{1}{2}) = \sqrt{\pi}$$

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

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

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)

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

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

Example 3.10

Tesi: arriviamo al modello esponenziale tramite funzione di funzione di X, variabile aleatoria uniforme in (0,1)

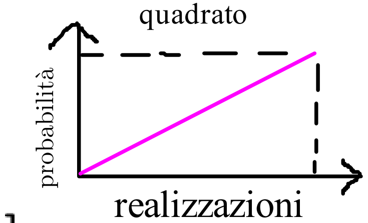
- Let X be uniformly distributed, *Unif(0,1)*
- Then, $Y = -\lambda^{-1} \ln(1-X)$ is $EXP(\lambda)$.

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

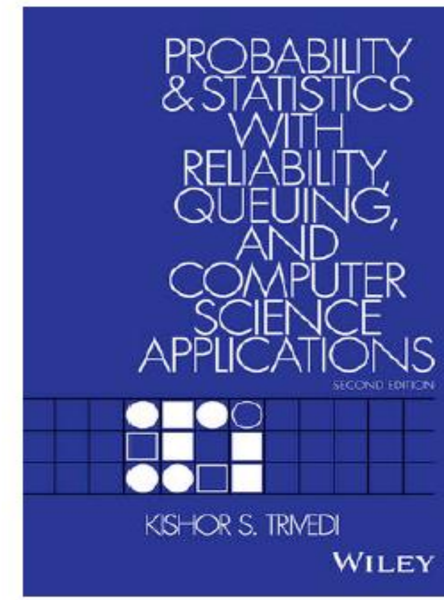
for $y > 0$,

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

Sarebbe un quadrato



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



Since X is $U(0,1)$, $F_X(x) = x, 0 \le x \le 1$. Therefore,

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

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

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

φ ē densità Gauss

Φ ē distrib. Gauss

log normale monotone gaussiana → log normale

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 ∞.

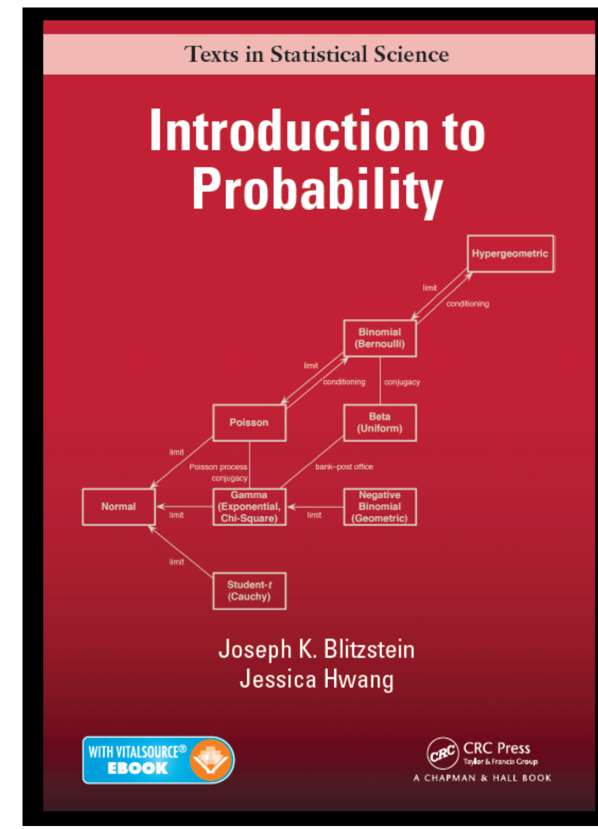
We can get the same result by working from the definition of the CDF, translating the event $Y \leq y$ into an equivalent event involving X. For $y > 0$,

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

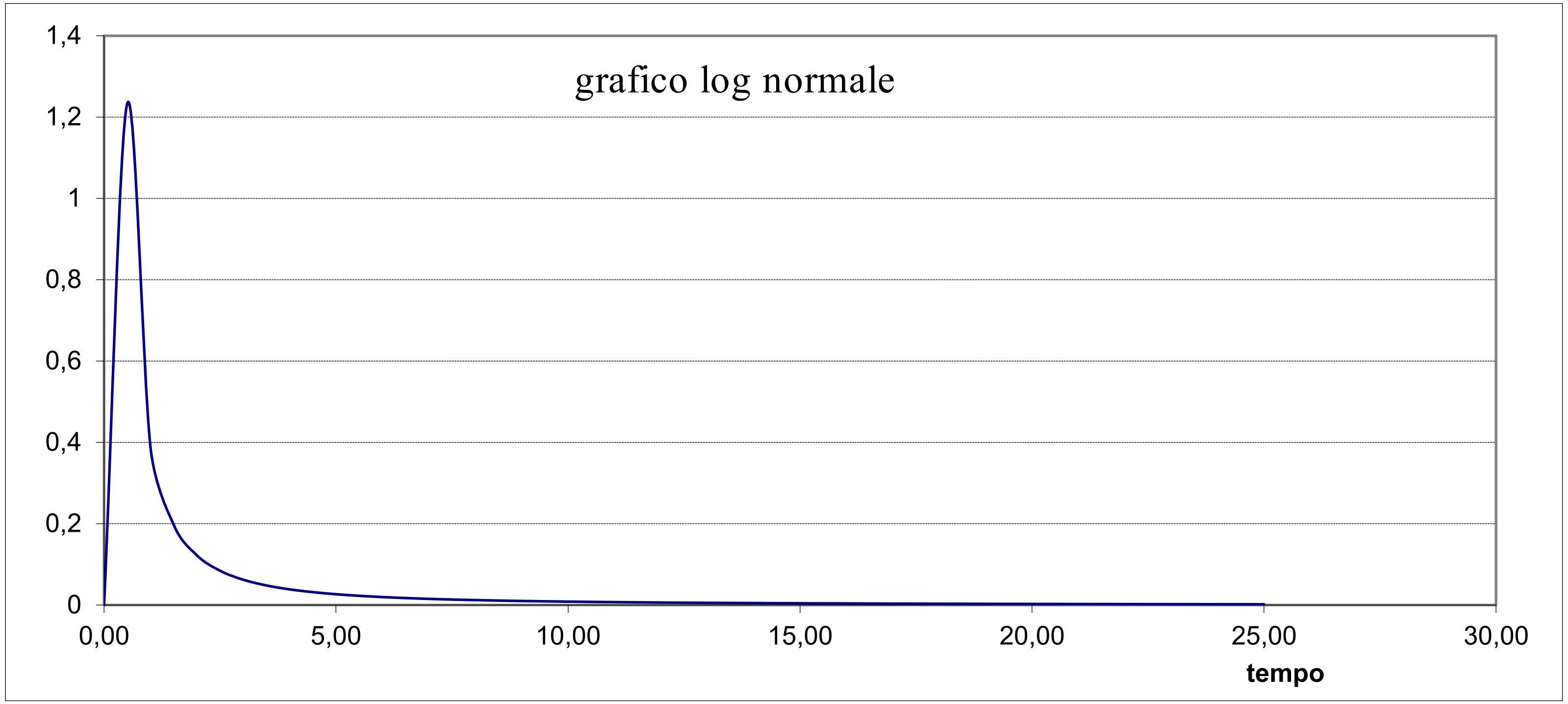
distribuzione Gaussiana

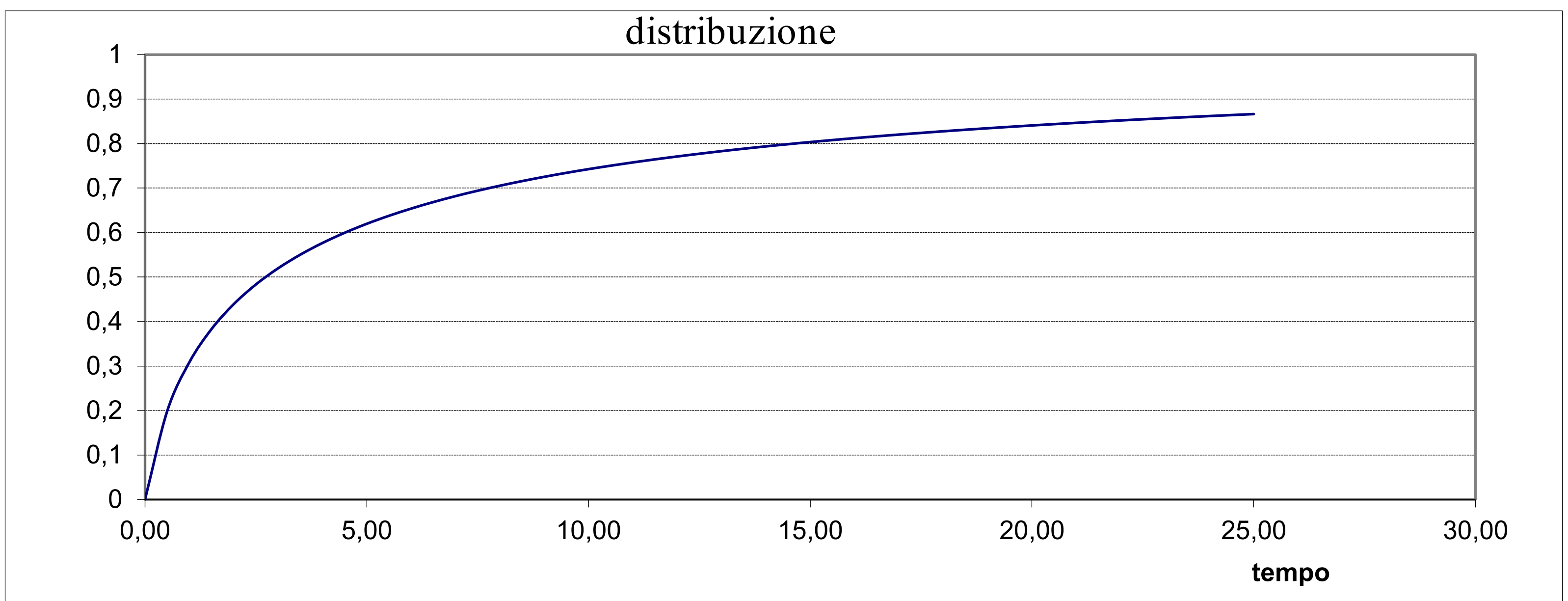
so the PDF is again

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



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

$\Phi \triangleq$ GAUSS DISTRIBUTION

inverte la x con y

X=lognorm Y=norm

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

dimostrazione (non richiesta)

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	$\frac{1}{2}+\frac{1}{2}\,\text{erf}\left[\frac{\ln x-\mu}{\sqrt{2}\sigma}\right]$
Quantile	$\exp(\mu+\sqrt{2\sigma^2}\,\text{erf}^{-1}(2F-1))$
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	$\exp(4\sigma^2)+2\exp(3\sigma^2)+3\exp(2\sigma^2)-6$

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

normale standardizzata

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

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

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

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

EXTRA

Appendice I

La funzione di distribuzione

Log-Logistic

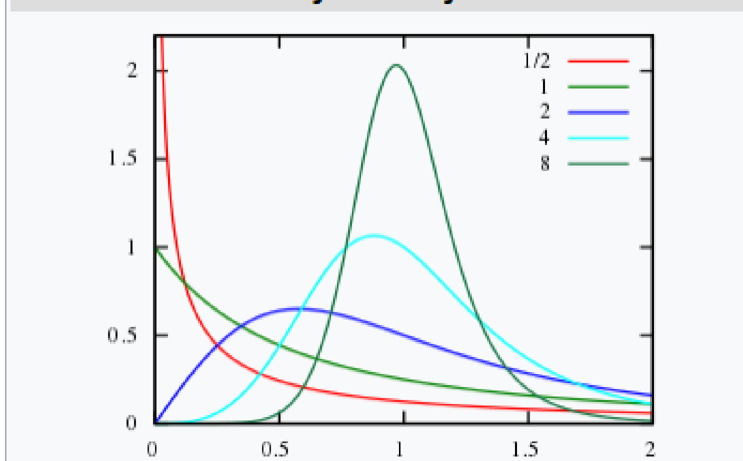
Log-logistic distribution

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.

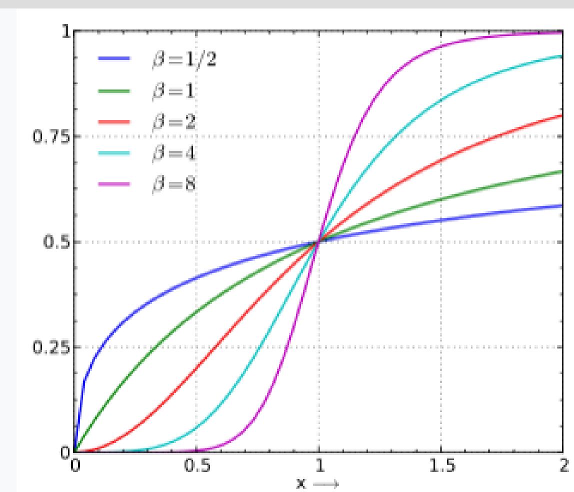
The log-logistic distribution is the probability distribution of a random variable whose logarithm has a logistic distribution. It is similar in shape to the log-normal distribution but has heavier tails. Unlike the log-normal, its cumulative distribution function can be written in closed form.

Probability density function



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

Cumulative distribution function



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

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

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

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

The probability density function is

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

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 usato 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 $F(x_p) = p$, $p \in (0,1)$ allora x_p è detto p-simo quantile della F

$p \triangleq 0.5$ è detto mediana

Moments [edit]

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

$$\begin{aligned} \operatorname{E}(X^k) &= \alpha^k \operatorname{B}(1 - k/\beta, 1 + k/\beta) \\ &= \alpha^k \frac{k\pi/\beta}{\sin(k\pi/\beta)} \end{aligned}$$

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

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

and the variance is

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

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]

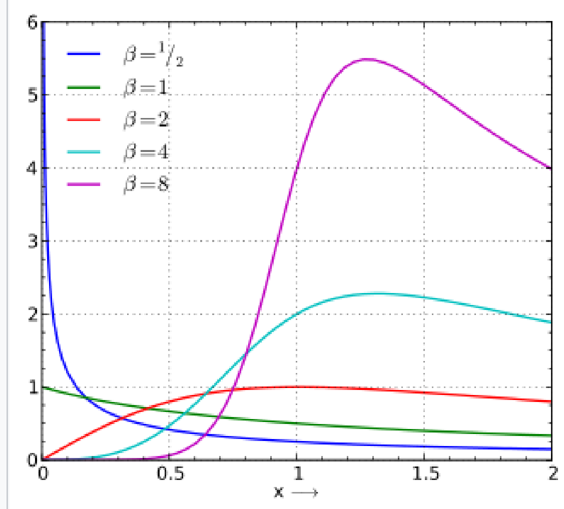
The survival function is

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

and so the hazard function is

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

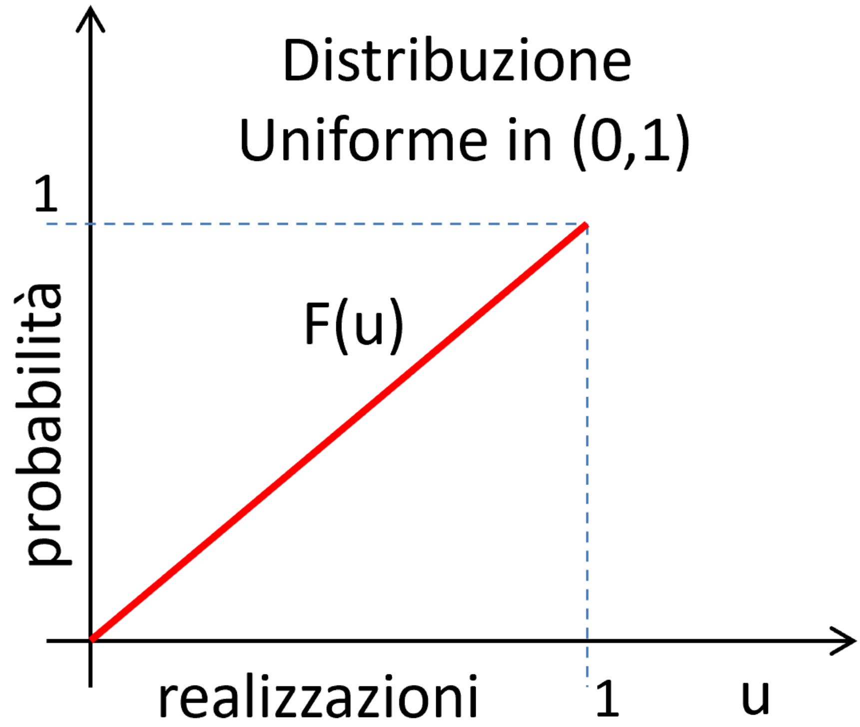
Anche questo grafico potete riprodurre con Excel !



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

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.

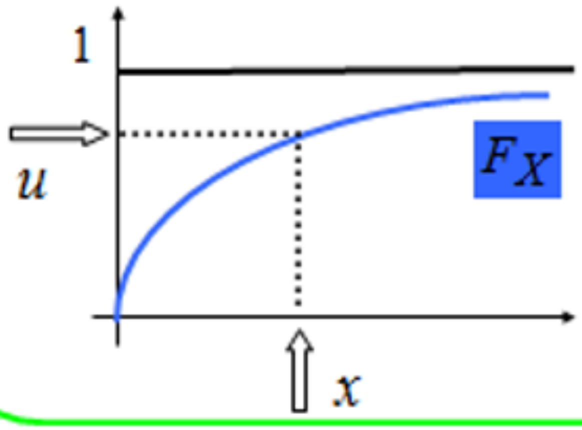
$\{ u_1, u_2, \ldots, u_n, \ldots \}$ realizzazioni indipendenti di $U\{0,1\}$

Generazione di realizzazioni delle Var. Al. d'ingresso

Metodo della trasformazione inversa

dimostrazione

$$\Pr\{X \le x\} \triangleq F_X(x) = F_X\left(F_X^{-1}(u)\right) = u = \Pr\{U \le u\}$$



$\Rightarrow \; x \triangleq F_X^{-1}(u)$

è realizzazione della X

A parole: presa una funzione di distribuzione invertibile $F_x(x)$ allora si ha che $F_X(F_X^{-1}(u)) = u$ dunque la funzione di distribuzione nell'inversa della distribuzione in u ritorna u. Esempio esponenziale con $\lambda = 1 \implies F_X(x) = 1 - e^{-x} \implies F_X^{-1}(x) = -ln(1-x)$

$F_X(F_X^{-1}(u)) = 1 - e^{-ln(1+u)} = 1 - (1-u) = u$ 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.

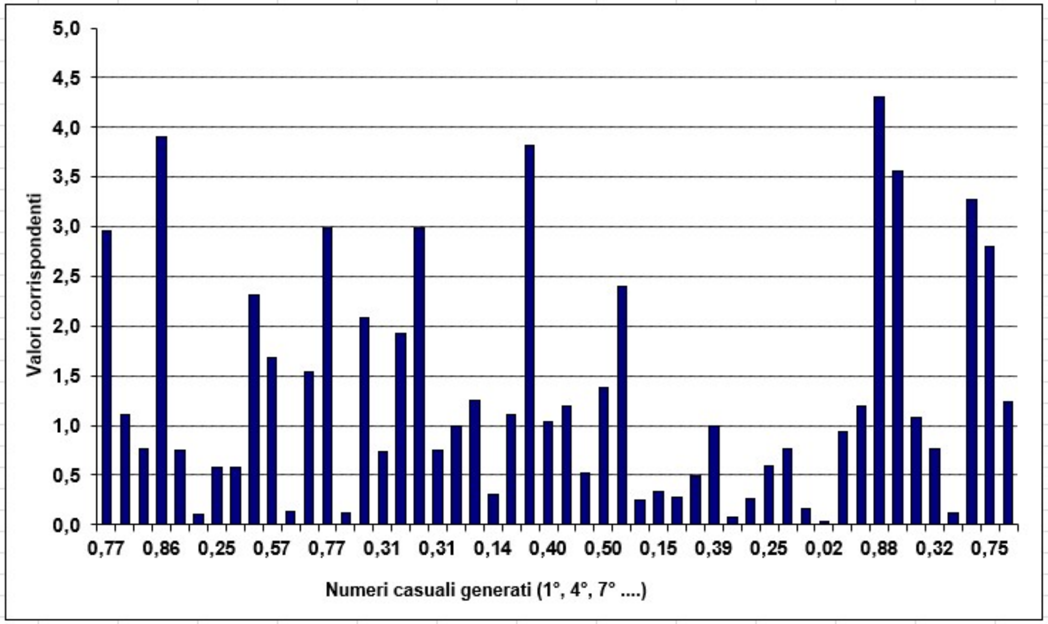
Ci mettiamo nell'ipotesi di modelli in cui la distribuzione è invertibile.

Il metodo non è utilizzabile per tutte le distribuzioni

lambda =	Media teorica	Media sperim.	Valori osservati min =	max =
0,5	2	1,3085	0,03402	4,29777

numeri casuali	valori realizzazioni
0,77128	2,950485
0,42497	1,10665
0,31886	0,767974
0,85769	3,899555
0,31025	0,742863
0,04833	0,099068
0,25017	0,575813
0,25206	0,580865
0,68434	2,30619
0,56897	1,683141
0,06673	0,138115
0,53733	1,541464
0,77464	2,98011
0,05832	0,120188
0,64648	2,079628
0,30505	0,72782
0,61717	1,920347
0,77447	2,978608
0,31108	0,74526
0,38972	0,987673
0,46311	1,243916
0,1386	0,298384
0,42707	1,113982
0,85168	3,816799
0,40272	1,030725
0,45122	1,200117
0,23151	0,526656
0,49758	1,37664
0,69791	2,394044
0,11341	0,240754
0,15382	0,334036
0,1283	0,27463
0,21925	0,494997
0,39157	0,993732
0,03818	0,077854
0,12015	0,255997
0,25392	0,585849
0,31854	0,76705
0,08024	0,167278
0,01686	0,034016

Realizzazioni casuali dalla distribuzione esponenziale di media = 2



n° realizz	n° classi	ampiezza classi (variabile)
50	10	0,42638
	fisso!	

Valore limite superiore di ogni classe

$\cong f_X(x)\Delta x$

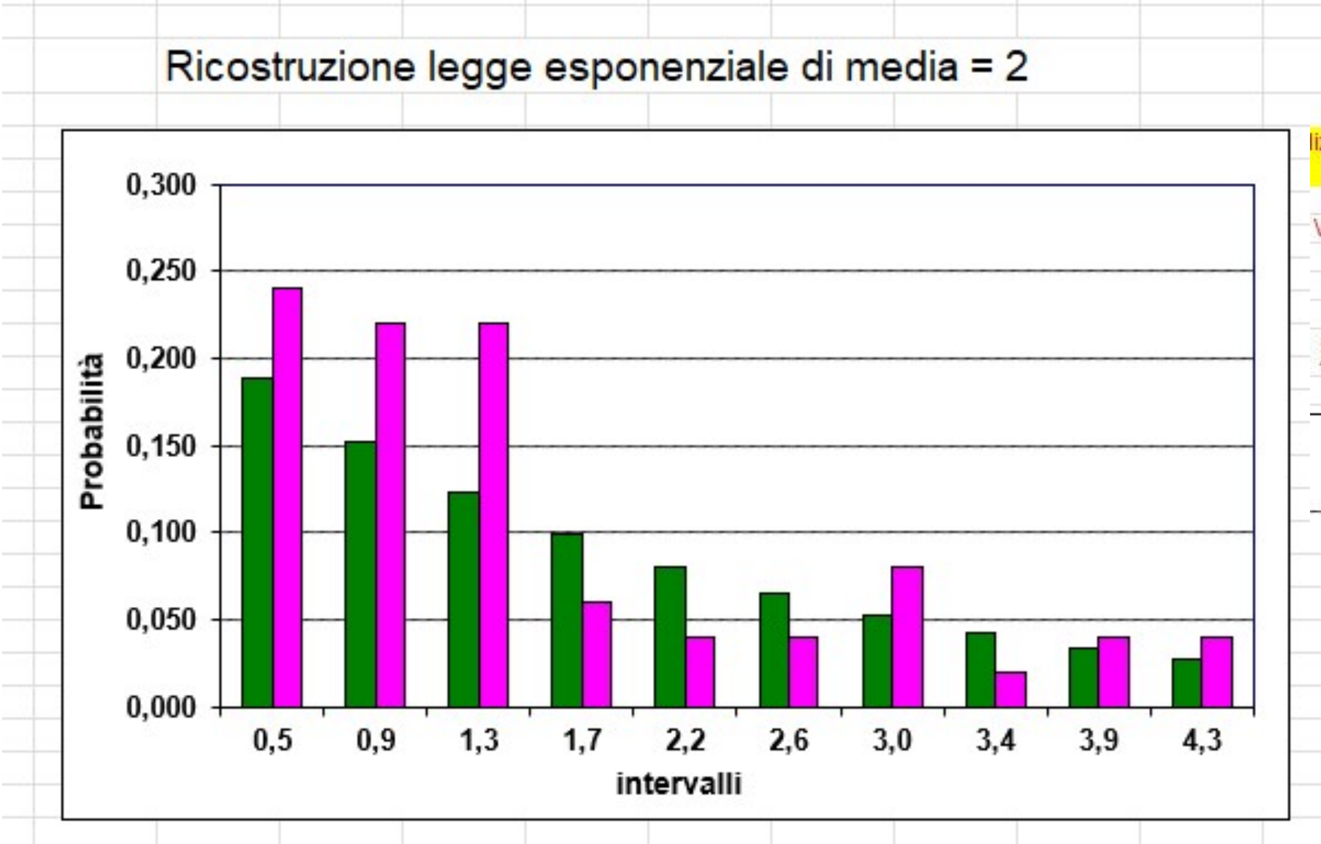
0,46	0,23	0,44564	0,19001
0,89	0,67	0,35703	0,15223
1,31	1,10	0,28848	0,123
1,74	1,53	0,23309	0,09939
2,17	1,95	0,18834	0,0803
2,59	2,38	0,15218	0,06489
3,02	2,81	0,12296	0,05243
3,45	3,23	0,09935	0,04236
3,87	3,66	0,08028	0,03423
4,30	4,08	0,06487	0,02766

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

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

$$\begin{aligned} &P(x \le X \le x+\Delta x) \\ &= F_X(x+\Delta x) - F_X(x) \\ &\cong f_X(x)\Delta x \end{aligned}$$

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.



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

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

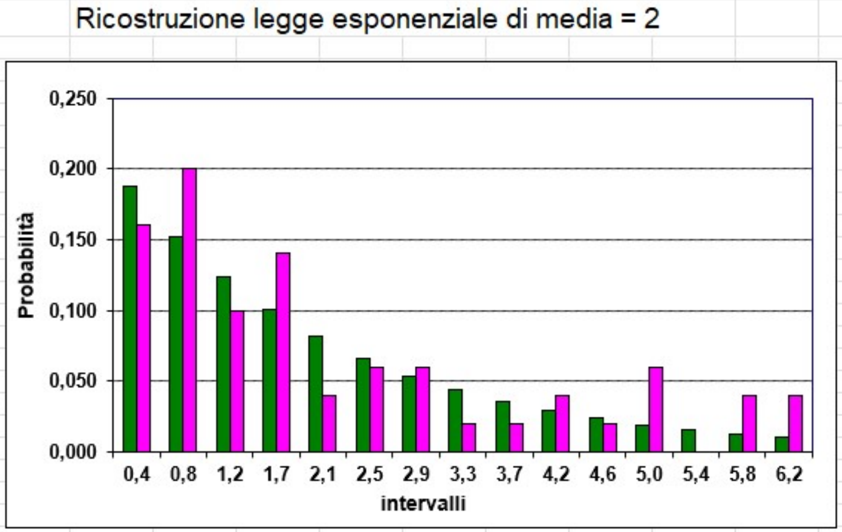
38		$\cong f_X(x)\Delta x$
classe		
0,23	0,44564	0,19001
0,67	0,35703	0,15223
1,10	0,28848	0,123
1,53	0,23309	0,09939
1,95	0,18834	0,0803
2,38	0,15218	0,06489
2,81	0,12296	0,05243
3,23	0,09935	0,04236
3,66	0,08028	0,03423
4,08	0,06487	0,02766

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

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

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:

n° classi	ampiezza classi (variabile)			
15	0,41535			$\cong f_X(x)\Delta x$
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



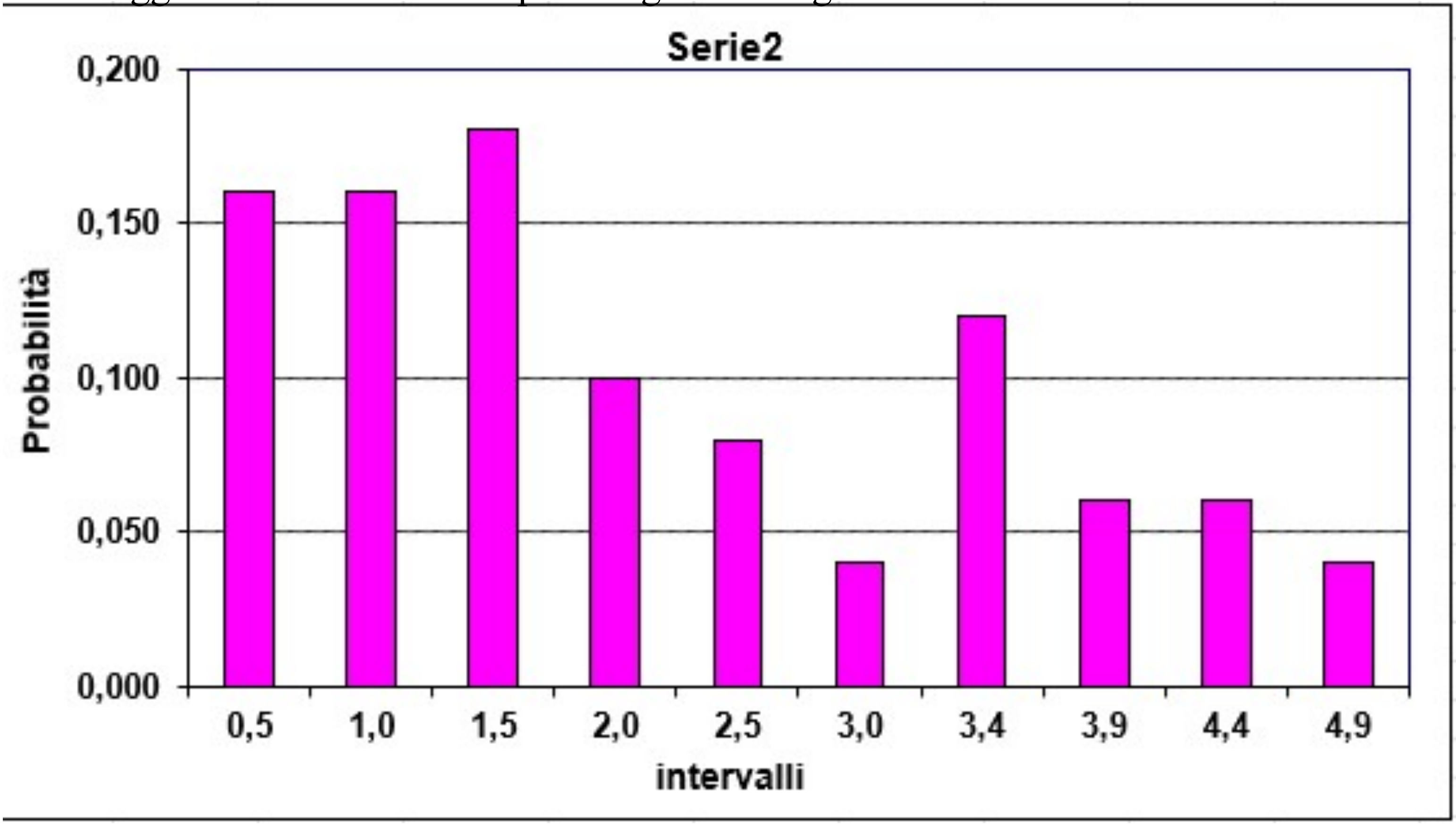
$$P(x \le X \le x+\Delta x)$$
$$= F_X(x+\Delta x) - F_X(x)$$
$$\cong f_X(x)\Delta x$$

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

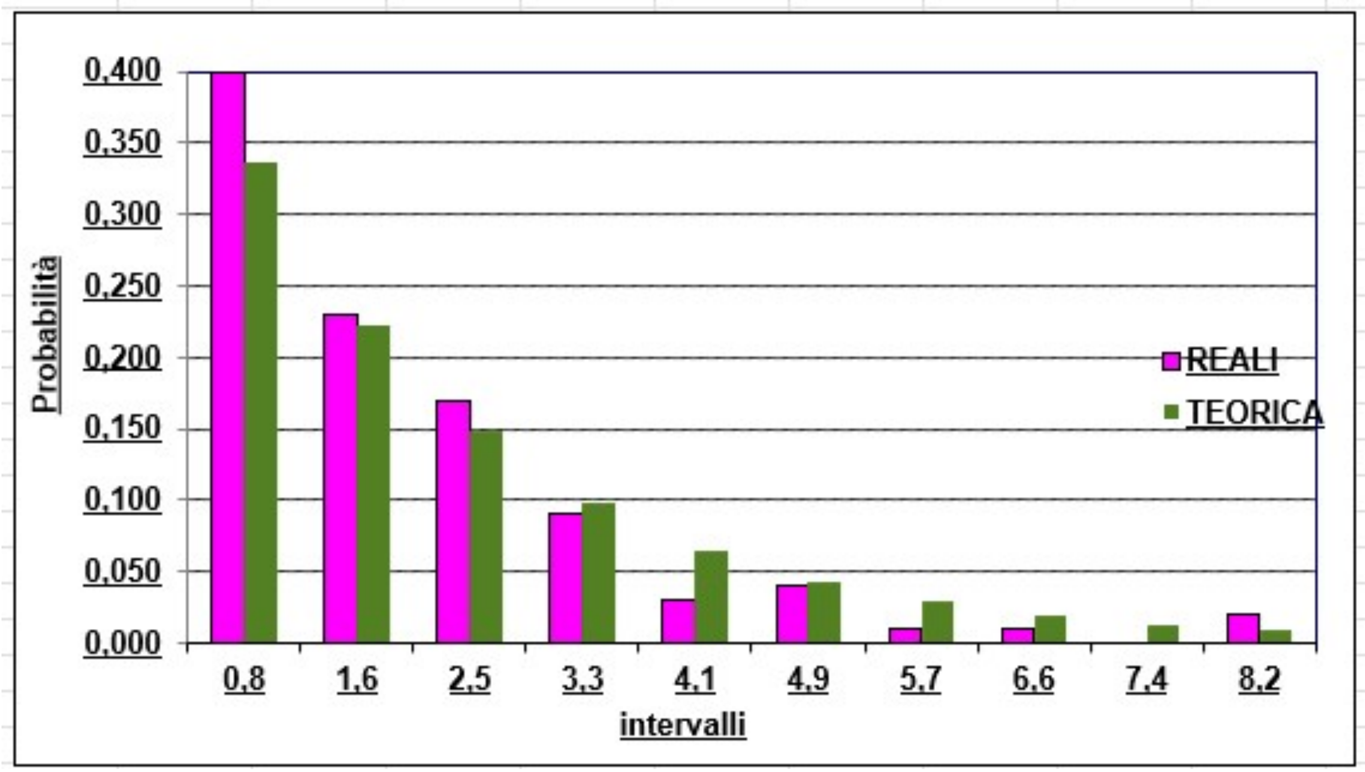
0,187	8	0,160
0,152	18	0,200
0,124	23	0,100
0,100	30	0,140
0,082	32	0,040
0,066	35	0,060
0,054	38	0,060
0,044	39	0,020
0,036	40	0,020
0,029	42	0,040
0,023	43	0,020
0,019	46	0,060
0,015	46	0,000
0,013	48	0,040
0,010	50	0,040

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

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.

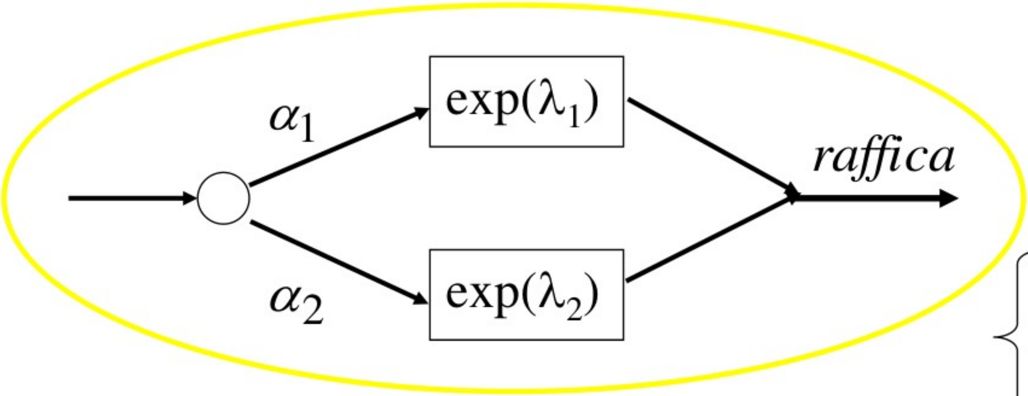


Nel grafico precedente si vede che anche se con media sperimentale 1.62(!), se uso 100 realizzazione quindi il doppio dei dati, per 10 classi la curva della legge esponenziale è abbastanza simile alla teorica. Prendiamo adesso la distribuzione iperesponenziale:

Trasformazione inversa per la distribuzione iperesponenziale

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

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



$$\begin{cases} \alpha_1 = 1-\varepsilon, \ \alpha_2 = \varepsilon \\ \lambda_1 >> \lambda_2 > 0 \end{cases}$$

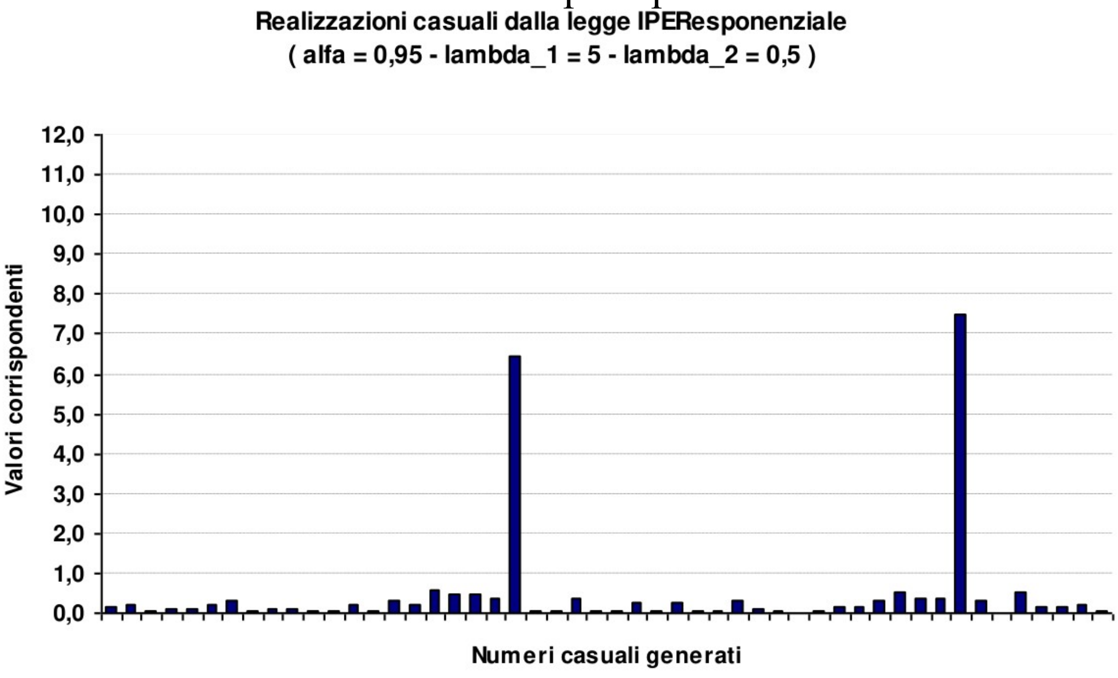
PROCEDURA:

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

$$if \quad u \leq \alpha_1 \quad then \quad genera\left(y \approx EXP(\lambda_1)\right)$$

$$else \quad genera\left(y \approx EXP(\lambda_2)\right)$$

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



Metodo della trasformazione inversa e legge di Erlang

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

Somma di esponenziali identiche e indipendenti

$$\begin{array}{ccccccc} 1 & & 2 & & & & n \\ \boxed{\exp(\lambda)} & \longrightarrow & \boxed{\exp(\lambda)} & \longrightarrow & \cdots\cdots & \longrightarrow & \boxed{\exp(\lambda)} \end{array}$$

$$X \hat{=} X_1 + \cdots + X_i + \cdots + X_n$$

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

$$u_i = 1 - e^{\lambda x_i}$$

$$x_i = -\frac{1}{\lambda} \ln(1 - u_i)$$

$$x \hat{=} \sum_{i=1}^{n} x_i$$

$$= \sum_{i=1}^{n} -\frac{1}{\lambda} \ln(1 - u_i) = -\frac{1}{\lambda} \ln\left(\prod_{i=1}^{n} (1 - u_i)\right)$$

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 ≥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. *X1 …Xn* e sia θ un parametro incognito di queste ultime; allora T(**X**) è detto **stimatore corretto di** θ se risulta E[T(**X**)] = θ.

- 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(**X**)] - θ) 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

$$MSE \triangleq E\left[(T(X_1,\ldots,X_n)-\theta)^2\right]$$ 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:

$$\lim_{n\to\infty} \Pr\left(\left|T(X_1,\ldots X_n)-\theta\right| \geq \varepsilon\right) = 0$$

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

quindi è consistente se per n tendente all'infinito, converge in probabilità al parametro da stimare

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"

X_1 pari alla realizzazione del primo run della monte carlo

E' detta <u>media campionaria</u> la var. al seguente :

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

singole osservazioni del campione

dimensioni del campione

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

media della popolazione

somma valori attesi

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

Valore atteso media campionaria

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

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

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

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

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.

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

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

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

varianza "bassa" allora stima più precisa

quadrato della dimensione del campione

somma degli scarti quadratici di ogni osservazione X_i del valore atteso della popolazione. Misura quanto ogni osservazione si discosta in media dal valore atteso

I passaggi algebrici saranno mostrati in una futura lezione!

$$= \quad \ldots \quad = \frac{\sigma^2}{n} \Rightarrow \bar{X}(n) \text{ stimatore di bontà crescente con "}n\text{"}$$

la varianza della media campionara

MODELLO PRODUTTORE-CONSUMATORE

MODELLO PRODUTTORE-CONSUMATORE

Tramite la simulazione monte carlo posso riprodurre la storia del sistema.

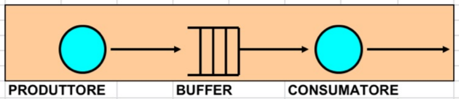
Tasso degli arrivi

t = 25	t = 50	100 =	200 =	400 =	800 =	1600 =	3200
1,16	1,1	1,07	1,01	1,045	1,001	1,048	1

Tasso delle partenze

t = 25	t = 50	100 =	200 =	400 =	800 =	1600 =	3200
0,68	0,84	1,05	0,89	1,013	0,998	1,034	1

$N(t)/t \qquad 1/E\,X \qquad t$



Si propone di studiare un classico modello gestionale produttore-consumatore (client-server) con distribuzione nota ed altrri utili vincoli come ad esempio impossibilità di guasti. La politica che si utilizzerà sarà FIFO e ci si pone di analizzare, il tempo medio di attesa in coda in un intervallo di tempo fissato. Dunque si andranno a valutare: il tempo di soggiorno (attesa) dell'i-esimo client, tempo di servizio dell'i-esimo client e l'interarrivo dell'i+1-esimo client, ovvero l'intervallo di tempo tra l'arrivo dell'i-esimo e quello dell'i+1-esimo. Graficamente:

In arancione l'arrivo dell'i-esimo, in rosso l'arrivo dell'i+1-esimo, in blu il tempo di servizio dell'i-esimo, in giallo l'interarrivo ed in verde il tempo di servizio dell'i+1-esimo.

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.

Tasso Interarrivi $\lambda = 1$ — Tasso di Servizio $\mu = \boxed{1,1}$ (Rate → velocità) — Tempi di Soggiorno

Valore Atteso in Input — Valore Atteso in Input — Valore Atteso Calcolato

Interarrivi (1/λ): 1 — Ordine — Tempi di Servizio (1/μ): 0,909090909 — E[S]=1/[μ•(1−ρ)]= 10

Media su 3200 Osservazioni — FIFO — Media su 3200 Osservazioni — Media su 3200 Osservazioni

Interarrivi: 0,958853087 — Tempi di Servizio: 0,907979337 — Tempi di Soggiorno: 13,92191877

$-\ln(1-u)\cdot\frac{1}{\lambda}$ — $-\ln(1-u)\cdot\frac{1}{\mu}$

Contatore		Numeri Casuali	Tempi di Interarrivo	Istanti di Arrivo	* Numeri Casuali	Tempi di Servizio	Istanti di Partenza	-0,760521178
1		0,514899008	0,663784499	0,663784499	0,385433153	0,866715914	1,530500413	
2		0,887865793	0,118934682	0,782719181	0,319521399	1,037210029	2,567710442	
3		0,887994107	0,118790172	0,901509353	0,202910298	1,449992071	4,017702513	
4		0,706357031	0,34763446	1,249143813	0,275110822	1,173255702	5,190958215	
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	10°	0,759480021	0,275121262	4,442132607	0,442506263	0,741182419	11,25288204	
11	11°	0,698349439	0,359035673	4,80116828	0,060268553	2,553586197	13,80646824	
12	12°	0,516391639	0,660889812	5,462058091	0,695429355	0,330205315	14,13667356	
13	13°	0,084551549	2,470393884	7,932451976	0,672166384	0,361135794	14,49780935	
14	14°	0,922743416	0,080404072	8,012856048	0,107281114	2,02936605	16,5271754	
15	15°	0,536793467	0,622141863	8,634997911	0,234827967	1,317183715	17,84435912	
16	16°	0,683086114	0,381134345	9,016132256	0,522356167	0,590368738	18,43472785	
17	17°	0,734663742	0,308342379	9,324474635	0,00711789	4,495585384	22,93031324	
18	18°	0,065606035	2,724087589	12,04856222	0,021809752	3,477634616	26,40794785	
19	19°	0,178758964	1,721716952	13,77027918	0,135614602	1,816307476	28,22425533	
20	20°	0,538244808	0,619441788	14,38972096	0,296575003	1,104959209	29,32921454	
21	21°	0,92482958	0,078145797	14,46786676	0,230974006	1,332227366	30,6614419	
22	22°	0,571644757	0,559237535	15,0271043	0,897498275	0,098312801	30,75975471	
23	23°	0,057883943	2,849315255	17,87641955	0,513670597	0,605611892	31,3653666	
24	24°	0,750335687	0,287234591	18,16365414	0,361302241	0,925491307	32,2908579	
25	25°	0,496853655	0,699459753	18,86311389	0,767112548	0,241019774	32,53187768	
26	26°	0,090907199	2,397916082	21,26102998	0,803401428	0,199000709	32,73087839	
27	27°	0,131174381	2,03122769	23,29225767	0,998592447	0,001280495	32,73215888	
28	28°	0,851756029	0,160455144	23,45271281	0,302429418	1,087188507	33,81934739	
29	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		0,217242197	1,526742431	29,34454424	0,504696894	0,621633854	40,75927878	

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 PERCHÈ 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

Da notare: tasso di interarrivi in questa simulazione è pari a 1, mentre il tasso di servizio è pari a 1.1

Tempo di Attesa in Coda
Valore Atteso Calcolato

ρ= E[W]=ρ/[μ∗(1−ρ)]= 9,090909091 — Modello analitico di riferimento

0,909090909 Media su 3200 Osservazioni

Tempo di Attesa in Coda → 13,01393943 13,01393943

passo 1,00

$W_i = W_{i-1} + t.servizio_{i-1} - t.interarrivo_i$

Tempi di Soggiorno	Tempi di Attesa in Coda	Equazione di Lindley W_i	Pearson	Correlazione	A(t) arrivi	D(t) departures	N(t) utenti presenti	t tempo
0,866715914	0	0	-0,152677476	-0,152677476				0,00
1,784991261	0,747781	0,747781232	-0,026784313		3	0	3	1,00
3,11619316	1,666201	1,666201089	-0,008168125		5	1	4	2,00
3,941814402	2,768559	2,7685587	0,01868795		7	2	5	3,00
4,124712027	3,940734	3,940734023	-0,026082237		8	2	6	4,00
3,986466977	2,870707	2,870706684	-0,023201384		11	3	8	5,00
5,447764384	3,646809	3,646808598	0,002299292		12	5	7	6,00
6,657270853	5,084018	5,084017765	-0,016934534		12	6	6	7,00
6,344688281	5,697894	5,697894041	-0,027880259		13	6	7	8,00
6,810749438	6,069567	6,069567018	-0,03519187		15	7	8	9,00
9,005299962	6,451714	6,451713765			17	8	9	10,00
8,674615465	8,34441	8,34441015			17	9	8	11,00
6,565357375	6,204222	6,204221581			17	10	7	12,00
8,514319352	6,484953	6,484953303			18	10	8	13,00
9,209361204	7,892177	7,892177489			19	11	8	14,00
9,418595597	8,828227	8,828226859			21	13	8	15,00
13,6058386	9,110253	9,110253218			22	13	9	16,00
14,35938563	10,881751	10,88175101			22	14	8	17,00
14,45397615	12,637669	12,63766868			23	15	8	18,00
14,93949357	13,834534	13,83453437			25	16	9	19,00
16,19357514	14,861348	14,86134778			25	16	9	20,00
15,73265041	15,634338	15,63433761						21,00
13,48894705	12,883335	12,88333515						22,00
14,12720376	13,201712	13,20171246						23,00
13,66876378	13,427744	13,42774401						24,00
11,46984841	11,270848	11,2708477						25,00
9,439901216	9,438621	9,438620721						26,00
10,36663458	9,279446	9,279446072						27,00
10,87855175	9,748204	9,748203972						28,00
11,22180258	7,919041	7,919040612						29,00
12,07879412	10,84444	10,84443974						30,00
12,3198431	11,66901	11,66900972						31,00

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 :

tempo d'attesa i-esimo + tempo servizio i-esimo - interarrivo i+1-esimo

I tempi di soggiorno sono variabili aleatorie, il primo valore coincide con il servizio. Immagino al sequenza temporale di soggiorno (X1,X2...Xn), 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

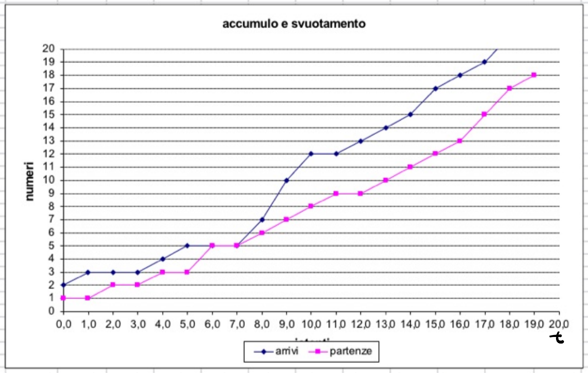
- Potrebbe servire la statistica? La realizzazione di S3 rispetto a quella di S5 è indipendente? O meglio, la sequenza dell'uscita del sistema di oseervazioni è indipendente? Gli input per ipotesi sono indipendenti. La realizzazione di S5 è correlata a quella di S3. Le due realizzazioni sono identicamente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

→ se dovessi studiare questo sistema, spero che per $n \to \infty$ 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)

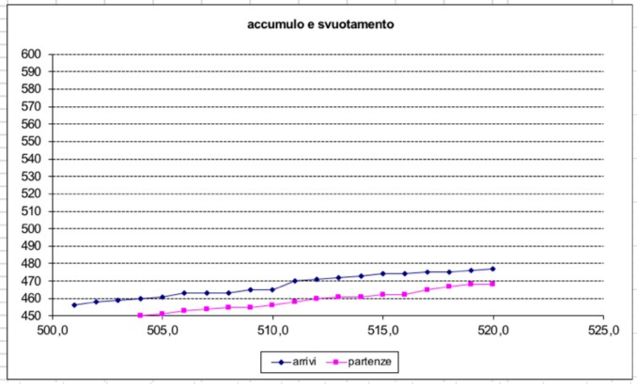
↳ Mi focalizzo sul tempo di attesa in coda (è una media aritmetica). Se clicco sulla sua casella mi rendo conto che salto da circa 10 a circa 5 con una "pigiata", quindi che varianza c'è? Come la stimo? Che "fiducia" posso dare a quel valore? Nascono le stime intervallari (intervallo di confidenza). 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:



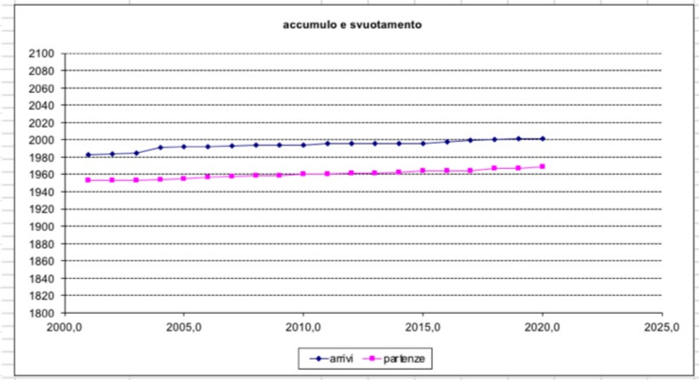
$0 \leq t \leq 20$ (inizio)

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)

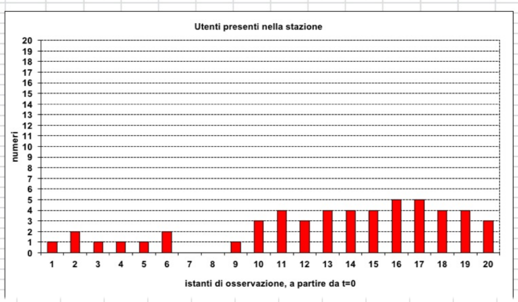
Quando il sistema arriva a regime il numero di arrivi e partenze tende a uguagliarsi

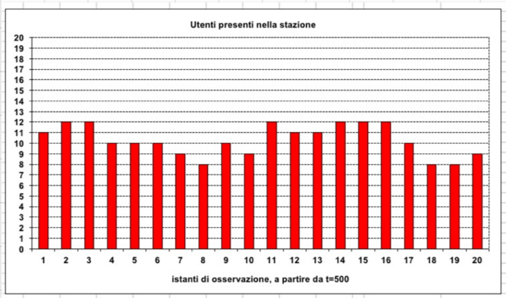
Arrivi (x_0, y_0) $-$ Partenze$(x_0, y_0) =$ numero di persone coinvolte

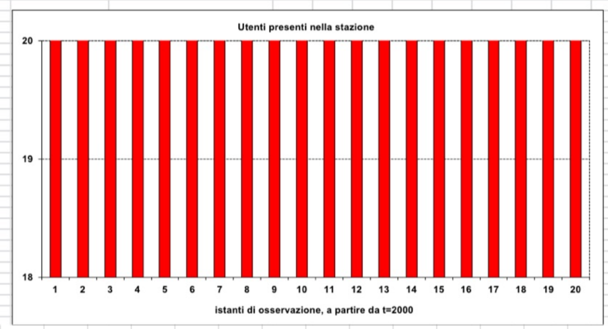
Tasso Interarrivi	1		Tasso di Servizio	1,1	Tempi di Soggiorno	
Valore Atteso in Input			Valore Atteso in Input		Valore Atteso Calcolato	
Interarrivi (1/λ)	1	Ordine	Tempi di Servizio (1/μ)	0,909090909	E[S]=1/[μ•(1−ρ)]=	10
Media su 3200 Osservazioni		FIFO	Media su 3200 Osservazioni		Media su 3200 Osservazioni	
Interarrivi	0,958853087		Tempi di Servizio	0,907979337	Tempi di Soggiorno	13,92191877

In questo caso, TASSO DI SERVIZIO > TASSO INTERARRIVI

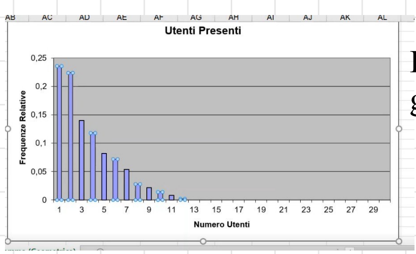
$\hookrightarrow$ 10% più veloce







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



Il grafico rappresenta proprio il modello geometrico di occupazione del buffer

Cosa accade se tasso di servizio<tasso interarrivi? Cosa accade se se tasso di servizio=tasso interarrivi? Cosa accade se tasso di servizio>>tasso interarrivi? Da notare che con un tasso 10% più alto di quello degli interarrivi, il sistema comunque tende a congestionarsi.

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

$$\tilde{Z}_n = \frac{\bar{X}(n) - E[\bar{X}(n)]}{\sqrt{VAR\left[\bar{X}(n)\right]}} = \frac{\bar{X}(n) - \mu}{\sqrt{\sigma^2 / n}}$$

ricordando che il TCL riguarda la somma di variabili aleatorie.

che tende alla normale standard per n → ∞ e, di conseguenza:

$$\bar{X}(n) = \frac{\sigma}{\sqrt{n}} \tilde{Z}_n + \mu$$

tende ad essere distribuita come una legge normale di media μ e varianza σ^2/n 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

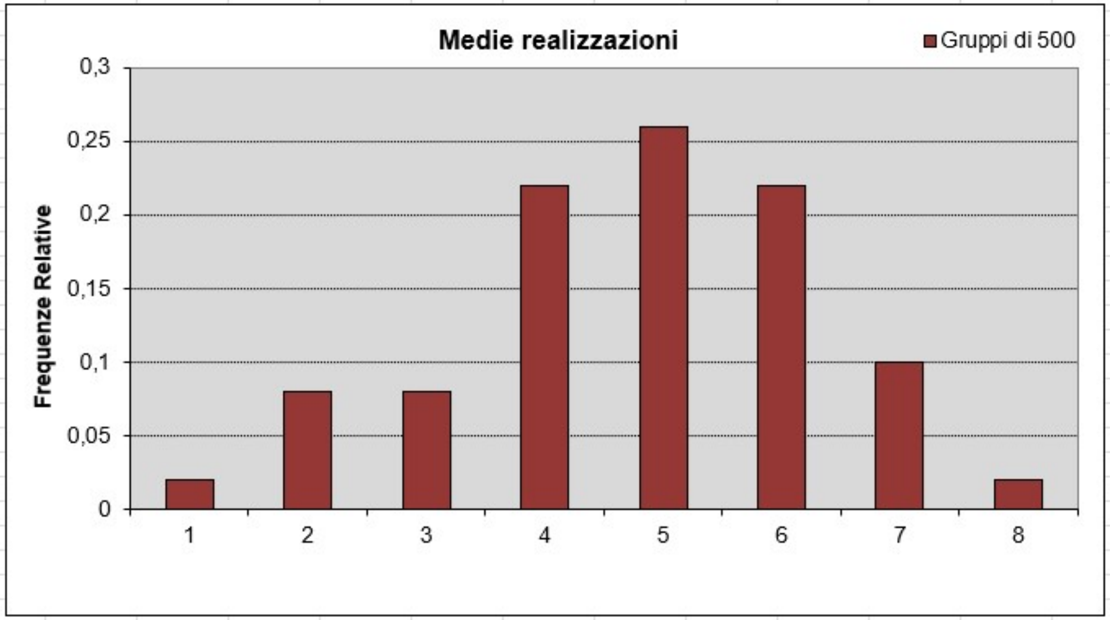
Per il modello produttore consumatore

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

4
11
13
11
5
1

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 200 e poi ancor di più con un numero di capioni 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 $X_1, X_2, \ldots, X_n$ variabili aleatorie tutte indipendenti e identicamente distribuite

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

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

$$\overline{X}(n) \triangleq \frac{\sum_{i=1}^{n} X_i}{n} \approx N(\quad, \frac{\sigma^2}{n})$$

per il "teorema di riproducibilità" della normale.

Rammentando che: (grazie all'indipendenza)

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

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:

$$\Pr[-\underset{\text{quantile}}{z_{\alpha/2}} \le Z \le \underset{\text{quantile}}{z_{\alpha/2}}] = 1 - \alpha$$

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

dove:

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

si ottiene: (isolando μ)

$$\Pr\left[\overline{X} - \frac{\sigma}{\sqrt{n}} z_{\alpha/2} \le \quad \le \overline{X} + \frac{\sigma}{\sqrt{n}} z_{\alpha/2}\right] = 1 - \alpha$$

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

$$\left[\overline{X} - \frac{\sigma}{\sqrt{n}} z_{\alpha/2},\ \overline{X} + \frac{\sigma}{\sqrt{n}} z_{\alpha/2}\right] \qquad \text{(IC)}$$

Da notare che l'intervallo 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'intervallo: ogni punto dell'intervallo ha la stessa probabilità di essere .

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

Infine, posto $d \triangleq z_{\alpha/2}\sigma / \sqrt{n}$, si tenga presente che la numerosità (n) del campione

richiesto per stimare con un intervallo di ampiezza $2d$ risulta:

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

Per progettare un intervallo di confidenza serve un parametro ed una statistica

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:

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

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

  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 <u>buona approssimazione</u> dell'intervallo vero già con $n \geq 30$.

<u>NOTA</u>:

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)

valori individuali di un campione di dimensione "n"

$$Siano \quad X_1, X_2, \ldots, X_n \quad I.I.D.$$

$E[X_i]$ è il valore atteso della popolazione da cui è estratto il campione

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

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

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

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

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

"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 vaarianza 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:

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

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

quadrato della varianza!

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

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	ESPONENZIALE	scarti	scarti^2
	0,25	B → tempi di soggiorno	A	A
0,907059		9,503177	5,649488	31,91671
0,071457	MEDIA	0,296555	-3,55713	12,6532
0,475876	3,853689	2,58411	-1,26958	1,61183
0,620852		3,879311	0,025622	0,000656
0,881526	SOM SCA	8,532253	4,678564	21,88896
0,242467	2152,032	1,110751	-2,74294	7,523707
0,62759		3,951039	0,097349	0,009477
0,94516		11,61336	7,759668	60,21245
0,210836		0,947125	-2,90656	8,448114
0,315209		1,514568	-2,33912	5,471486
0,750039		5,545806	1,692117	2,863259
0,293832		1,39161	-2,46208	6,061832
0,202667		0,905933	-2,94776	8,689265
0,146378		0,633068	-3,22062	10,3724
0,844105		7,434278	3,580589	12,82061
0,571085		3,385983	-0,46771	0,218749
0,3715		1,857675	-1,99601	3,984071
0,997946		24,75135	20,89767	436,7124
0,815052		6,750719	2,89703	8,392782
0,010066		0,040468	-3,81322	14,54065
0,256869		1,187533	-2,66616	7,10839
0,464483		2,49809	-1,3556	1,83765

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

stima della media e della varianza	n	n-1	alfa	alfa/2	1-alfa/2
A	100	99	0,05	0,025	0,975

Intervallo di confidenza per la media

$z_{\alpha/2}$ = 1,96

Media capionaria ← **Stimatore puntuale della media**
3,8536892
Varianza campionaria ← **Stimatore puntuale della varianza**
21,737698
scarto quadratico medio — semi ampiezza / quantile
4,6623705 — 0,0466237
intervallo bilaterale per la media al 95%
delta 0,91380782
limite inferiore
2,9398813
limite superiore
4,767497

$$\overline{X}-\frac{S}{\sqrt{n}}z_{\alpha/2}<\mu<\overline{X}+\frac{S}{\sqrt{n}}z_{\alpha/2}$$

2,9398813 ≤ μ ≤ 4,767496981

$\frac{1}{\lambda}$ = 4

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.

$z_{\frac{\alpha}{2}}$ valori della Gaussiana in corrispondenza del quantile $\frac{\alpha}{2}$

1.96 è l'ascissa entro cui realizzo $\frac{\alpha}{2}$

Lavoro con un modello esponenziale applicato ad una realizzazione di una variabile aleatoria di forma nota. Nel file vengono riportati 2 casi: quello normale ed esponenziale. Il primo rappresenta gli intervalli di confidenza per la Gaussiana. In verità i tempi di soggiorno non sono né gaussiani né esponenziali. Ipotizzo che siano esponenziali (input e output esponenziale)

casuale	media	10	normale(μ,σ)	scarti	scarti^2
	varianza	4	B	A	A
0,93641			13,05063868	2,99304	8,9583
0,90393			12,60859815	2,551	6,50761
0,38798			9,430834362	-0,62676	0,39283
0,73052			11,22875651	1,17116	1,37161
0,66088			10,82973808	0,77214	0,5962
0,18201			8,184505265	-1,87309	3,50847
0,45214			9,759509861	-0,29809	0,08886
0,96925			13,73990095	3,6823	13,5594
0,46606			9,829635879	-0,22796	0,05197
0,53401			10,17071052	0,11311	0,01279
0,79892			11,67553457	1,61794	2,61772
0,51266			10,06350167	0,0059	3,5E-05
0,06273			6,935544614	-3,12205	9,74721
0,79957			11,68020263	1,62261	2,63285
0,00776			5,159899376	-4,8977	23,9874
0,95412			13,37227692	3,31468	10,9871
0,68681			10,97364108	0,91604	0,83914
0,05664			6,832797054	-3,2248	10,3993
0,05435			6,7918386	-3,26576	10,6652
0,16272			8,033329709	-2,02427	4,09766
0,85231			12,09279673	2,0352	4,14204
0,65749			10,81127127	0,75367	0,56802

stima della media e della varianza	n	n-1	alfa	alfa/2	1-alfa/2
A	50	49	0,05	0,025	0,975

dimensione del campione

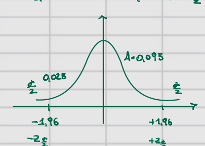
Intervallo di confidenza per la media

$z_{\alpha/2}$ = 1,96

ascissa entro cui realizzo $\frac{\alpha}{2}=0{,}025$

valori della Gaussiana in corrispondenza del quantile $\frac{\alpha}{2}$

Media capionaria ← **Stimatore puntuale della media**
10,057597
Varianza campionaria ← **Stimatore puntuale della varianza**
4,0654709
scarto quadratico medio
2,0163013
intervallo bilaterale per la media al 95%
delta 0,55887993
limite inferiore
9,4987172
limite superiore
10,616477



se integro la Gaussiana da $-\infty$ a $-1{,}96$ ottengo $\frac{\alpha}{2}=0{,}025$

$$\overline{X}-\frac{S}{\sqrt{n}}z_{\alpha/2}<\mu<\overline{X}+\frac{S}{\sqrt{n}}z_{\alpha/2}$$

9,4987172 ≤ μ ≤ 10,6164771

10

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

aumento la varianza da 10 a 20 per aggiornare i valori

Media capionaria ← **Stimatore puntuale della media**
10,085473
Varianza campionaria ← **Stimatore puntuale della varianza**
13,985031 → ho 30% di errore
scarto quadratico medio
3,7396565
intervallo bilaterale per la media al 95%
delta 1,03656085
limite inferiore
9,0489125
limite superiore
11,122034

$$\overline{X}-\frac{S}{\sqrt{n}}z_{\alpha/2}<\mu<\overline{X}+\frac{S}{\sqrt{n}}z_{\alpha/2}$$

9,0489125 ≤ μ ≤ 11,12203421

10

La media campionaria prende la varianza del processo e la lascia immutata. La varianza inve l'amplifica

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.

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

Stima centrale che oscilla tra il suo valore meno qualcosa ed il suo valore sommato a qualcosa

SOTTO QUALI IPOTESI L'INTERVALLO È CORRETTO? → LEZIONE 16

$\overline{X}$ è la media campionaria

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

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

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

L'ampiezza dell'intervallo è proporzionata 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

⚠ 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.

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:

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

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 rispettivamenta 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".

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

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

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

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

$$= \int_0^z \left(1/\Gamma(1/2)2^{1/2}\right) \cdot v^{1/2-1} \cdot \exp(-v/2) \cdot dv$$ 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: $X_{\gamma_1}^2, X_{\gamma_2}^2 \ldots X_{\gamma_n}^2$ indipendenti, con i rispettivi gradi di libertà: $\gamma_1^2, \gamma_2^2 \ldots \gamma_n^2$ allora:

$X_{\gamma_1}^2 + X_{\gamma_2}^2 + \ldots + X_{\gamma_n}^2$ è ancora una chi-quadrato, con gradi di libertà pari a: $\gamma 1, \gamma 2 \cdots \gamma n$

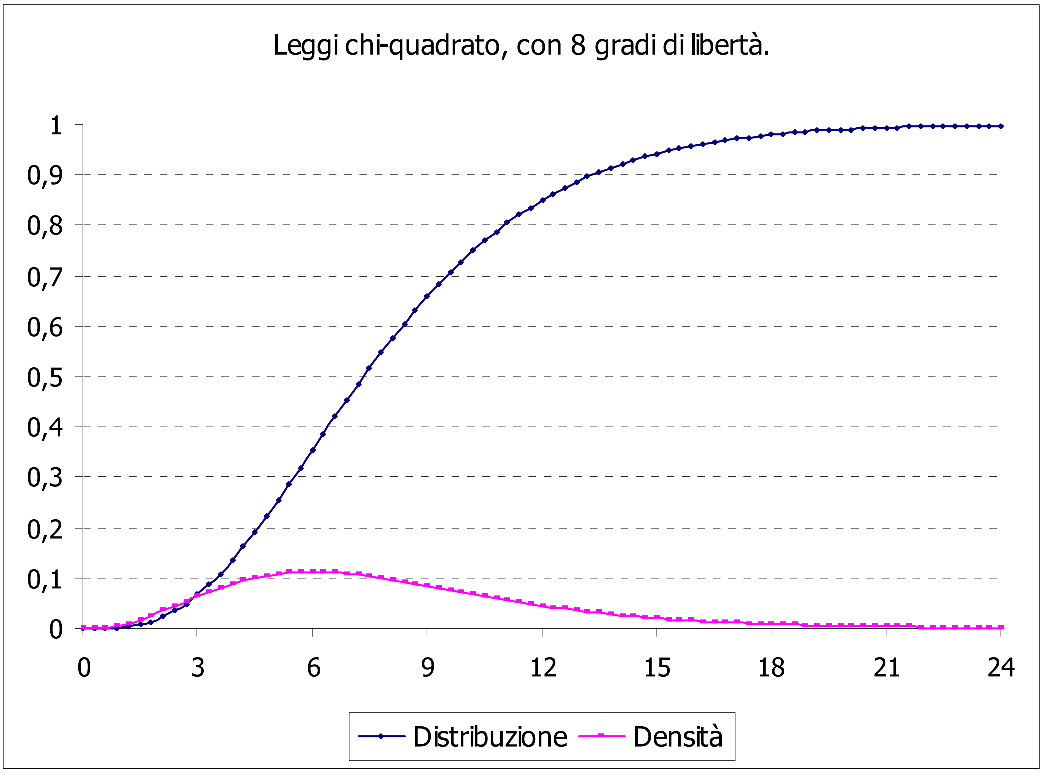
Attraverso le trasformate di Laplace ed al prodotto di convoluzione si prova il teorema.

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

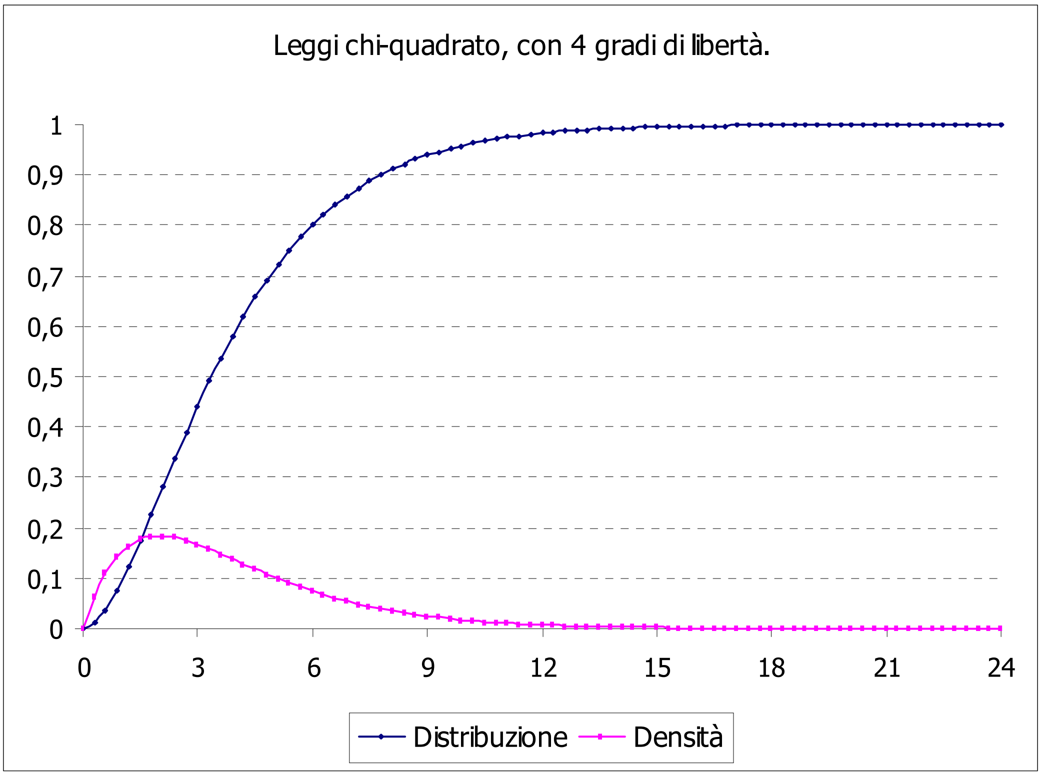
In parole povere il teorema afferma che la somma di chi-quadrato è ancora una chi-quadrato dove i gradi di libertà si sommano. Come detto è importante per "stimare" la varianza e la bontà della forma. (Si vedrà in futuro)(Test sulla forma con chi-square)

Bisognerebbe separare il grafico,
ad esempio il punto di intersezione
non si capisce ed una curva "sovrasta"
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

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 , si dovrebbe riconoscere che è stata fondamentale la disponibilità di una variabile aleatoria di distribuzione nota che conteneva quel parametro:

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

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 (nascita della chi-square) Stessa ipotesi della media campionaria

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

ricordando la varianza campionaria

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

statistica $\longrightarrow$ $$\frac{(n-1)S^2}{\sigma^2} = \frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \overline{X})^2 = X^2_{\gamma=n-1} \quad \text{(Tesi)}$$

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

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

ricorda che S è la stima della varianza

probabilità che la realizzazione della statistica cada nell'intervallo/quantili

da cui:

isolando sigma

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

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.

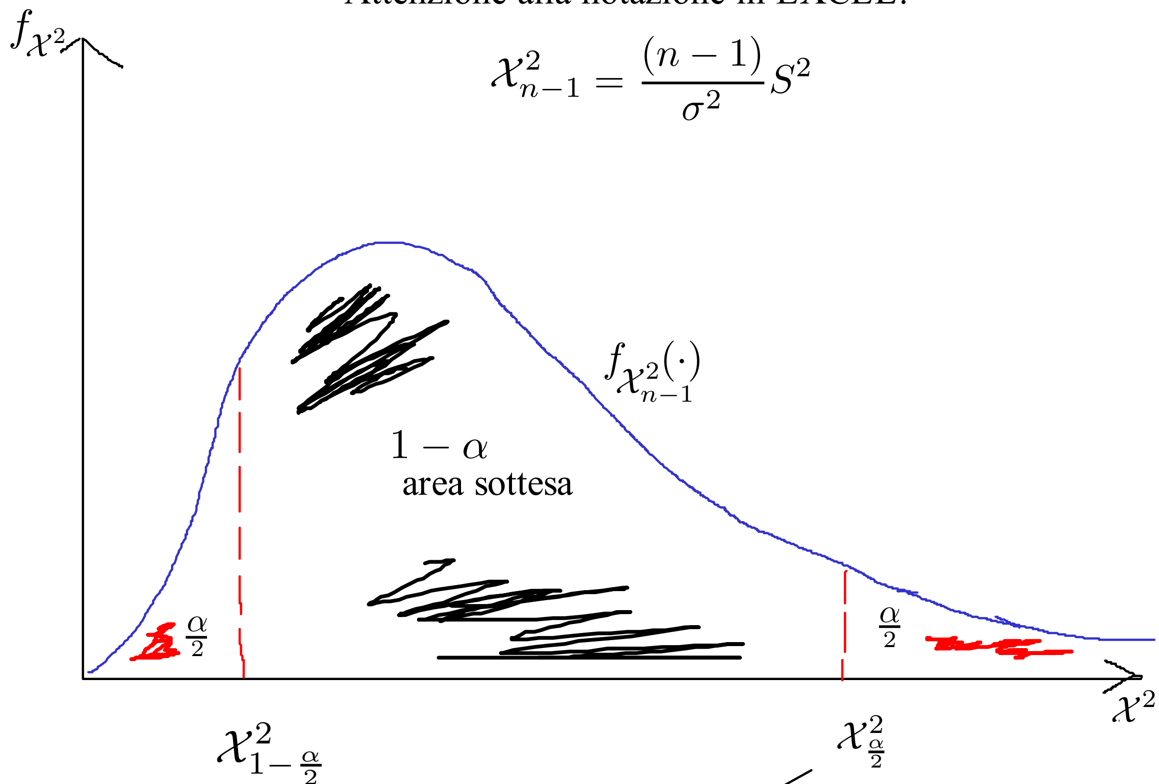
Da notare che in particolare gli intervalli di confidenza sulla varianza a differenza di quelli sulla media campionaria, escono "larghi" già in partenza. Si possono stringere quanto si vuole ma resteranno comunque "larghi". Vedi file Excel

INOLTRE NON SONO SIMMETRICI COME NEL CASO DELLA MEDIA CAMPIONARIA basta guarda i due denominatori dell'intervallo.

Attenzione alla notazione in EXCEL:

$$\mathcal{X}^2_{n-1} = \frac{(n-1)}{\sigma^2} S^2$$

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



complemento a 1 del quantile

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

Relazione base per la dimostrazione della tesi

$$\frac{(n-1)S^2}{\sigma^2} = \sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 - \left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)^2$$

$$\frac{(n-1)S^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\bar{X})^2 = \frac{1}{\sigma^2}\sum_{i=1}^{n}[(X_i-\mu)-(\bar{X}-\mu)]^2$$

$$= \frac{1}{\sigma^2}\left[\sum_{i=1}^{n}(X_i-\mu)^2 - 2(\bar{X}-\mu)\sum_{i=1}^{n}(X_i-\mu) + n\cdot(\bar{X}-\mu)^2\right]$$

$$= \frac{1}{\sigma^2}\left[\quad\cdots\quad -2(\bar{X}-\mu)\left(\sum_{i=1}^{n}X_i - n\mu\right) + \quad\cdots\right]$$

$$= \frac{1}{\sigma^2}\left[\quad\cdots\quad -2n(\bar{X}-\mu)^2 + \quad\cdots\quad\right] = \sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 - \left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)^2$$

Completamento della dimostrazione

A partire dalla relazione base:

$$\frac{(n-1)S^2}{\sigma^2} = \sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 - \left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)^2 \qquad (+)$$

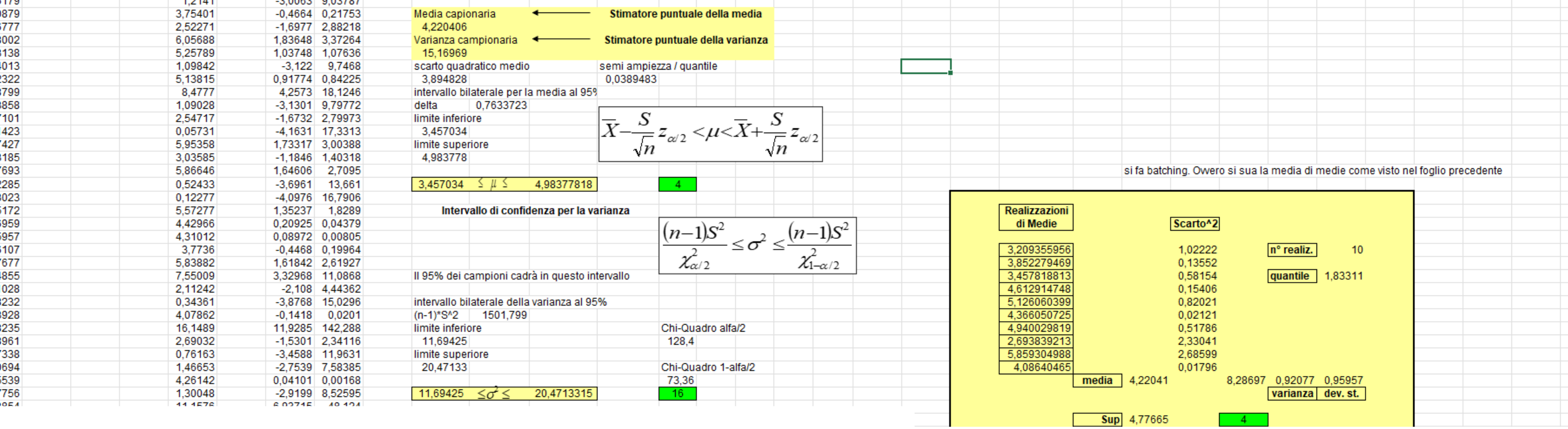
Si sfruttano i seguenti:

Teorema 1:
La variabile aleatoria "chi-quadrato" con "1" grado di libertà corrisponde al quadrato della "normale standard".

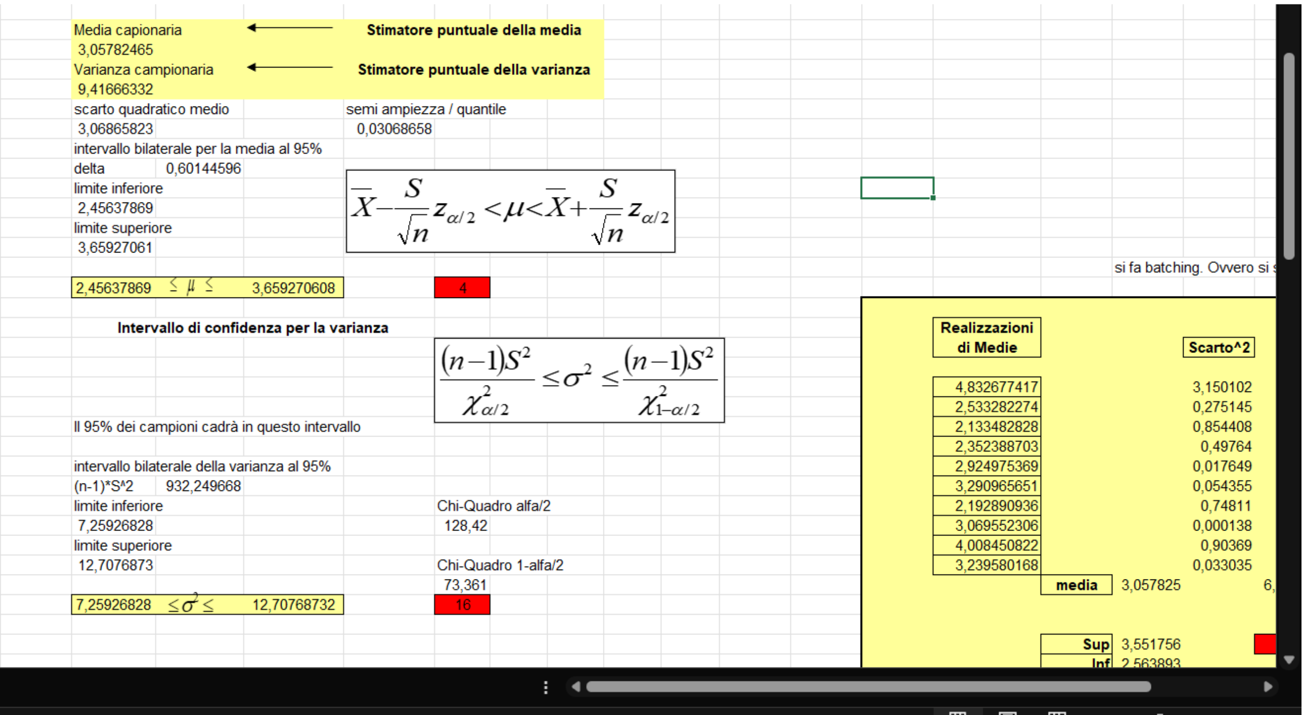
Teorema 2:
La somma di "n" variabili "chi-quadrato" con "1" grado di libertà corrisponde ad una "chi-quadrato" con "n" gradi di libertà.

Per interpretare la (+) alla seguente maniera:

$$X^2_{\gamma=n-1} = X^2_{\gamma=n} - X^2_1$$



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'intervallo 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

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

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

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

stimatore CORRETTO varianza

per n grande il rapporto si semplifica. quindi asintoticamente corretto

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

Quindi uno stimatore è asintoticamente corretto se per n tendente all'infinito risulta corretto

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

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

deviazione tra la stima e valore atteso

Lemma di Chebyshev:

$$\Pr\left\{ |T - \theta| \geq n \cdot \sqrt{Var[T]} \right\} \leq \left(1/n^2 \right)$$

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

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

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.

$$\mathsf{E}[S^2] = \mathsf{E}\left[\frac{\sum_{i=1}^{n} X_i^2 - n\bar{X}^2}{n-1}\right] = \frac{\sum_{i=1}^{n}\mathsf{E}[X_i^2] - n\mathsf{E}[\bar{X}^2]}{n-1}$$

$$= \frac{n}{n-1}\left(\mathsf{E}[X_1^2] - \mathsf{E}[\bar{X}^2]\right)$$ si usa: $VAR[X] = E[X^2] - E[X]^2$

$$= \frac{n}{n-1}\left(\mathsf{Var}(X_1) + (\mathsf{E}[X_1])^2 - \mathsf{Var}(\bar{X}) - (\mathsf{E}[\bar{X}])^2\right)$$

$$= \frac{n}{n-1}(\sigma^2 - \sigma^2/n) = \sigma^2. \quad \text{Done.}$$

Risultato fondamentale sull' errore quadratico medio

$$MSE \triangleq E\left[(T(X_1,\ldots,X_n)-\theta)^2\right]$$ è l'**errore quadratico medio**

<u>ALLORA RISULTA</u>:

$$\begin{aligned}
\mathrm{MSE}(T(\mathbf{X})) &= \mathrm{E}[T^2] - 2\theta\mathrm{E}[T] + \theta^2 \\
&\text{trucco per ottenere la varianza} \\
&= \mathrm{E}[T^2] - (\mathrm{E}[T])^2 + (\mathrm{E}[T])^2 - 2\theta\mathrm{E}[T] + \theta^2 \\
&\text{il bias sposta centro del intervallo ed è indotto dal casualità} \\
&= \underbrace{\mathrm{Var}(T)}_{\text{intervallo più largo o stretto}} + (\underbrace{\mathrm{E}[T] - \theta}_{\text{Bias}})^2 .
\end{aligned}$$

vista nell'intervallo di confidenza

Per il momento non fornisco alcun esempio di calcolo di MSE.
Mi limito a definire "<u>stima puntuale</u>" (del parametro θ) qualunque realizzazione della variabile aleatoria stimatore. Pertanto, maggiore è il MSE dello stimatore e più rischioso risulta affidarsi a quella stima.

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; distorto se MSE=$E[T(X)] \neq \theta$
2) Var $[T_1(X)] <=$ Var $[T_2(X)]$, per ogni valore di θ;
3) Var $[T_1(X)] <$ Var $[T_2(X)]$, per qualche valore di θ; per qualche valore è più preciso

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

<u>ESEMPIO</u>: la media campionaria è lo stimatore più efficiente tra quelli definibili tramite una combinazione lineare delle X1,…,Xn, di coefficienti a1,…,an.

Infatti:

essendo indipendenti

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

Da qui:

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

$$\begin{cases} \underset{a_1,\ldots,a_n}{MIN}: & \sigma^2 \sum_{i=1}^n a_i^2 \\ sub.to: & \sum_{i=1}^n a_i = 1, \end{cases} \Rightarrow \boxed{a_i = 1/n, \quad i = 1,\ldots,n.}$$

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.

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

binomiale ovvero
0,1. cade o non cade nell'intervallo

Grazie al teorema di Bernoulli, caso particolare della LGN:

$$\lim_{n\to\infty} P\left(\left|\bar{B}(n) - E\left[\bar{B}(n)\right]\right| \geq \varepsilon\right) = 0$$

appunto DEFINIZIONE stimatore consistente

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

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

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

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

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

Stima di parametri di una distribuzione (in input)
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”.

Ad esempio voglio calcolare il valore atteso in un modello che sembra un exp, allora modello a priori lo pongo appunto come exp.

DALLA FORMULA DI BAYES ALL'INFERENZA BAYESIANA

Ricordando Bayes

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

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

$B_1,\ldots,B_n$ è una partizione di Ω

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

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!

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

$P(\theta)$ è la prob. "a priori"

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

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

Il metodo dei momenti

$$Siano \quad X_1, X_2, \ \ldots, X_n \quad indip. \ e \ id. \ distr.$$
$$con \quad \mu^k \ momento\, di\, ordine\, "k", \quad k = 1,2,\ldots$$

Allora, pensando alla legge dei grandi numeri: (che vedremo)

$$\sum_{i=1}^{n} \frac{X_i}{n} \to E[X_i] \hat{=} \mu, \quad \sum_{i=1}^{n} \frac{X_i^2}{n} \to E[X_i^2] \hat{=} \mu_2, \quad \ldots$$

media campionaria
converge al valore atteso
come visto tramite monte carlo

Spesso è utile la seguente :

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

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

Il metodo dei momenti (MOM)

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

momento di ordine k

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

momento k-esimo

- Per il modello esponenziale di parametro λ, siamo alla banalità!

perchè è k=uno!

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

senza saperlo su Excel si stava attuando il metodo dei momenti

M O M per la distribuzione normale:

- **X1, …, Xn estratte da N(μ,σ^2)**
- **2 parametri da stimare $\Rightarrow$ sistema in 2 equazioni**:

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

appunto utilizzando: $VAR[X] = E[X^2] - E[X]^2$

- Dunque:

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

(stessi risultati
della stima MLE)

Es3: MOM per la distribuzione **iperesponenziale** (vedi anche file Excel)

Il metodo di massima verosimiglianza per la iperesponenziale è complicato da attuare

- Distribuzione iperesponenziale

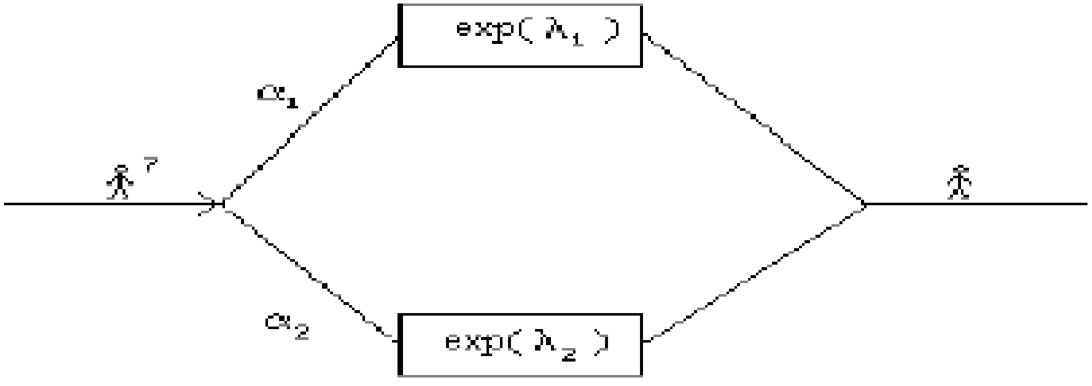
$$F_y(y) = \sum_{i=1}^{n} \alpha_i \left(1 - e^{-\lambda_i y}\right) \qquad y \geq 0$$

parametri della distribuzione

- 3 parametri da stimare: α , λ_1 ,λ_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 imporre un'ulteriore condizione per rendere determinato il sistema:

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

a questo punto sostituendo con gli stimatori di media e varianza

$\overline{X}$ 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 λ_1 e condizionato da λ_2

- 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

- Varianza:

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

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

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

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

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

$$\begin{cases} \dfrac{1}{\lambda_1+\lambda_2} = \dfrac{\dfrac{\alpha}{\lambda_1}+\dfrac{1-\alpha}{\lambda_2}}{2} & \text{condizione aggiunta} \\ \bar{X} = \dfrac{\alpha}{\lambda_1}+\dfrac{1-\alpha}{\lambda_2} & \text{media ottenuta} \\ S^2 = 2\left(\dfrac{\alpha}{\lambda_1^{\ 2}}+\dfrac{1-\alpha}{\lambda_2^{\ 2}}\right)-\left(\dfrac{\alpha}{\lambda_1}+\dfrac{1-\alpha}{\lambda_2}\right)^2 & \text{varianza ottenuta} \end{cases}$$

risoluzione che puoi "saltare"

- Osservando che il secondo membro della prima equazione è uguale a quello della seconda: $\lambda_1 = \frac{2}{\overline{X}} - \lambda_2$

- Dalla seconda equazione ricaviamo: $\hat{\alpha} = \frac{\lambda_1\left(\lambda_2 \overline{X} - 1\right)}{\lambda_2 - \lambda_1}$

- Sostituendo λ_1: $\alpha = 1 - \frac{\lambda_2 \overline{X}}{2}$

- Osserviamo che :

$$\frac{\alpha}{\lambda_1} = \frac{1-\alpha}{\lambda_2} = \frac{\overline{X}}{2} \qquad \frac{\alpha}{\lambda_1^{\,2}} = \frac{\overline{X}^2}{2\left(2 - \lambda_2 \overline{X}\right)} \qquad \frac{1-\alpha}{\lambda_2^{\,2}} = \frac{\overline{X}}{2\lambda_2}$$

- La terza equazione diventa:

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

da cui:

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

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

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

si arriva ad un problema di analisi numerica che si risolve

- … che ammette come soluzioni:

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

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

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

- Per completezza riscriviamo il terzo coefficiente: $$\boxed{\hat{\alpha} = \frac{\lambda_1\left(\lambda_2 \overline{X} - 1\right)}{\lambda_2 - \lambda_1}}$$

Il metodo della massima verosimiglianza.

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

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

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

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

La stima di massima verosimiglianza (MLE) di θ è il valore di θ che massimizza la funzione di verosimiglianza, ponendo uguale a zero la sua derivata rispetto il parametro o nel caso di più parametri ponendo le derivate parziale rispetto ad ogni parametro. Se all'interno della funzione inserisco anche i differenziali: $dx_1, dx_2, \dots, dx_n$ ottengo una produttoria tra probabilità. Quindi si mira a massimizzare la probabilità che le realizzazioni sono figlie della forma con il parametro incognito. Per facilitare la ricerca del massimo viene utilizzata la seguente funzione:

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

che è nota come funzione di **verosimiglianza ridotta**.
I massimi restano sempre uguali ed essendo il logaritmo una funzione monotona è corretto l'uso della funzione di verosimiglianza ridotta. Si utilizza il logaritmo anche perché distribuzioni comuni sono espresse con esponenziali e prodotti. Infatti come già detto per l'iperesponenziale, il logaritmo di una somma non facilita nulla mentre il logaritmo di prodotti esponenziali si.
Per essere sicuri che θ sia un massimo, piuttosto che un minimo o un punto di flesso, la derivata di $l(\theta)$, valutata in θ deve essere negativa.

Es. 1: la densità esponenziale,

A partire da: $f_x(X) = \lambda e^{-\lambda x}$ ricerco λ

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

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

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

non è detto che i due metodi diano la stessa stima

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

Es. 2: la distribuzione normale

Ho 2 parametri da stima, valore atteso e varianza quindi andrò a risolvere le derivate parziali per ognuno.

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

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

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

- Stima del parametro μ

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

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

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

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

- Stima del parametro σ^2

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

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

la stima qui risulta non distorta

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

Es. 3: la ripartizione bernoulliana

Example: Suppose $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Bern}(p)$. 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 $x_i, \quad i = 1, \ldots, n$

Lo stimatore di massima verosimiglianza:

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

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

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 campionara

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

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

Es. 4: la ripartizione poissoniana

Si parte dalla densità della Poissoniana per ricavare λ

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

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

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

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

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

$$\begin{cases} \dfrac{n}{\lambda}-\displaystyle\sum_{i=1}^{n} x_i^{\alpha}=0 \\ \dfrac{n}{\alpha}+\displaystyle\sum_{i=1}^{n}\ln x_i-\lambda\sum_{i=1}^{n} x_i^{\alpha}\ln x_i=0 \end{cases}$$

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

lambda di WEIB = 0,2500

alfa 2 assegnato. Per ipotesi stimato

somma x-i ^α 388,3024

lambda_cap 0,2575

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

media camp di real Weib ovvero media aritmetica la sua realizzazione = 1,753074621

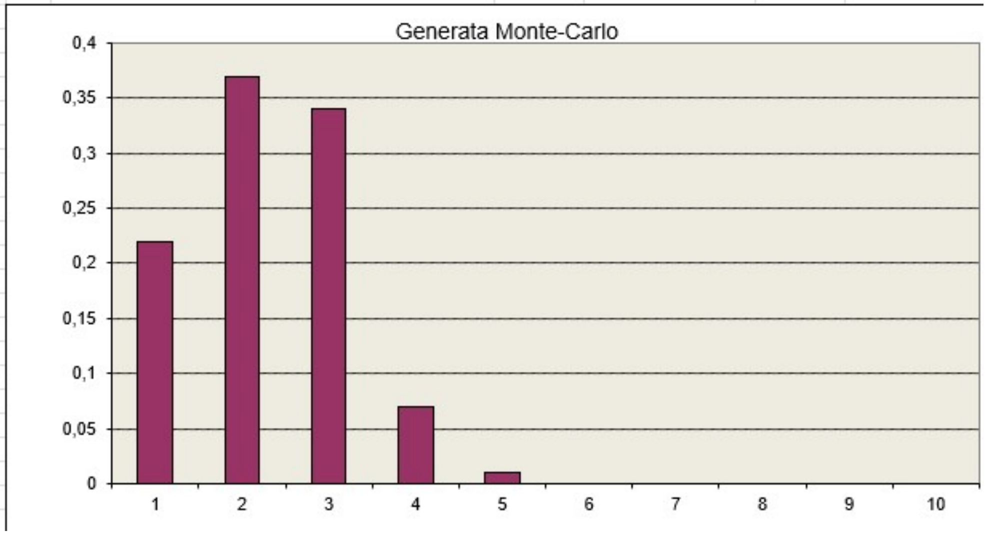
num intervalli = 10

num realizzazioni = 100

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

beta 2

Realizz di una Weibull	Real Weib ^alpha
2,278184076	5,190122686
2,980528663	8,883551111
1,374309393	1,888726308
4,09979484	16,80831773
2,178630612	4,746431344
0,513064168	0,26323484
1,272460527	1,619155792
0,729703423	0,532467086
0,371302031	0,137865198
2,035413355	4,142907527
1,164327462	1,355658439
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



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 densità "gamma"

Si ottiene dalla funzione gamma ed è un'estensione della densità di Erlang dove il parametro di forma diventa un reale positivo. Serve indirettamente a costruire altre funzioni utili nella statistica

- La funzione gamma (II integrale di Eulero):

$$\Gamma(\alpha) \hat{=} \int_0^{\infty} x^{\alpha-1} e^{-x} dx, \qquad \alpha \in \Re^+$$

E' facile verificare che: $\Gamma(1) = 1,$ meno facile che: $\Gamma(1/2) = \sqrt{\pi}$

- Integrando per parti, si ottiene la seguente:

$$\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1), \quad se \ \alpha > 1$$

e quindi: $\underbrace{\Gamma(n) = (n-1)!}_{\text{importante}}, \quad se \ \alpha \equiv n \in \mathrm{N}$

- E' utile riscriverla nella forma seguente, con $x = \lambda y$ e $\lambda>0$:

$$\Gamma(\alpha) = \int_0^{\infty} (\lambda y)^{\alpha-1} \cdot e^{-\lambda y} \cdot d(\lambda y) = \int_0^{\infty} (\lambda y)^{\alpha-1} \cdot e^{-\lambda y} \cdot \lambda dy$$

$$\Rightarrow 1 = \int_0^{\infty} \frac{(\lambda y)^{\alpha-1} \cdot \lambda e^{-\lambda y}}{\Gamma(\alpha)} \cdot dy \quad (*)$$

mostra come costruire una densità

La densità di Erlang

Erlang si ottiene a partire da una somma di esponenziali identiche e indipendenti

- Dalla densità di una Erlang $\{n,\lambda\}$:

La densità di Erlang può quindi essere riscritta con la funzione Gamma e passando da n numero naturale ad alpha numero reale si possono generare realizzazioni

$$f_X(x) = \frac{(\lambda x)^{n-1}}{(n-1)!} \lambda e^{-\lambda x}, \quad \lambda > 0, x \geq 0$$

la seguente <u>rispetta i requisiti di una densità</u> grazie alla (*):

$$f_X(x) = \frac{(\lambda x)^{\alpha-1}}{\Gamma(\alpha)} \lambda e^{-\lambda x}, \quad \alpha > 0, \lambda > 0, \; x \geq 0$$

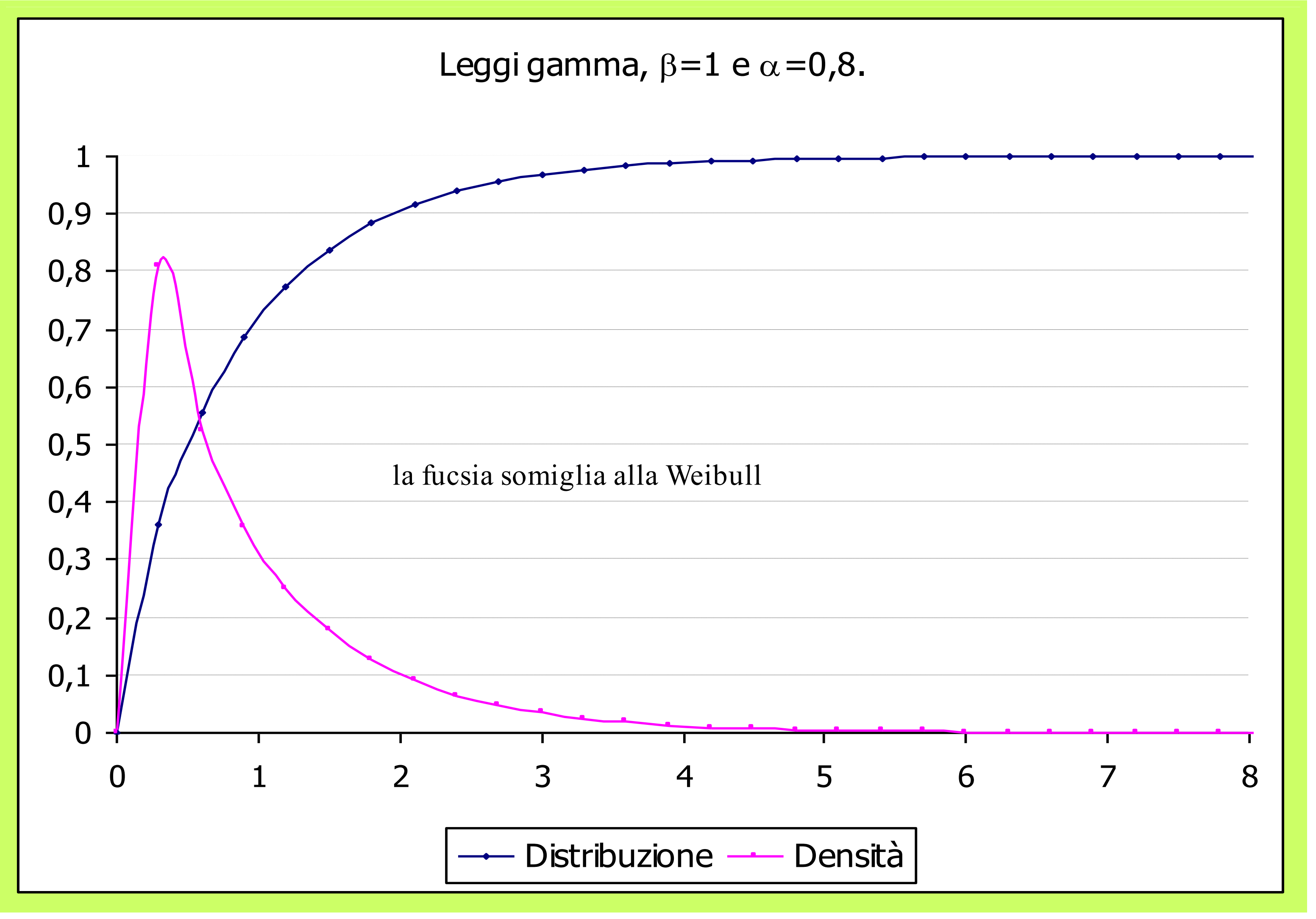
IN EXCEL: $$f_X(x) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta}, \quad \beta \hat{=} \frac{1}{\lambda}$$

parametro di scala perché rapporto tra 1 e valore atteso. Quindi alpha sarà parametro di forma

Valore atteso e varianza: $E[X] = \alpha\,\beta, \; Var[X] = \alpha\,\beta^2$

ricorrendo a: $L_X(s) \triangleq \int_{x=0}^{\infty} e^{-sx} f_X(x)dx = (1-\beta s)^{-\alpha}, \quad 0 \leq s < \beta^{-1}$

La densità e la distribuzione “gamma”



Ilenia Epifani (Poli_MI)
Appunti di Statistica
del prof. Invernizzi

MOM PER LA GAMMA

Esempio di applicazione n. 1: distribuzione gamma

Si supponga di disporre di un campione di n misurazioni, e si ipotizzi che le misurazioni siano tutte modellabili con variabili aleatorie indipendenti con distribuzione di probabilità $\Gamma(\alpha, \beta)$. Si ipotizzi inoltre che α e β siano parametri incogniti e che si desideri fornirne delle stime $\hat{\alpha}$ e $\hat{\beta}$.

Allora, seguendo il procedimento descritto poco fa, calcoliamo il momento primo $\bar{X}$ e il momento secondo M_2 (che saranno semplicemente dei numeri ottenuti a partire dai dati sperimentali), e costruiamo il sistema:

$$\begin{cases} \mu_1(\alpha, \beta) = \bar{X} & \text{media} \\ \mu_2(\alpha, \beta) = M_2 & \text{scarto quadratico} \end{cases}$$

Ricordando che in una distribuzione gamma la media è data dal prodotto tra i due parametri:

$$\mu_1(\alpha, \beta) = \alpha\beta$$

E che vale la relazione:

$$\mu_2(\alpha, \beta) - [\mu_1(\alpha, \beta)]^2 = Var(X) = \sigma^2 \quad \rightarrow \quad \mu_2(\alpha, \beta) = \sigma^2 + [\mu_1(\alpha, \beta)]^2 = \sigma^2 + (\mathrm{E}[X])^2$$

Dove, nel caso di distribuzione gamma:

$$\sigma^2 = \alpha\beta^2$$

In arancione la nota relazione
che la varianza è esprimibile come
momento del secondo ordine meno
momento del primo ordine al quadrato.

Andando a creare il sistema di equazioni

Otteniamo:

$$\begin{cases} \alpha\beta = \bar{X} \\ \alpha\beta^2 + (\mathrm{E}[X])^2 = M_2 \end{cases} \rightarrow \begin{cases} \alpha\beta = \bar{X} \\ \alpha\beta^2 = M_2 - (\mathrm{E}[X])^2 \end{cases} \rightarrow \begin{cases} \alpha\beta = \bar{X} \\ \alpha\beta^2 = M_2 - \bar{X}^2 \end{cases}$$

Notando che:

$$\sum_{j=1}^{n} (X_j - \bar{X})^2 = \sum_{j=1}^{n} X_j^2 - 2\bar{X}\sum_{j=1}^{n} X_j + n\bar{X}^2 = nM_2 - 2n\bar{X}^2 + n\bar{X}^2 = nM_2 - n\bar{X}^2$$

Possiamo riscrivere il sistema come segue:

$$\begin{cases} \alpha\beta = \bar{X} \\ \alpha\beta^2 = \dfrac{1}{n}\displaystyle\sum_{j=1}^{n} (X_j - \bar{X})^2 \end{cases}$$

Ricordando poi che lo stimatore varianza campionaria è definito come:

$$S^2 = \frac{1}{n-1}\sum_{j=1}^{n} (X_j - \bar{X})^2$$

Otteniamo facilmente:

$$\begin{cases} \alpha\beta = \bar{X} \\ \alpha\beta^2 = \dfrac{n-1}{n} S^2 \end{cases} \rightarrow \begin{cases} \dfrac{n-1}{n}\dfrac{S^2}{\beta} = \bar{X} \\ \alpha = \dfrac{n-1}{n}\dfrac{S^2}{\beta^2} \end{cases} \rightarrow \begin{cases} \hat{\beta}_{mom} = \dfrac{S^2}{\bar{X}}\dfrac{n-1}{n} \\ \hat{\alpha}_{mom} = \dfrac{\bar{X}}{\beta} = \dfrac{n}{n-1}\dfrac{\bar{X}^2}{S^2} \end{cases}$$

La disuguaglianza di Chebyshev

Sia Y variabile aleatoria di media μ e varianza σ^2 con funzione di densità definita da $-\infty$ a $+\infty$ quindi Gaussiana e non.
Si definisce disuguaglianza di Chebyshev quella sottostante, che è utilizzata per calcolare la probabilità che una realizzazione si trovi entro una certa distanza dalla media espressa in multipli della deviazione standard o meglio la probabilità che si trovi all'interno di un intervallo di confidenza.

h>0 costante

disuguaglianza vista nella lezione degli stimatori

$$\Pr\left\{\left|Y-\mu\right|\geq h\right\}\leq\frac{\sigma^2}{h^2}$$

scarto, ancora v.a.

se $k\sigma \hat{=} h \Rightarrow$ $\Pr\left\{\left|Y-\mu\right|\geq k\sigma\right\}\leq\frac{1}{k^2}$

considero
multipli della deviazione standard

ovvero:

esplicitando il valore assoluto
ottengo appunto un intervallo

$$\Pr\left\{\mu-k\sigma\leq Y\leq\mu+k\sigma\right\}\geq 1-\frac{1}{k^2}$$

con $k>0 \; : (1/k)\leq 1$

quantità arbitrariamente vicina a 1

Cantelli si usa per il caso di v.a. continue e non negative come Weibull, exp ecc.

Cantelli per densità in $[0,+\infty)\rightarrow$

estensione di Chebyshev

$$\Pr\left\{Y\leq\mu+k\sigma\right\}\geq 1-\frac{1}{1+k^2}$$

appunto prendo solo la disuguaglianza destra maggiore di zero

Prova:

partendo dalla def. di varianza oppure
ricordando il collegamento tra scarti e varianza

$$\sigma^2 \hat{=} \int_{-\infty}^{+\infty}(y-\mu)^2 f(y)dy$$

maggiore uguale perchè si taglia un pezzo di area, e dato che f(y)> 0, allora l'area è >=0 quindi anche $(y-\mu)^2 f(y)\geq 0$

$$\geq\int_{-\infty}^{\mu-h}(y-\mu)^2 f(y)dy+\int_{\mu+h}^{+\infty}(y-\mu)^2 f(y)dy$$

sia $h = y-\mu$

non si perde e non si deve perdere la disuguaglianza

$$\geq h^2\left(\int_{-\infty}^{\mu-h}f(y)dy+\int_{\mu+h}^{+\infty}f(y)dy\right)$$

con $h \hat{=} k\sigma \Rightarrow$ tesi

Il primo integrale è una distribuzione il secondo no!

1° integrale vale $Pr\{Y<\mu-h\}$
2° integrale vale $Pr\{Y>\mu+h\}$

$$\Rightarrow\sigma^2\geq h^2\left(\Pr\left\{\left|Y-\mu\right|\geq h\right\}\right)=\Pr\left\{\left|Y-\mu\right|\geq h\right\}\leq\frac{\sigma^2}{h^2}$$

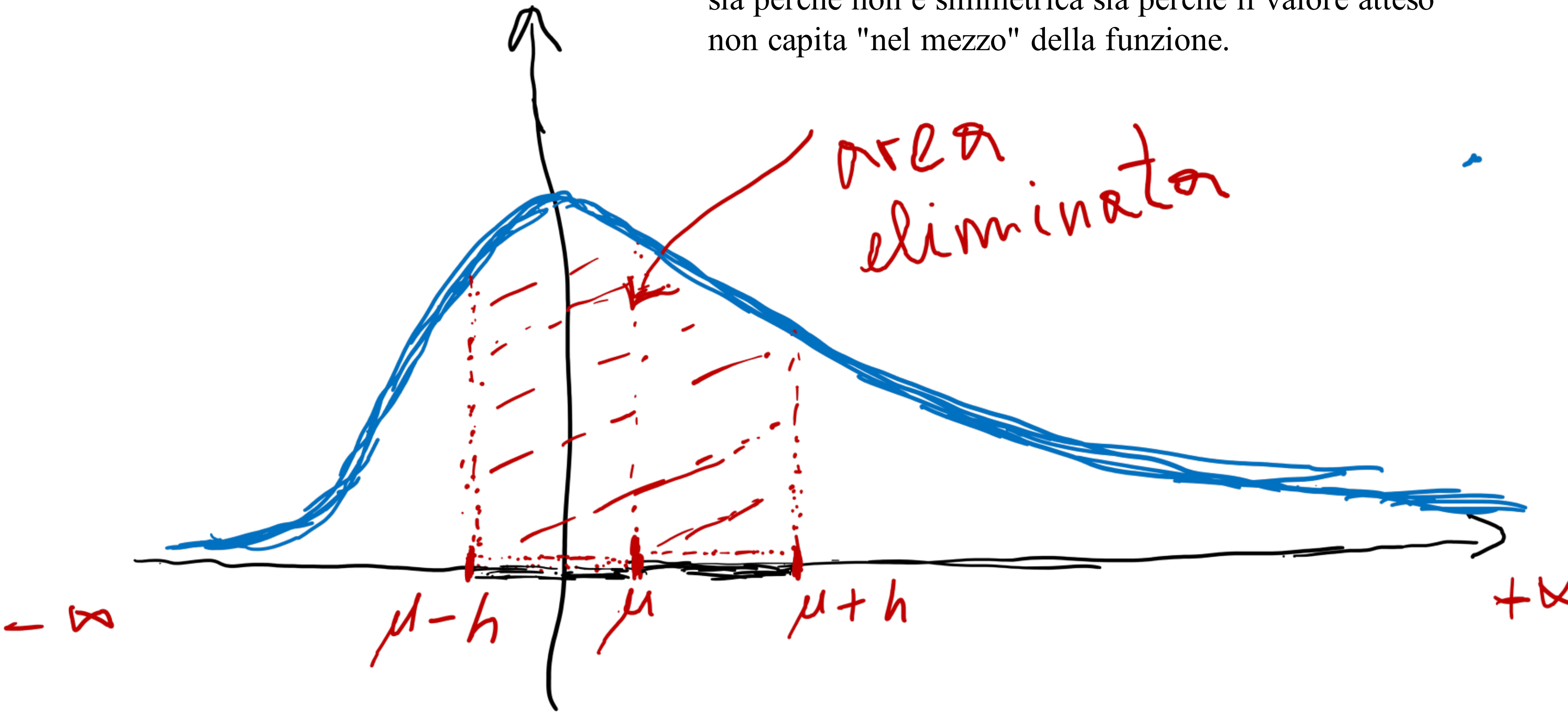
a questo punto ci si esprime con la deviazione standard

Se Y è la variabile aleatoria media campionaria, allora $VAR[Y] = \frac{\sigma^2}{n}$ e sostituendo nella disuguaglianza di Chebyshev, si ottiene che

$Pr\{|Y-\mu| \geq h\} \leq \frac{VAR[Y]}{h^2} = \frac{\sigma^2}{n} \cdot \frac{1}{h^2}$ se $h = \sigma$ $Pr\{|Y-\mu| \geq \sigma\} \leq \frac{1}{n}$ dunque la dimensione del campione n, influisce.

Al crescere del campione, la probabilità che una realizzazione della media campionaria abbia una distanza apprezzabile dal valore atteso diventa sempre più piccola.

La funzione si nota che non è una Gaussiana sia perchè non è simmetrica sia perchè il valore atteso non capita "nel mezzo" della funzione.



CONVERGENZA IN PROBABILITA'

Nata grazie a disuguaglianza di Chebyshev

Si dice che una generica sequenza di variabili aleatorie:

(non dice se indipendenti ed identicamente distribuite e o altro)

$$Y_1, Y_2, \ldots, Y_n \quad \rightarrow \quad \vartheta \ \ (valore\ finito)$$

(che convergono, tutte
o a meno di poche che si ignorano)

Converge in probabilità ad un valore finito.

e si scrive: $\lim_{n\to\infty} \Pr\left\{ \left| Y_n - \vartheta \right| < \delta \right\} = 1, \quad \forall \delta > 0$ notazione

sarebbe diverso
il lim della sequenza di v.a.

sono diverse

quando: $fissati \quad \varepsilon > 0 \ e \ quindi \ \delta > 0 \quad \exists \ \tilde{n}$ tale che

NON DA N TILDE IN POI!

$\delta \triangleq \delta_\varepsilon$

δ dipende da ε!

ε piccolo a piacere

grazie a Chebyshev

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

probab. che da una v.a. la distanza dal valore finito è minore di delta è molto vicino a 1, quindi quasi certa!

Si osservi che l'affermazione "risulta Pr …" si riferisce ad una sola disuguaglianza e non può essere estesa alla seguente:

sarebbe un AND $\left| Y_{\tilde{n}+1} - \vartheta \right| \cap \left| Y_{\tilde{n}+2} - \vartheta \right| \cap \left| Y_{\tilde{n}+3} - \vartheta \right| \cap \ \cdots$

avranno valori variabili. chi più vicino e chi meno vicino

che esprime un tipo di convergenza detta "forte". Che quindi vale per più variabili aleatorie

Teorema di Bernoulli - Teorema di convergenza debole

Prese $X_1, X_2, \ldots$ variabili aleatorie indipendenti BERNOULLIANE ovvero: $X_i = \begin{cases} 1 & con\ prob = p \\ 0 & con\ prob = 1-p \end{cases}, \quad i = 1,2,\ldots,n$

$B \triangleq \sum_{i=1}^{n} X_i$ è la binomiale con E[B]=np e VAR[B]=np(1-p) allora:

LA BINOMIALE È UNA SOMMA!

(B rappresenta i casi favorevoli, B/n è la media campionaria che rappresenta gli esiti di n prove di Bernoulli ovvero la proporzione di successi su n prove.)

$lim_{n\to\infty} Pr\{|\frac{B}{n} - p| < \delta\} = 1, \forall \delta > 0$ Il limite per n che tende a infinito della probabilità che B/n è abbastanza vicina a p è minore di delta è pari a 1 ovvero:

il rapporto tra casi favorevoli e casi possibili può approssimare la probabilità vera, su un numero infinitamente grande di esperimenti indipendenti.

Prova: parto dalla distanza dal valore atteso espressa rispetto a k deviazioni standard. (quante deviazioni standard sono lontano dal valore atteso?)

$$\Pr\left\{\left| B - E[B]\right| < k\sqrt{Var[B]}\right\} > 1 - \frac{1}{k^2}$$

Visto che non è negativa la binomiale uso:

(Chebyshev-Cantelli)

$$k^2 := 1 + k^2$$

assegnazione non uguale

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

$$\text{con } \varepsilon \hat{=} 1/k^2$$

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

$$\text{con } \delta \hat{=} \sqrt{\frac{p(1-p)}{n\,\varepsilon}}$$

delta dipende da epsilon ed è inversamente proporzionale.
più piccolo epsilon, più è grande delta e viceversa.

Per n infinitamente grande, non si esclude il caso in cui, la differenza $B - np$ fra il numero di successi realizzati e numero di successi attesi (testa che escono e testa che mi aspetto ovvero 50%) diverge (non ho 50% di fare testa) con la velocità della radice di n mentre n cresce linearmente verso l'infinito. In parole povere se B-np cresce alla velocità della radice il limite per n che tende a infinito, di radice di n su n fa 0 e non p la probabilità a cui si vuole convergere. Quindi Bernoulli cattura debolmente la convergenza, diciamo "che non copre la realtà" ovvero che tutto questo non accade. Per n che tende a infinito si ha quel 50% di probabilità. Borel si pone di eliminare questa debolezza e rafforzare la convergenza

IL TEOREMA DI BOREL

da guardare alla fine

Stabilisce un risultato di convergenza "forte":

$$\boxed{\Pr\left\{\lim_{n\to\infty}\left|\frac{B}{n}-p\right|<\delta\right\}=1, \quad \forall\delta>0}$$

Si dice che $\frac{B}{n}\to p$ *"con prob.* 1", per n → ∞

e si intende che $\forall\,\varepsilon>0\; e\; quindi\; \delta>0 \quad \exists\; \tilde{n}, \tilde{n}+1, \tilde{n}+2,\ldots$ tale che

qui debbo dire da n tilde in poi!

$$Prob\left[\left|\frac{B}{\tilde{n}}-p\right|<\delta\cap\left|\frac{B}{\tilde{n}+1}-p\right|<\delta\cap\left|\frac{B}{\tilde{n}+2}-p\right|<\delta\cap\ldots\right]<1-\varepsilon$$

Questo risultato *forte* della legge dei grandi numeri garantisce che a partire da un "certo n" in poi le realizzazioni della variabile aleatoria *B/n* che risulteranno tanto vicine a *p* quanto più si vuole corrisponderanno ad un sottospazio di Ω che ha probabilità 1 (convergenza "*con prob* 1").

quasi sempre negli esperimenti
quasi tutti i risultati elementari
quasi ovunque nello spazio

In questo senso, il verificarsi di realizzazioni apprezzabilmente distanti da *p* è un evento che ha probabilità 1-P(Ω), dunque nulla. Cioè tutte le realizzazioni convergono tranne al più un sottoinsieme dello spazio dei risultati a cui si da probabilità nulla, perchè non si da importanza.

Legge (debole) dei grandi numeri (per la media campionaria)

Che si può dimostrare

La legge dei grandi numeri si applica per somme di variabili aleatoria, ed essendo la media somma di variabili diviso una costante non ci sono problemi nell'applicarla.

Prima ipotesi:

•) $X_1, X_2, \ldots$ var. al. ind., con valore atteso finito e identico, μ e varianze sia pur non identiche ma finite, $\sigma_k^{\,2}, k = 1, \ldots, n$ e tutte limitate da una stessa costante, C.

Oppure seconda ipotesi:

(non indipendenti)

•) $X_1, X_2, \ldots$ var. al. con val. att. identico e finito, μ

Quindi la varianza dovrà essere in gerarchia degli infiniti sotto n quadro così da rispettare la condizione di limite pari a 0

e con $\lim\limits_{n \to \infty} \dfrac{Var\left[\sum_{i=1}^{n} X_i\right]}{n^2} = 0$

non somma delle varianze perchè non sono indipendenti!

<u>OSSERVAZIONE</u>: La prima ipotesi serve sulla sequenza di input al modello P-C (tempi di interarrivo, ad esempio), mentre la seconda serve sulla sequenza di output (tempi di soggiorno, ad esempio).

perché output non sono indipendenti

Legge (debole) dei grandi numeri (per la media campionaria)

TESI comune alle due ipotesi :

Per la v.a. med camp: $\bar{X}(n) \hat{=} (X_1 + \cdots + X_n)/n$ ma anche per somme di v.a.

Il limite significa che dopo molto tempo è certo (=1) che la distanza tra media campionaria (stima del valore atteso) e valore atteso è molto piccola

risulta:

$$\lim_{n\to\infty} Prob\left[\ \left|\bar{X}(n) - E[\bar{X}(n)]\right| < \delta\ \right] = 1, \quad \forall \delta > 0$$

Il problema è che nella realtà non si può aspettare per sempre.

e s'intende che:

$$fissati\ \ \varepsilon > 0\ e\ quindi\ \delta_\varepsilon > 0 \quad \exists\ \ \tilde{n}\ \ tale\ che$$

$$risulta:\ Prob\left[\ \left|\bar{X}(\tilde{n}) - E\left[\bar{X}(\tilde{n})\right]\right| < \delta_\varepsilon\ \right] > 1 - \varepsilon$$

rivediamo il concetto di convergenza in probabilità

Quindi dopo un certo n tilde, la probabilità che la distanza tra stima del valore atteso e valore atteso sia minore di delta arbitrariamente piccolo è arbitrariamente vicino a 1, ovvero alla certezza. Ad esempio dopo 500 tiri di moneta, si sa dalla vita che la probabilità che esca testa è 50%.
La legge ci dice che facendo 500 tiri quasi sicuramente sarai sarà uscito testa circa il 50% dei tiri. Infatti empiricamente dopo 500 tiri si stima che la probabilità è attorno al 50%, magari 49.87%, 48%. Questa convergenza è debole perchè è maggiore di $1-\varepsilon$ non è esattamente 1!

Dimostrazione della legge debole ***(per var. al. correlate)***

(non indipendenti)

A partire dalla seconda ipotesi:

$$VAR\left[\overline{X}(n)\right] = \frac{1}{n^2}\, VAR\left[\sum_{i=1}^{n} X_i\right] \triangleq v^2(n)$$

funzione che rispetta l'ipotesi

La tesi da validare è che valgano:

$$Prob\left[\,\left|\overline{X}(\tilde{n}) - E\left[\overline{X}(\tilde{n})\right]\right| < \delta_\varepsilon\,\right] > 1-\varepsilon$$

$$\lim_{n\to\infty} Prob\left[\,\left|\overline{X}(n) - E[\overline{X}(n)]\right| < \delta\right] = 1, \quad \forall \delta > 0$$

basta riprendere la disuguaglianza di Chebyshev-Cantelli:

essendo $v^2(n)$ sempre positiva

$$\Pr\left\{\left|\,\overline{X}(n) - \mu\,\right| \geq k \cdot v(n)\right\} \leq \frac{1}{k^2}$$

e porre: $\varepsilon \triangleq (1/k^2)$ e $\delta \triangleq k \cdot v(n)$

posto ciò si può riscrivere alla seguente maniera:

$$\Pr\left\{\left|\,\overline{X}(n) - \mu\,\right| < k \cdot v(n) = \delta\right\} > 1 - \left(v^2(n)/\delta^2\right) = 1 - \varepsilon$$

e riconoscere che l'ipotesi fatta:

$$\lim_{n\to\infty} v^2(n) = 0$$

→ serve per validare

$$\lim_{n\to\infty} Prob\left[\,\left|\overline{X}(n) - E[\overline{X}(n)]\right| < \delta\right] = 1, \quad \forall \delta > 0$$

è sufficiente a provare la tesi.

perchè così facendo $\varepsilon = \frac{v^2(n)}{\delta^2} = 0$ quindi $1 - \varepsilon = 1$

Dimostrazione della legge debole ***(per var. indipendenti)***

Per la PRIMA IPOTESI

partendo ancora da Chebyshev

$$\Pr\left\{ \left| \bar{X}(n) - E[\bar{X}(n)] \right| \geq h \right\} \leq \left(VAR[\bar{X}(n)] / h^2 \right) \qquad (h \triangleq n\sqrt{VAR[\bar{X}(n)]})$$

riscrivo il tutto sotto Chebyshev-Cantelli

$$\Pr\left\{ \left| \bar{X}(n) - E[\bar{X}(n)] \right| < h \right\} \geq 1 - \left(Var[\bar{X}(n)] / h^2 \right)$$

indipendenti e e sfruttando l'ipotesi di varianze finite, limitate da una medesima costante C

$$Var\left[\bar{X}(n)\right] = \frac{1}{n^2} \sum_{k=1}^{n} Var[X_k]$$

sommatoria pari a nC

con:

$$Var\left[\bar{X}(n)\right] = \frac{1}{n^2} \sum_{k=1}^{n} \sigma_k^2 < \frac{C}{n}$$

l'ipotesi di costanza di C servirà per il limite.

andando a sostituire si ottiene che:

$$\Pr\left\{ \left| \bar{X}(n) - E[\bar{X}(n)] \right| < h \right\} > 1 - \left(\frac{C}{n \cdot h^2} \right) > 1 - \varepsilon$$

dunque se n tende a infinito il rapporto è 0

ovvero:

$$\lim_{n \to \infty} \Pr\left\{ \left| \bar{X}(n) - E[\bar{X}(n)] \right| < h \right\} = 1, \quad \forall h > 0$$

tollero di meno lo sbaglio di tanto, ma tollero di più lo sbaglio di poco.

Legge <u>forte</u> dei grandi numeri (per la media campionaria)

non dimostrabile

IPOTESI ALTERNATIVE:

•) $X_1, X_2, \ldots \quad v.a.\ ind.\, e\, id.\, distr.,\ con\, media\ \ \mu\ finita$

OPPURE

•) $X_1, X_2, \ldots \quad v.a.\ ind.\ \begin{cases} con\, media\, identica,\, \mu,\ finita \\ e\, non\, ident.\, varianze\ \ \sigma_1^2, \sigma_2^2, \ldots \end{cases}$

$$con \quad \sum_{k=1}^{\infty} \frac{\sigma_k^2}{k^2} < \infty$$

TESI COMUNE: Allora, con: $\bar{X}(n) \hat{=} \left(X_1 + \cdots + X_n \right) / n$ media campionaria

$$Prob\left[\lim_{n \to \infty} \left| \bar{X}(n) - \mu \right| < \delta \right] = 1, \quad \forall \delta > 0$$

probabilità di limite non limite di probabilità!

E s'intende che:

$$\forall \quad \varepsilon > 0\,, \delta_\varepsilon > 0 \quad \exists \ \tilde{n} \mid \ \forall\, m \ \in [1,2,\ldots]$$

$$Prob\Big[\ \left|\bar{X}(\tilde{n}) - E[\bar{X}(\tilde{n})]\right| < \delta_\varepsilon \ \cap \left|\bar{X}(\tilde{n}+1) - E[\bar{X}(\tilde{n}+1)]\right| \ \cap$$

$$\cap \ \left|\bar{X}(n+2) - E[\bar{X}(n+2)]\right| \ \ldots \ < \delta_\varepsilon \ \Big] \ > 1-\varepsilon$$

AND visto nella slide della convergenza in probabilità

Di grande interesse pratico: (si lavora con)

•) $X_1, X_2, \ \ldots \quad v.a.\ ind.\, e\, id.\, distr. \quad con\ \ E\left[X\right] = \mu$

$$ALLORA \quad \bar{X}(n) \triangleq \frac{\sum_{i=1}^{n} X_i}{n} \xrightarrow[\text{converge}]{} \mu \quad \text{"}con\ pr\ 1\text{"} \quad per \quad n \to \infty$$

quasi sempre

quasi ovunque

Applicazione della legge forte

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.

$$Siccome: \quad E[X] = 0 \cdot \Pr(0) + 1 \cdot \Pr(1)$$

$$\Rightarrow \; Prob(A) \equiv E[X]$$

$$allora: \quad \overline{X}(n) \triangleq \frac{\sum_{i=1}^{n} X_i}{n} \xrightarrow{pr\,1} Prob(A), \qquad n \to \infty$$

IMPORTANTE

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 $Prob(A)$.

unico e finito

Seconda applicazione della Legge forte dei G.N. :

$$\Pr\left\{\lim_{t\to\infty}\frac{N(t)}{t}=\frac{1}{E[X]}\hat{=}\lambda\right\}=1$$

Sperimentazione della convergenza forte:

$con:$

$N(t)\hat{=}$ $numero\ di\ arrivi\ in\ (0,t]$

se uso t1, t2, ecc. sarebbe un processo stocastico

$X_1, X_2, \ldots, X_{N(t)}$ $INTERARRIV$ $ind.\,e\ id.\,distr.\ \equiv X$

$S_{N(t)}\hat{=}X_1+X_2+\ \ldots\ +X_{N(t)}$

somma di tempi aleatori. se N(t) fissato ho somma di tempi finiti

PROVA:

t istante tra l'ultimo arrivato e quello che deve arrivare

$$\frac{S_{N(t)}}{N(t)}\le\frac{t}{N(t)}<\frac{S_{N(t)+1}}{N(t)}\quad\Rightarrow\quad\frac{S_{N(t)}}{N(t)}\le\frac{t}{N(t)}<\frac{S_{N(t)+1}}{N(t)+1}\,\frac{N(t)+1}{N(t)}$$

non è lim di prob, ma prob di lim

$$\lim_{t\to\infty}\frac{S_{N(t)}}{N(t)}=\lim_{t\to\infty}\frac{X_1+X_2+\cdots+X_{N(t)}}{N(t)}=E[X]\quad \textit{con probabilità 1}$$

quasi ovunque, quasi sempre

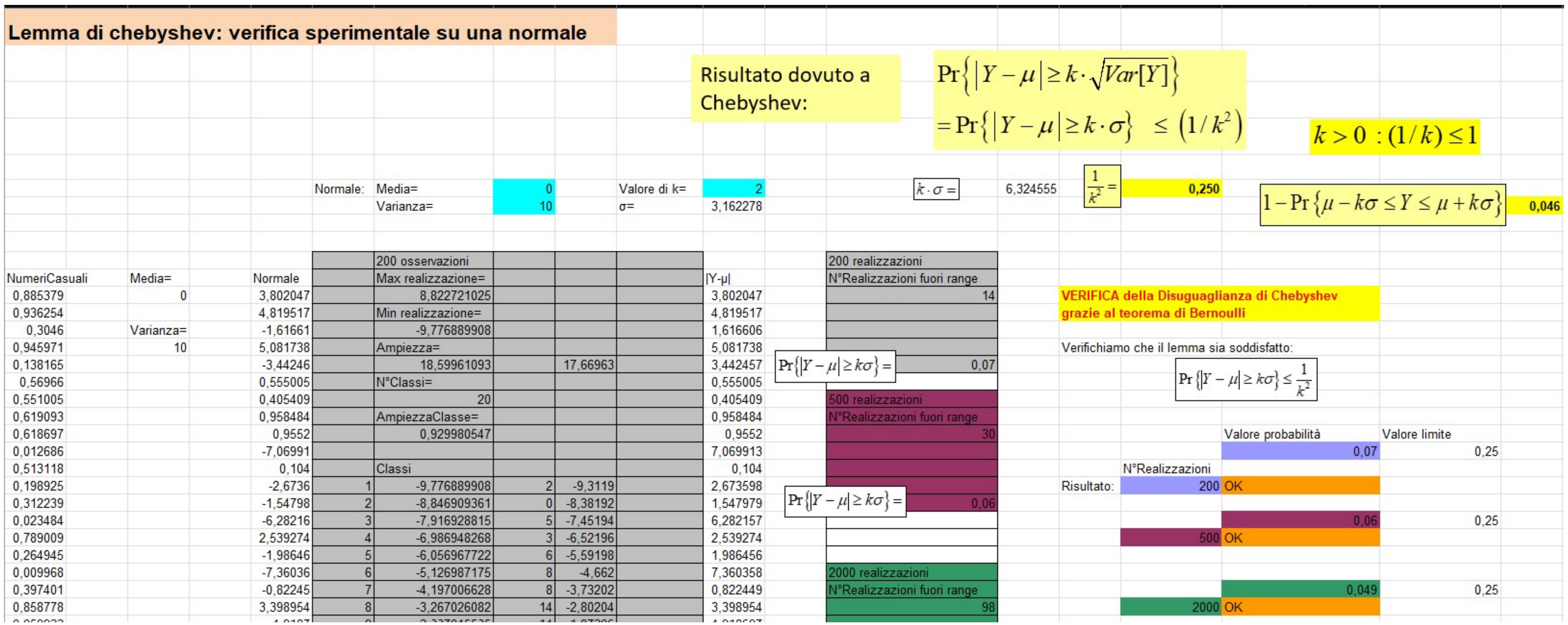
grazie alla legge forte converge al valore atteso con prob 1

$$\lim_{t\to\infty}\left[\frac{S_{N(t)+1}}{N(t)+1}\,\frac{N(t)+1}{N(t)}\right]=E[X]\lim_{t\to\infty}\frac{N(t)+1}{N(t)}=E[X]\quad \textit{con prob 1}$$

tramite teorema del confronto

$$\Rightarrow\quad E[X]\le\lim_{t\to\infty}\frac{t}{N(t)}<E[X]\quad \textit{con probabilità 1}$$

Tramite monte carlo andiamo a provare la disuguaglianza di Chebyshev su una normale con:



Lemma di chebyshev: verifica sperimentale su una normale

Risultato dovuto a Chebyshev:

$$\Pr\left\{\left|Y-\mu\right|\geq k\cdot\sqrt{Var[Y]}\right\} = \Pr\left\{\left|Y-\mu\right|\geq k\cdot\sigma\right\} \leq \left(1/k^2\right)$$

$$k>0:(1/k)\leq 1$$

Normale:	Media=	0	Valore di k=	2	$k\cdot\sigma=$	6,324555
	Varianza=	10	σ=	3,162278		

$\frac{1}{k^2}=$ **0,250**

$1-\Pr\left\{\mu-k\sigma\leq Y\leq\mu+k\sigma\right\}$ **0,046**

NumeriCasuali		Normale		200 osservazioni			|Y-μ|
	Media=			Max realizzazione=			
0,885379	0	3,802047		8,822721025			3,802047
0,936254		4,819517		Min realizzazione=			4,819517
0,3046	Varianza=	-1,61661		-9,776889908			1,616606
0,945971	10	5,081738		Ampiezza=			5,081738
0,138165		-3,44246		18,59961093		17,66963	3,442457
0,56966		0,555005		N°Classi=			0,555005
0,551005		0,405409		20			0,405409
0,619093		0,958484		AmpiezzaClasse=			0,958484
0,618697		0,9552		0,929980547			0,9552
0,012686		-7,06991					7,069913
0,513118		0,104		Classi			0,104
0,198925		-2,6736	1	-9,776889908	2	-9,3119	2,673598
0,312239		-1,54798	2	-8,846909361	0	-8,38192	1,547979
0,023484		-6,28216	3	-7,916928815	5	-7,45194	6,282157
0,789009		2,539274	4	-6,986948268	3	-6,52196	2,539274
0,264945		-1,98646	5	-6,056967722	6	-5,59198	1,986456
0,009968		-7,36036	6	-5,126987175	8	-4,662	7,360358
0,397401		-0,82245	7	-4,197006628	8	-3,73202	0,822449
0,858778		3,398954	8	-3,267026082	14	-2,80204	3,398954

Realizzazioni	N°Realizzazioni fuori range	$\Pr\left\{\left\|Y-\mu\right\|\geq k\sigma\right\}=$
200 realizzazioni	14	0,07
500 realizzazioni	30	0,06
2000 realizzazioni	98	

VERIFICA della Disuguaglianza di Chebyshev grazie al teorema di Bernoulli

Verifichiamo che il lemma sia soddisfatto:

$$\Pr\left\{\left|Y-\mu\right|\geq k\sigma\right\}\leq\frac{1}{k^2}$$

	N°Realizzazioni	Valore probabilità	Valore limite
		0,07	0,25
Risultato:	200	OK	
		0,06	0,25
	500	OK	
		0,049	0,25
	2000	OK	

Media=0
Varianza=10
k=2

Ricordando che k non deve essere intero.

$$\frac{1}{k^2}=0.25$$

L'esperimento funziona perché della normale conosciamo la distribuzione, quindi possiamo calcolare il valore vero!

Perciò con questi parametri Chebyshev afferma che la probabilità che la realizzazione differisca dal valore atteso per 2 volte (k=2) la deviazione standard è minore uguale di 0.25. Ma utilizzando la distribuzione della normale andando a calcolare su Excel 1-(DISTRIB.NORM(mu+ksigma;mu;sigma;VERO)-DISTRIB.NORM(mu-ksigma;mu;sigma;VERO)) si ottiene 0.046! Si deduce che il bound di Chebyshev è lasco! Si vogliono bound sufficientemente stretti altrimenti se troppo stretti si hanno problemi con il bias (ovvero lo spostamento a destra o sinistra rispetto al valore). Si è poi andato a verificare la disugualianza di Chebyshev su 200, 500 e 2000 realizzazioni. Da notare che con varianza 10 con 2000 realizzazioni l'istogramma assomiglia alla normale, ma se aumento le realizzazioni ovviamente la rappresentazione viene più fine. Con una varianza diversa magari servono meno o più osservazioni per rappresentare bene la normale. Su Excel c'è anche la verifica di Chebyshev, o meglio Cantelli per l'esponenziale con una modifica sulla disuguaglianza.

n	p	q
50	0,3	0,7

media
15

varianza
10,5

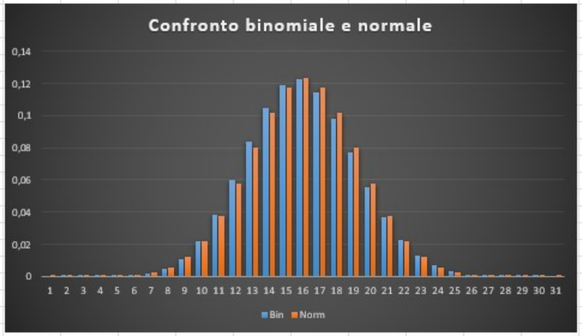
deviazione st.
3,24037

mostro come la binomiale di media n e varianza p tende ad una normale con valore atteso np e varianza np(1-p)

(X~B(n,p))-->(X~N(np,np(1−p)))

se media>=5 e varianza >=5

N(np,npq)
2,73628E-06
1,08868E-05
3,93803E-05
0,000129508
0,000387215
0,001052559
0,002601235
0,005844553
0,011938826
0,022172292
0,037436753
0,057467775
0,080202741
0,101763646
0,11739098
0,123116261
0,11739098
0,101763646
0,080202741
0,057467775
0,037436753
0,022172292
0,011938826
0,005844553
0,002601235
0,001052559
0,000387215
0,000129508
3,93803E-05
1,08868E-05
2,73628E-06



passo 1

k	Coeff.Bin.	B(n,p)
0	1	1,79847E-08
1	50	3,85385E-07
2	1225	4,04655E-06
3	19600	2,77477E-05
4	230300	0,00013973
5	2118760	0,000550934
6	15890700	0,00177086
7	99884400	0,004770481
8	536878650	0,010989144
9	2505433700	0,021978287
10	10272278170	0,038618991
11	37353738800	0,06018544
12	1,214E+11	0,08382972
13	3,54861E+11	0,105017451
14	9,37846E+11	0,118948338
15	2,25083E+12	0,122346862
16	4,92369E+12	0,114700183
17	9,84738E+12	0,098314443
18	1,80535E+13	0,077247062
19	3,04059E+13	0,055757278
20	4,71292E+13	0,037038763
21	6,73274E+13	0,022676794
22	8,87498E+13	0,012810916
23	1,08043E+14	0,006683956
24	1,21549E+14	0,003222622
25	1,26411E+14	0,001436369
26	1,21549E+14	0,00059191
27	1,08043E+14	0,00022549
28	8,87498E+13	7,93815E-05
29	6,73274E+13	2,58088E-05

la poissoniana converge alla gaussiana, la binomiale converge alla gaussiana. Ma comunque la binomiale converge alla poissoniana

n1	p1	q1
20	0,1	0,9

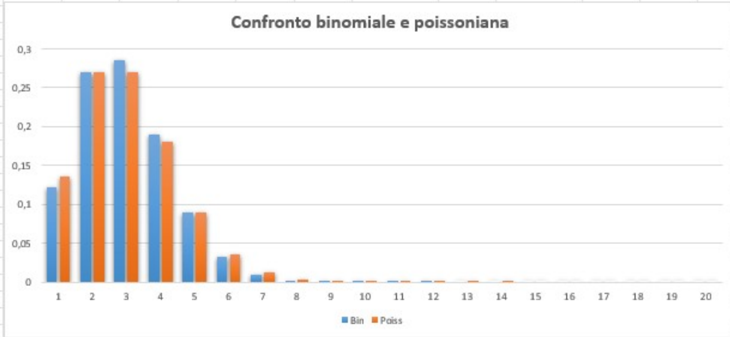
media
2

varianza
1,8

deviazione st.
1,34164

passo 1

k	Coeff.Bin.	B(n,p)	P(λ)
0	1	0,121576655	0,135335283
1	20	0,270170344	0,270670566
2	190	0,285179807	0,270670566
3	1140	0,190119871	0,180447044
4	4845	0,089778828	0,090223522
5	15504	0,031921361	0,036089409
6	38760	0,008867045	0,012029803
7	77520	0,001970454	0,003437087
8	125970	0,000355776	0,000859272
9	167960	5,27076E-05	0,000190949
10	184756	6,44204E-06	3,81899E-05
11	167960	6,50711E-07	6,94361E-06
12	125970	5,4226E-08	1,15727E-06
13	77520	3,70776E-09	1,78041E-07
14	38760	2,05987E-10	2,54345E-08
15	15504	9,15496E-12	3,39126E-09
16	4845	3,1788E-13	4,23908E-10
17	1140	8,3106E-15	4,98715E-11
18	190	1,539E-16	5,54128E-12
19	20	1,8E-18	5,83292E-13



se invece p si avvicina a 1, si può contare il numero di insuccessi anzichè quello dei successi in modo tale che la probabilità di insuccesso 1−p sia piccola e si può ancora usare la distribuzione di Poisson

Confronto tra binomiale e normale.