Variabili aleatorie scalari

Sommario della Esercitazione 2

- Introduzione
- CDF e PDF: definizione
- CDF e PDF: proprietà
- Distribuzioni uniforme e Gaussiana
- Gaussiana: grafici della CDF e PDF
- Gaussiana: calcolo probabilità
- Esercizio

Richiami

- DEFINIZIONE FORMALE: Una funzione $Y: S \rightarrow R$ è una VA se e solo se $P\{s \in S : Y(s) \leq y\}$ esiste per qualunque $y \in R$, ed inoltre $P\{Y=\infty\}=0$ e $P\{Y=-\infty\}=0$

- In altre parole, una VA è una variabile il cui valore non è fissato, ma può assumerne uno qualunque preso da un intervallo di valori con probabilità prefissata

- Y è una funzione mentre i valori assunti da Y, $y=Y(s)$, sono numeri reali

- Le variabili aleatorie le indicheremo con lettere maiuscole (Y, X, …) mentre i risultati con lettere minuscole (y, x, …)

Motivazione

- In questa esercitazione vogliamo rispondere a domande del tipo:
  - qual'è la probabilità che Y sia minore di a?
  - qual'è la probabilità che Y sia compresa tra a e b?

- La probabilità si riferisce ad un evento per cui la frase "…Y minore di a…" oppure "…Y compreso tra a e b…" DEVE corrispondere ad un evento

- Gli eventi li indicheremo come segue: $\{Y < a\}$
  - indicando quel sottoinsieme del dominio S contenente i risultati y_i tali che $Y(y_i)<a$

- ESEMPI:
  - $P\{Y \leq y\}$: probabilità che la VA Y sia minore o uguale ad y
  - $P\{y1 < Y < y2\}$: probabilità che la VA Y sia compresa tra y1 e y2
  - $P\{Y = y\}$: probabilità che la VA Y sia uguale ad y

Esempio

- Consideriamo l'esperimento che consiste nel lancio di una moneta non truccata tre volte consecutive.
- Lo spazio campionario consiste di otto eventi elementari egualmente probabili:

- S={HHH, HHT, HTH, THH, HTT, TTH, THT, TTT }

- Se X è la VA che fornisce il numero di teste (H) uscite determinate P(X=2) e P(X<2)
- A=(X=2)={HHT, HTH, THH} quindi P(X=2)=P(A)=3/8

- B=(X<2)={HTT, TTH, TTH, TTT} quindi P(X<2)=P(B)=1/2

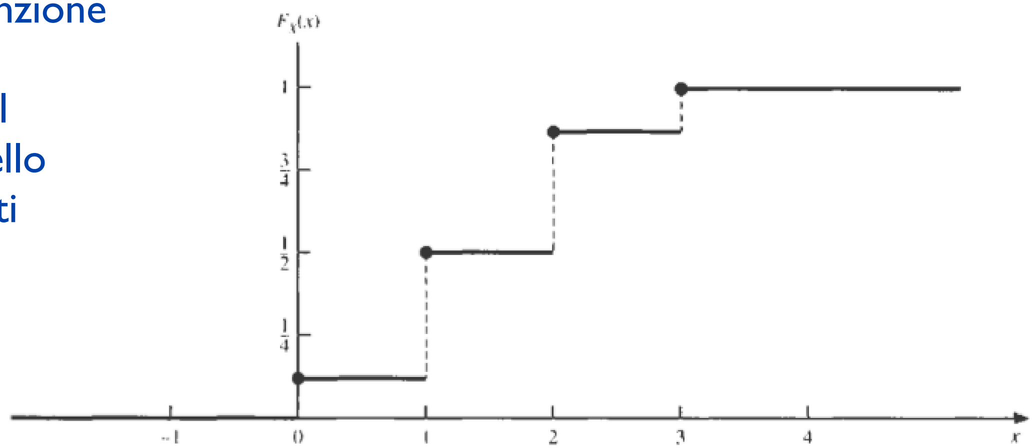
- Determinare il grafico della $F_X(x)$

Esempio

- Costruiamo la tabella di $F_X(x)=P(X \leq x)$ per x=-1,0,1,2,3,4.

x	$(X \leq x)$	$F_X(x)$
-1	$\varnothing$	0
0	(TTT)	$\frac{1}{8}$
1	$\{TTT, TTH, THT, HTT\}$	$\frac{4}{8} = \frac{1}{2}$
2	$\{TTT, TTH, THT, HTT, HHT, HTH, THH\}$	$\frac{7}{8}$
3	S	1
4	S	1

Notate che la funzione ha un salto ad x=0,1,2,3 e che il valore alto è quello corretto in questi punti



CDF e PDF

- Dagli esempi appena considerati è ovvio che $P\{Y \leq y\}$ è una funzione di y

- Questo ci consente di definire la seguente funzione:

$$F_Y(y) = P\{Y \leq y\}$$

- La funzione $F_Y(y)$ si chiama funzione di distribuzione della probabilità oppure funzione cumulativa della distribuzione di probabilità o in forma abbreviata CDF

- Spesso è più comodo utilizzare la derivata della CDF che prende il nome di funzione densità di probabilità, $fY(y)$, o in forma abbreviata PDF
- La PDF è più comoda da usare in quanto in molti casi esistono espressioni analitiche
- E' possibile passare dalla CDF alla PDF attraverso le relazioni:

$$\boxed{F_Y(y) = \int_{-\infty}^{y} f_Y(\xi)\, d\xi} \qquad \boxed{f_Y(y) = \frac{dF_Y(y)}{dy}}$$

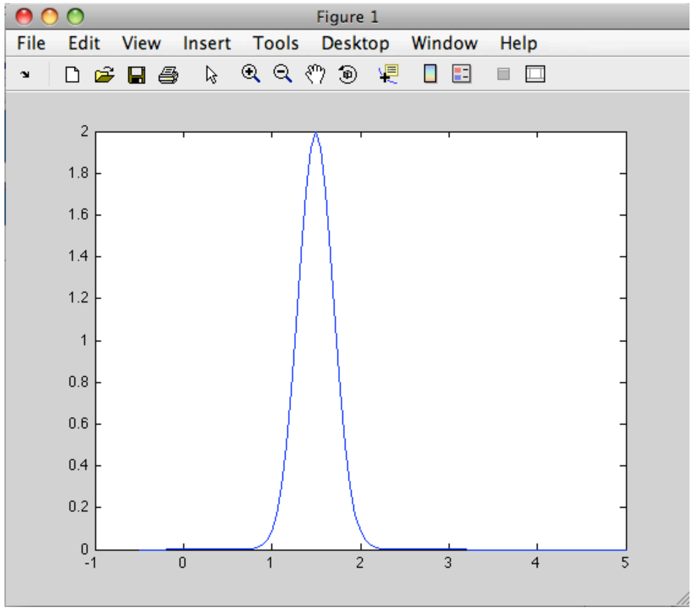
Tipo di VA

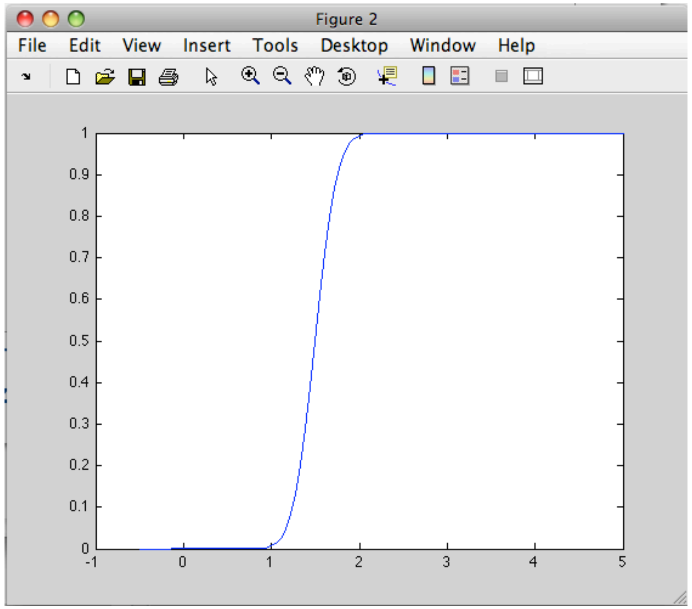
- Una VA si caratterizza a seconda del tipo ovvero a seconda del tipo di CDF o PDF.
- Costruiamo un file per graficare la gaussiana

Operazioni
1) Inserire i parametri della Gaussiana
2) Inserire l'intervallo di ascissa
3) Calcolare la pdf
4) Diagrammare la funzione
5) Calcolare la cdf
6) Diagrammare la funzione

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%                                                                        %%
%%      In questo M-file si diagrammano le funzioni CDF e PDF di una      %%
%%     distribuzione Gaussiana di media mu e deviazione standard sigma    %%
%%                                                                        %%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
clc; % Pulizia dello schermo.
clear all; % Tutte le variabili vengono cancellate dalla memoria.

% L'istruzione "disp" scrive su video il testo tra (' e ').
disp('Generazione del grafico della PDF e CDF di una Gaussiana')

% L'istruzione "input" scrive su video il testo tra (' e ') e resta in
% attesa di un valore da parte dell'utente. Nell'istruzione successiva tale
% valore viene memorizzato nella variabile "mu".
mu = input('Inserire il valore della media:  ')
sigma = input('Inserire il valore della deviazione standard:  ')
y1 = input('Inserire l''estremo inferiore della variabile y:  ')
y2 = input('Inserire l''estremo superiore della variabile y:  ')

% Definizione del vettore y. L'intervallo [y1, y2] viene suddiviso in 100
% parti.
y = linspace(y1, y2, 100);

% Definizione della PDF Gaussiana.
f = 1/(sqrt(2*pi)*sigma)*exp(-0.5*(y-mu).^2/sigma^2);

% Plottaggio della PDF.
plot(y, f)

% Definizione della CDF. "erf" è la funzione "error function" già inclusa
% in Matlab. Si veda l'help per ulteriori dettagli.
F = 0.5*(1+erf((y-mu)/(sqrt(2)*sigma)))

% Il comando "figure" genera una nuova finestra per la successiva figura.
% Ciò evita che la figura precedente sia sostituita da quella successiva.
figure

%Plottaggio della CDF.
plot(y, F);
```

Gaussiane con Matlab





- Provate a stimare a occhio media e varianza
  - Come procedereste?

Gaussiane con Matlab

- Il toolbox statistico di Matlab ha due funzioni predefinite che calcolano direttamente la CDF e la PDF di una Gaussiana:

$\boxed{f = \text{normpdf}(y, mu, sigma)}$

$\boxed{F = \text{normcdf}(y, mu, sigma)}$

PDF e CDF di una Gaussiana di media mu e deviazione standard sigma

programma rifrasato in modo da usare le funzioni predefinite

1) Inserire i parametri della Gaussiana
2) Inserire l'intervallo di ascissa
3) Calcolare la pdf
4) Diagrammare la funzione
5) Calcolare la cdf
6) Diagrammare la funzione

NB se mu e sigma non sono passate la gaussiana è standard

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%                                                                      %%
%%      In questo M-file si diagrammano le funzioni CDF e PDF di una     %%
%%     distribuzione Gaussiana di media mu e deviazione standard sigma  %%
%%     Si utilizzano le funzioni predefinite di Matlab per la PDF e CDF  %%
%%                                                                      %%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
clc; % Pulizia dello schermo.
clear all; % Tutte le variabili vengono cancellate dalla memoria.

disp('Generazione del grafico della PDF e CDF di una Gaussiana')

mu = input('Inserire il valore della media:  ')
sigma = input('Inserire il valore della deviazione standard:  ')
y1 = input('Inserire l''estremo inferiore della variabile y:  ')
y2 = input('Inserire l''estremo superiore della variabile y:  ')

y = linspace(y1, y2, 100);

% Definizione della PDF Gaussiana. Si utilizza la funzione predefinita
% "normpdf"
f = normpdf(y, mu, sigma);

% Plottaggio della PDF.
plot(y, f)

% Definizione della CDF Gaussiana. Si utilizza la funzione predefinita
% "normcdf"
F = normcdf(y, mu, sigma);

figure

%Plottaggio della CDF.
plot(y, F);
```

Probabilità di eventi per VA gaussiane

- Siamo ora interessati a calcolare la probabilità di eventi nell'ipotesi che la VA sia di tipo Gaussiano

- In particolare, vogliamo calcolare la probabilità dei seguenti eventi nell'ipotesi che Y sia una GAUSSIANA STANDARD:

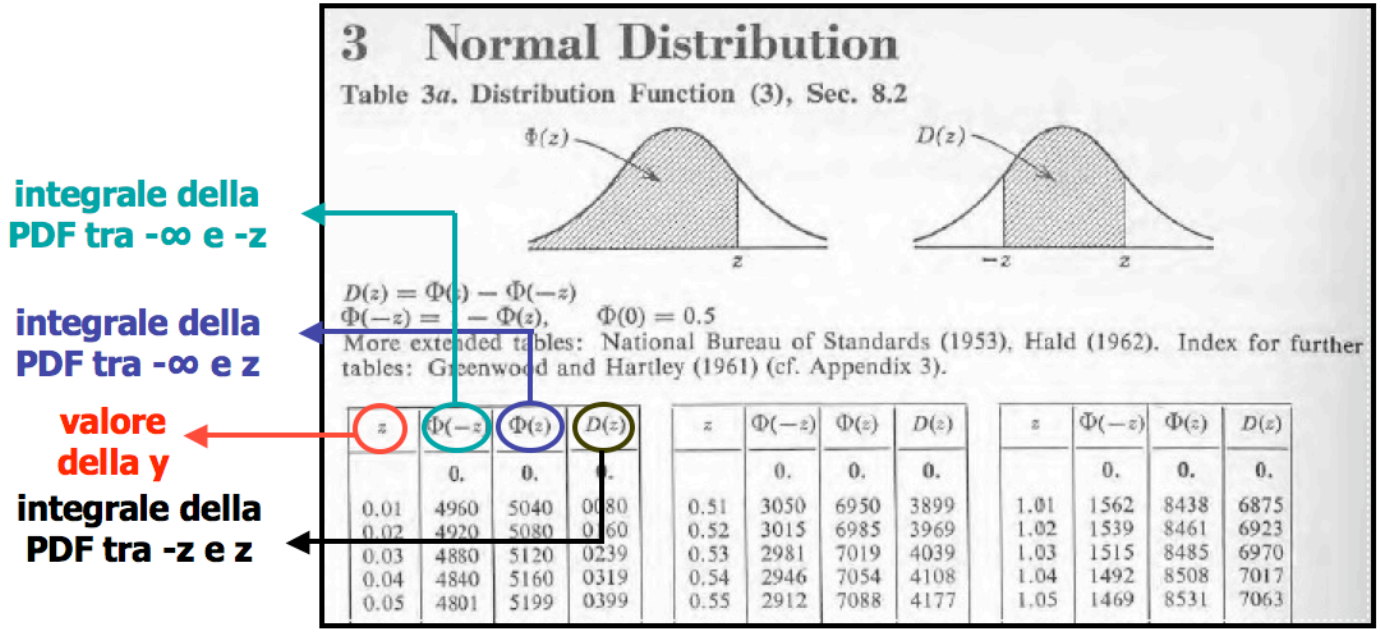
1. $P(Y \leq 2.44)$
2. $P(Y \geq 1)$
3. $P(1.1 \leq Y \leq 2)$

- Inoltre, vogliamo calcolare la probabilità degli stessi eventi nel caso in cui Y sia GAUSSIANA di media 0.8 e varianza 2

Uso delle tabelle

- Quando i calcolatori elettronici non esistevano ancora, si utilizzavano delle tabelle

- Nelle tabelle sono riportati gli integrali della PDF di una distribuzione GAUSSIANA STANDARD, per diversi valori della y (= z nelle tabelle)



3 Normal Distribution

Table 3*a*. Distribution Function (3), Sec. 8.2

$D(z) = \Phi(z) - \Phi(-z)$
$\Phi(-z) = 1 - \Phi(z), \qquad \Phi(0) = 0.5$
More extended tables: National Bureau of Standards (1953), Hald (1962). Index for further tables: Greenwood and Hartley (1961) (cf. Appendix 3).

z	$\Phi(-z)$	$\Phi(z)$	$D(z)$	z	$\Phi(-z)$	$\Phi(z)$	$D(z)$	z	$\Phi(-z)$	$\Phi(z)$	$D(z)$
	0.	0.	0.		0.	0.	0.		0.	0.	0.
0.01	4960	5040	0080	0.51	3050	6950	3899	1.01	1562	8438	6875
0.02	4920	5080	0160	0.52	3015	6985	3969	1.02	1539	8461	6923
0.03	4880	5120	0239	0.53	2981	7019	4039	1.03	1515	8485	6970
0.04	4840	5160	0319	0.54	2946	7054	4108	1.04	1492	8508	7017
0.05	4801	5199	0399	0.55	2912	7088	4177	1.05	1469	8531	7063

Uso delle tabelle

- $P(Y \leq 2.44)$
- Siamo interessati al valore dell'integrale sotto la PDF tra $-\infty$ e $y(=z)=2.44$. E' sufficiente trovare il valore $\Phi(z)$ per $z = 2.44$ ⇨ 0.9927

- $P(Y \geq 1)$
- Siamo interessati al valore dell'integrale sotto la PDF tra $y(=z)=1$ e $+\infty$. E' sufficiente trovare il valore $\Phi(-z)$ per $z = 1$ ⇨ 0.1587

- $P(1.1 \leq Y \leq 2)$
- Siamo interessati al valore dell'integrale sotto la PDF tra $y_1(=z_1)=1.1$ e $y_2(=z_2)=2$. Si ottiene per differenza: $\Phi(z_2) - \Phi(z_1)$ ⇨ 0.1129

Uso di Matlab

- $P(Y \leq 2.44)$
- normcdf(2.44, 0, 1) ⇨ 0.9927

- $P(Y \geq 1)$
- 1 – normcdf(1, 0, 1) (PERCHE'?) ⇨ 0.1587

- $P(1.1 \leq Y \leq 2)$
- normcdf(2, 0, 1) – normcdf(1.1, 0, 1) ⇨ 0.1129

Gaussiana non standard

- Nelle tabelle sono riportati i dati solo per la GAUSSIANA STANDARD

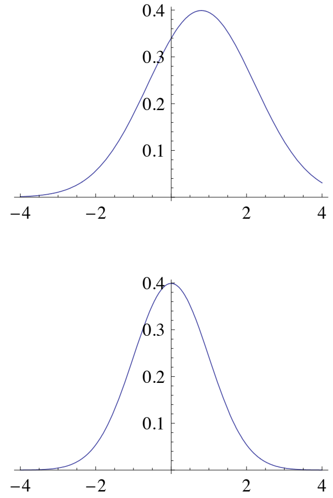
- La seconda parte dell'esercizio è banale con Matlab
- $P(Y \leq 2.44)$
- normcdf(2.44, 0.8, sqrt(2)) ⇨ 0.8769

- $P(Y \geq 1)$
- 1 – normcdf(1, 0.8, sqrt(2)) ⇨ 0.4438

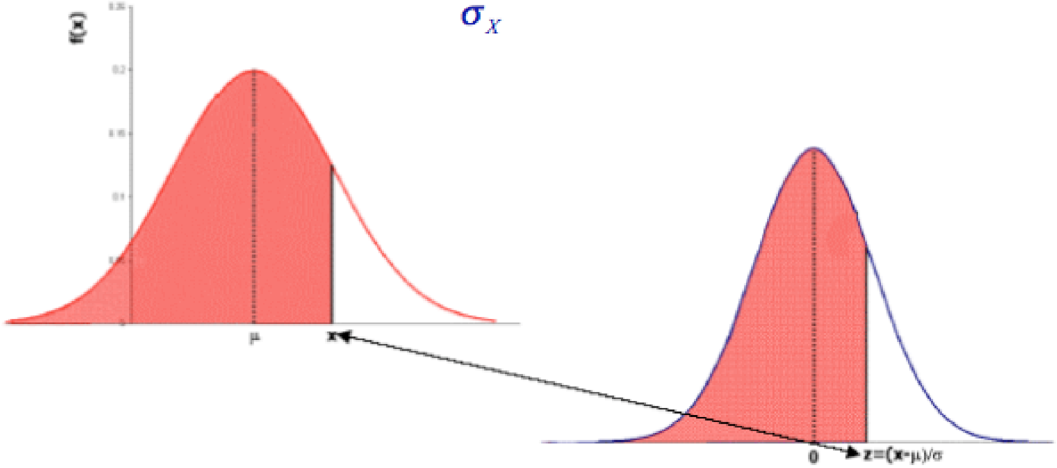
- $P(1.1 \leq Y \leq 2)$
- normcdf(2, 0.8, sqrt(2)) – normcdf(1.1, 0.8, sqrt(2)) ⇨ 0.2179

Gaussiana non standard

- Per usare le tabelle dobbiamo trasformare la nostra gaussiana nella standard
- Quindi verificate con le tabelle con Gaussiana standard
  - P(Z ≤ (2.44-0.8)/sqrt(2)) etc...



$$Z = \frac{X - \mu_X}{\sigma_X}$$



Problemi inversi

- Potremmo dover risolvere anche problemi inversi ovvero problemi in cui è nota la probabilità di un evento

- In particolare, per una Gaussiana standard vogliamo calcolare le costanti c in modo tale che la probabilità dei seguenti eventi sia:

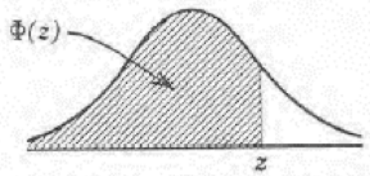
1. $P(Y \geq c) = 0.2$
2. $P(0 \leq Y \leq c) = 0.45$
3. $P(-c \leq Y \leq c) = 0.99$

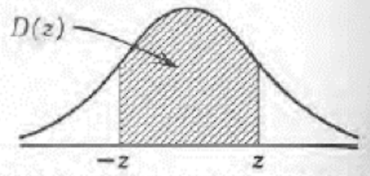
- Inoltre, vogliamo calcolare la costante c nel caso in cui Y sia GAUSSIANA di media -2 e varianza 0.25

Problemi inversi

- Per la Gaussiana standard possiamo usare le tabelle
- Nella tabella 3b sono riportati i valori di z corrispondenti a varie percentuali delle aree tratteggiate in figura (ovvero le probabilità)

Table 3b. Normal Distribution. Values of z for Given Values of (3), Sec. 8.2, and $D(z)$





$D(z) = \Phi(z) - \Phi(-z)$
Example. $\Phi(z) = 61\%$
for $z = 0.279$,
$D(z) = 61\%$ for $z = 0.860$
More extended tables: Comrie (1949), Fisher and Yates (1957), Hald (1962), Kelley (1948) (cf. Appendix 3).

%	$z(\Phi)$	$z(D)$	%	$z(\Phi)$	$z(D)$	%	$z(\Phi)$	$z(D)$
1	−2.326	0.013	41	−0.228	0.539	81	0.878	1.311
2	−2.054	0.025	42	−0.202	0.553	82	0.915	1.341
3	−1.881	0.038	43	−0.176	0.568	83	0.954	1.372
4	−1.751	0.050	44	−0.151	0.583	84	0.994	1.405
5	−1.645	0.063	45	−0.126	0.598	85	1.036	1.440

% dell'area tratteggiata

valore di z per $\Phi(z)$

valore di z per D(z)

Problemi inversi

- Torniamo all'esercizio. Per la Gaussiana standard calcoliamo c tale che:
- $P(Y \geq c) = 0.2$
  - Siamo interessati al valore di c(=z) tale che l'integrale sotto la PDF tra c e $+\infty$ sia uguale a 0.2. E' sufficiente trovare il valore –z tale che % = 20 ⇨ 0.842

- $P(0 \leq Y \leq c) = 0.45$
  - Siamo interessati al valore di c tale dell'integrale sotto la PDF tra 0 e c sia 0.45. E' sufficiente trovare il valore di z tale che D(z) = 2*0.45 = 0.9 = 90% ⇨ 1.645

- $P(-c \leq Y \leq c) = 0.99$
  - Siamo interessati al valore di c tale dell'integrale sotto la PDF tra -c e c sia 0.99. E' sufficiente trovare z(D) per % = 99 ⇨ 2.576

Tabelle

- Dalla tabella appena presentata si ottengono anche importanti informazioni (già viste):

- $P(\mu - \sigma < Y \leq \mu + \sigma) = 68\%$
- $P(\mu - 1.96\,\sigma < Y \leq \mu + 1.96\,\sigma) = 95\%$
- $P(\mu - 2\,\sigma < Y \leq \mu + 2\,\sigma) = 95.5\%$
- $P(\mu - 2.58\,\sigma < Y \leq \mu + 2.58\,\sigma) = 99\%$
- $P(\mu - 3\,\sigma < Y \leq \mu + 3\,\sigma) = 99.7\%$
- $P(\mu - 3.29\,\sigma < Y \leq \mu + 3.29\,\sigma) = 99.9\%$

Problema inverso con Matlab

- Risolviamo il problema con Matlab

- In linea di principio dovremmo risolvere un integrale (per la PDF) o un'equazione non lineare (per la CDF)

- Per esempio, il calcolo di c per P(Y ≥ c)=0.2 si effettua risolvendo l'integrale:

$$\int_c^{+\infty} f_Y(y)dy = \int_c^{+\infty} \frac{1}{\sqrt{2\pi}\sigma_Y} exp\left[-\frac{1}{2}\frac{(y-\mu_Y)^2}{\sigma_Y^2}\right] dy = 0.2$$

- oppure l'equazione: $F_Y(-c) = \frac{1}{2}\left[1 + Erf\left(\frac{-c-\mu_Y}{\sqrt{2}\sigma_Y}\right)\right] = 0.2$

- Si noti che in entrambi i casi l'incognita è c

Problema inverso con Matlab

- Matlab mette a disposizione anche in questo caso alcune funzioni predefinite che risolvono direttamente l'equazione per la CDF appena vista (la funzione si applica a qualunque Gaussiana)

- Il comando da utilizzare è: *norminv(P, mu, sigma)*

  - dove P la probabilità di un evento

1. Quindi, per calcolare c tale che $P(Y \geq c) = 0.2$: *norminv(0.8, 0, 1)*

2. Per calcolare c tale che $P(0 \leq Y \leq c) = 0.45$:
   norminv(0.95, 0, 1) – norminv(0.5, 0, 1)

3. Per calcolare $P(-c \leq Y \leq c) = 0.99$: *norminv(0.99+0.005, 0, 1)*

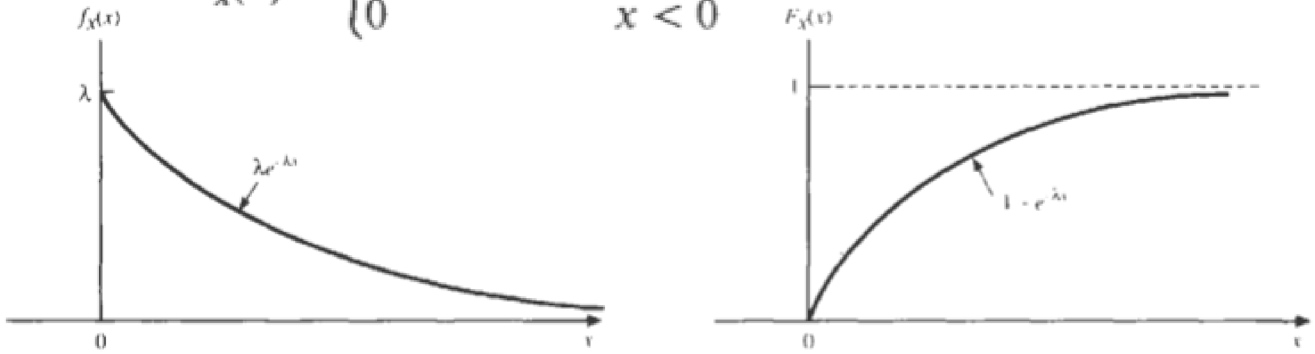
Esercizio 1

- Sia Y una VA esponenziale caratterizzata da una PDF $f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & x < 0 \end{cases}$

- Verificate che la funzione in esame può essere una PDF

- Calcolate la CDF corrispondente

$$F_X(x) = \begin{cases} 1 - e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$



(a) (b)

Esercizio I

- Calcolate la media e la varianza di questa distribuzione

$$\mu_X = E(X) = \frac{1}{\lambda}$$

$$\sigma_X{}^2 = \mathrm{Var}(X) = \frac{1}{\lambda^2}$$

Esercizio II

- Una linea di produzione produce resistori da 1000Ω con una tolleranza del 10%. Sia X la resistenza di un resistore. Ipotizzando che X si a una VA Gaussiana con media 1000 e varianza 2500, determinate la probabilità che un resistore preso a caso sia fuori specifica.

- Sia A l'evento resistore fuori specifica A={X<900}∪{X>1100}.
- Dal momento che {X<900}∩{X>1100}=∅ si ha:

$$P(A) = P(X < 900) + P(X > 1100) = F_X(900) + [1 - F_X(1100)]$$

- Dalle tavole

$$F_X(900) = \Phi\left(\frac{900 - 1000}{50}\right) = \Phi(-2) = 1 - \Phi(2)$$

$$F_X(1100) = \Phi\left(\frac{1100 - 1000}{50}\right) = \Phi(2)$$

$$P(A) = 2[1 - \Phi(2)] \approx 0.045$$