Trattamento delle Osservazioni
ESERCITAZIONE

Esercizio 1

Sia data la seguente funzione:

$$f(x) = \begin{cases} k(4x - x^2) & 0 \le x \le 4 \\ 0 & x \le 0; x > 4 \end{cases}$$

Determinare k in omdo tale che f(x) possa essere considerata una funzione densità di probabilità e ricavare poi la funzione di distribuzione corrispondente. Calcolare inoltre la P(2≤x≤3).

____

Affinché f(x) possa essere considerata una funzione densità di probabilità è necessario che l'integrale dei valori assunti nel dominio reale sia pari all'unità.

$$\int f(x)dx = 1$$

Ponendo questa condizione si ricava il valore di k:

$$\int_{-\infty}^{+\infty} k(4x - x^2)dx = \int_0^4 k(4x - x^2)dx = 1$$

$$k\left|4 \cdot \frac{x^2}{2} - \frac{x^3}{3}\right|_{0\to 4} = 1 \quad \Rightarrow \quad k\left(32 - \frac{64}{3}\right) = 1 \quad \Rightarrow \quad k = \frac{3}{32}$$

La funzione densità di probabilità assume la forma:

$$f(x) = \begin{cases} \frac{3}{32}(4x - x^2) & 0 \le x \le 4 \\ 0 & x \le 0; x > 4 \end{cases}$$

La rappresentazione seguente mostra f(x) e la funzione cumulata ottenuta per integrazione della f(x).

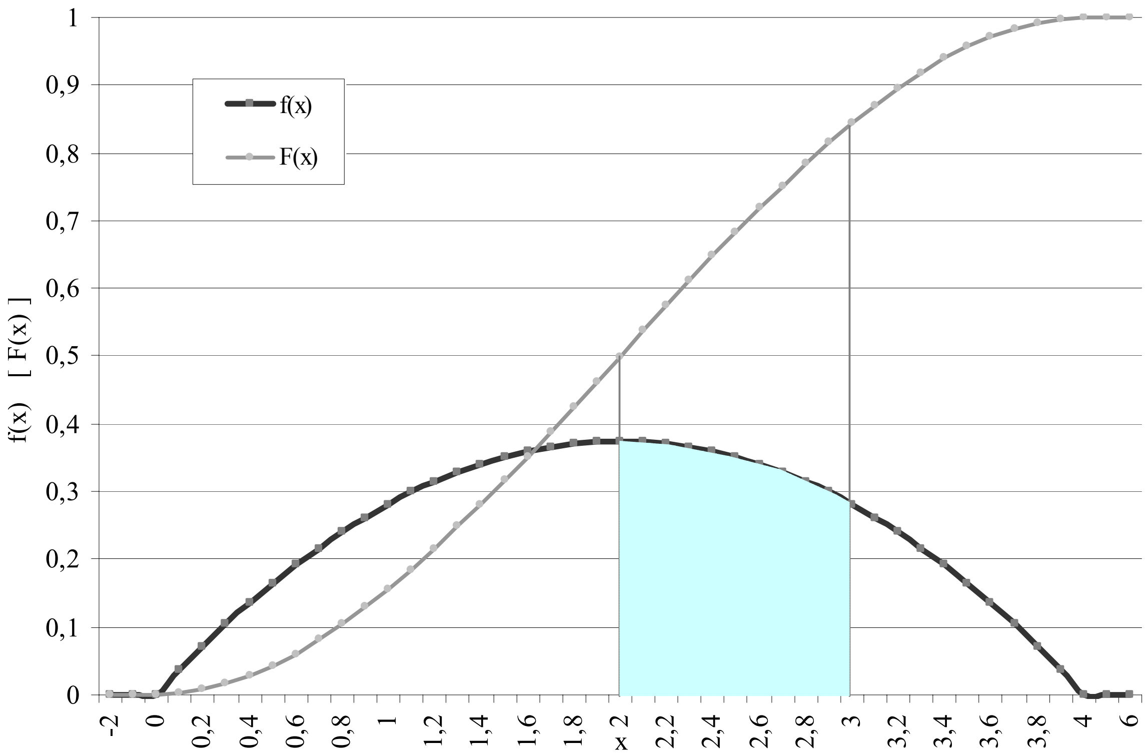


Figura 1 - *Funzione densità e funzione cumulata di probabilità*

E’ evidente, data la definizione di f(x), che:

Per x<0

$$F(x) = \int_{-\infty}^{x} f(z) \cdot dz = 0$$

Per x>4

$$F(x) = \int_{-\infty}^{x} f(z) \cdot dz = 1$$

Per $0 \le x \le 4$

$$F(x) = \int_{-\infty}^{x} f(z) \cdot dz = \frac{3}{32} \left| \frac{4z^2}{2} - \frac{z^3}{3} \right|_{0 \to x} = \frac{1}{32}\left(6x^2 - x^3\right)$$

Per rispondere alla seconda domanda è necessario calcolare l'area sottesa dalla funzione f(x) tra i valori 2 e 3.

$$P(2 \le x \le 3) = \int_{2}^{3} f(x) \cdot dx = \int_{2}^{3} \frac{3}{32}(4x - x^2) \cdot dx = \frac{3}{32} \left| \frac{4x^2}{2} - \frac{x^3}{3} \right|_{2 \to 3} = F(3) - F(2) = \frac{11}{32} = 0.3438$$

Esercizio 2

Sia data la variabile statistica semplice riportata in tabella:

-0,38	-0,03	-0,28	-0,31	-0,24
2,61	-0,50	-0,95	-0,19	-1,22
-1,20	-1,81	0,01	0,40	1,09
1,09	0,73	-0,56	-1,35	0,00
-0,87	1,11	0,15	-0,35	1,42

Calcolare media, varianza e gli indici di Skewness β (asimmetria) e di curtosi.

Tracciare l'istogramma della variabile statistica standardizzata e confrontare le ordinate con i valori corrispondenti della v.c. normale standardizzata.
Ricavare la funzione cumulata di frequenza della v.s. e la funzione di distribuzione della v.c. normale standardizzata corrispondente.

La media del campione si calcola con la formula

$$m_x = \sum_i x_i \cdot f_i = \sum_i x_i \cdot \frac{1}{N} = \sum_1^{25} x_i \cdot \frac{1}{25} = -0{,}274$$

La varianza della popolazione con la

$$s^2 = \sum_i (x_i - m_x)^2 \cdot f_i = \sum_{i=1}^{25} (x_i - m_x)^2 \cdot \frac{1}{25} = 0{,}8825$$

L'indice di asimmetria o Skewness è espresso dalla

$$\beta = \frac{\sum_i (x_i - m_x)^3 \cdot f_i}{s^3} = \sum_{i=1}^{25} \frac{\sum_i (x_i - m_x)^3}{(\sqrt{s^2})^3} \cdot \frac{1}{25} = -0{,}2786$$

L'indice di asimmetria curtosi

$$\gamma = \frac{\sum_i (x_i - m_x)^4 \cdot f_i}{s^4} = \sum_{i=1}^{25} \frac{\sum_i (x_i - m_x)^4}{(\sqrt{s^2})^4} \cdot \frac{1}{25} = 3{,}0132$$

Per effettuare analisi di tipo statistico diffuso sul campione preso in esame è conveniente procedere ad una classificazione dei valori assunti dalla v.c.
I dati possono essere classificati in modi diversi, teoricamente infiniti. Il modo più semplice di procedere prevede la suddivisione in classi entro un intervallo rappresentato dal range (o intervallo di variazione):

range = $x_{max} - x_{min}$ = 1,42 – (-2,61) = 4,03

Il valore ottenuto permette di scegliere agevolmente il numero e quindi l'ampiezza degli intervalli di classificazione, che nella fattispecie valgono:

n_{int} = 5 ➔ amp_{int} = 4,03/5 = 0,806

Gli estremi delle classi sono calcolati di seguito:
x_0= min = -2,61
$x_1 = x_0 + amp_{int}$ = -1,804
x_2 = -0,998
x_3 = -0,192
x_4 = 0,614
x_5 = 1,420
e riportati in tabella, con le frequenze assolute e le frequenze relative:

Classe	Estremo inf.	Estremo sup.	f_{ass}	$f=f_{ass}/N$
1	-2,61	-1,804	2	0,08
2	-1,804	-0,998	3	0,12
3	-0,998	-0,192	9	0,36
4	-0,192	0,614	6	0,24
5	0,614	1,42	5	0,20

La variabile standardizzata assume particolari valori in alcuni indici statistici, in particolare ha media nulla e varianza unitaria.
Standardizzare il campione significa renderne i valori argomentali tali da rappresentare

l'estrazione di una v.c. standard; in pratica ad ogni valore si sottrae la media e si divide la differenza per lo s.q.m.

$$z = \frac{x - m_x}{s} \qquad amp_{\text{int}}(z) = \frac{x_{\max} - x_{\min}}{N \cdot s}$$

Anziché standardizzare ogni valore del campione si procede alla standardizzazione della classificazione effettuata precedentemente:

$$amp_{\text{int}}(z) = \frac{0{,}806}{0{,}939415} = 0{,}8580$$

$$z_0 = \frac{-2{,}61 + 0{,}274}{0{,}939415} = -2{,}487 \qquad z_1 = z_0 + amp_{\text{int}}(z) = -1{,}629$$

z_2 = -0,771 z_3 = 0,087 z_4 = 0,945 z_5 = 1,803

La funzione densità di probabilità standardizzata assume i valori puntuali seguenti:

$$Y = f(z) = \frac{f(x)}{amp_{\text{int}}(z)}$$

$$Y_1 = \frac{0{,}08}{0{,}8580} = 0{,}0932 \qquad Y_2 = \frac{0{,}12}{0{,}8580} = 0{,}1398 \qquad Y_{31} = \frac{0{,}36}{0{,}8580} = 0{,}4195$$

$$Y_4 = \frac{0{,}24}{0{,}8580} = 0{,}2797 \qquad Y_{51} = \frac{0{,}20}{0{,}8580} = 0{,}2331$$

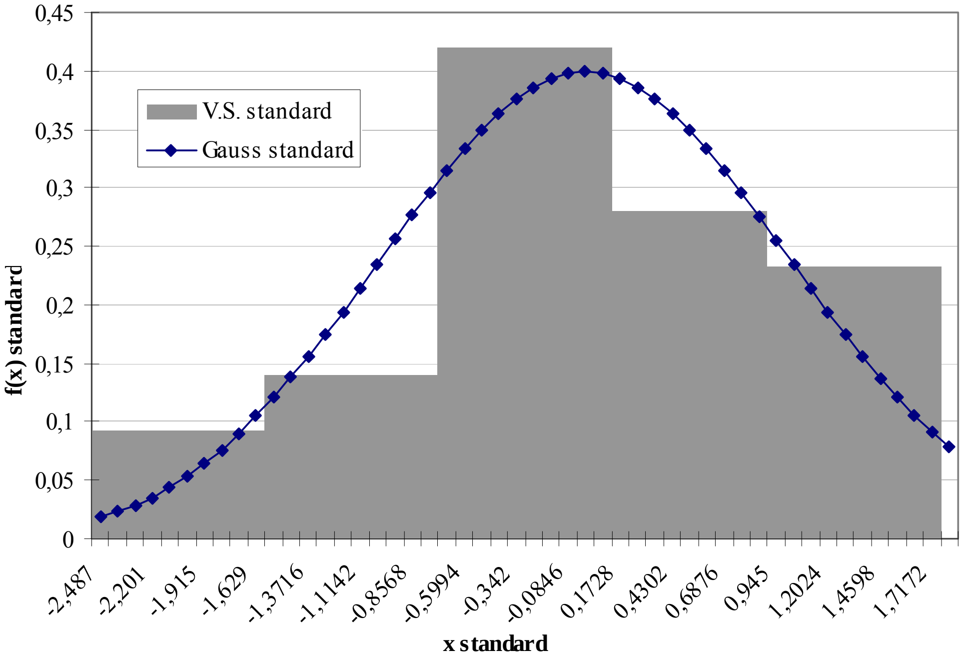
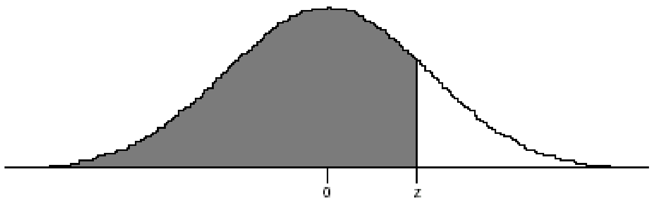


Figura 2 - *Confronto tra le densità della V.C. standardizzata e della pdf Gaussiana standard*

La determinazione delle cumulate della distribuzione normale standardizzata viene effettuata per ciascuna classe ricorrendo alla tabella dell'integrale di probabilità della distribuzione normale standardizzata, che riporta i valori per intervalli discreti della variabile z:

$$f(z) = \frac{e^{-\frac{z^2}{2}}}{\sqrt{2\pi}}$$



Normal Deviate z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-4.0	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
-3.9	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
-3.8	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
-3.7	.0001	.0001	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
-3.6	.0002	.0002	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001
-3.5	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002
-3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
-3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
-3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
-3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
-3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
-2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
-2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
-2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
-2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
-2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
-2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
-2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
-2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
-2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
-2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
-1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
-1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
-1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
-1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
-1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
-1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
-1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
-1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
-1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
-1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
-.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
-.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
-.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
-.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
-.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
-.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
-.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483

$P(1) = P(z < -1.629) - P(z < -2.487) = 0.0516 - 0.0064 = 0.0452$

$P(2) = P(z < -0.771) - P(z < -1.629) = 0.2206 - 0.0516 = 0.1690$

$P(3) = 0.3153$

$P(4) = 0.2930$

$P(5) = 0.1352$

Esercizio 3

Sono stati misurati i diametri di 20 sferette in uscita da una linea produttiva: le misure in cm sono le seguenti:

2.08	1.72	1.92	1.95	1.89	1.85	1.80	1.84	1.82	1.84
1.93	1.86	2.00	1.80	1.82	2.08	1.90	1.85	2.02	2.00

Per poter effettuare un'analisi statistica sui dati è necessario procedere alla loro classificazione.

Si calcola quindi il **range** (o intervallo di variazione):

$x_{max} - x_{min} =$ 2,08 – 1,72 = 0,36 cm

e si sceglie convenientemente di suddividere il range in 6 classi di ampiezza 0,06 cm.

Ne risulta la seguente tabella:

Classe n°	Intervallo	Freq. assoluta	Freq. relativa	Freq.
1	1,72 < x < 1,78	1	0,05	0,05
2	1,79 < x < 1,84	6	0,30	0,35
3	1,85 < x < 1,90	5	0,25	0,60
4	1,91 < x < 1,96	3	0,15	0,75
5	1,97 < x < 2,02	3	0,15	0,90
6	2,03 < x < 2,08	2	0,10	1,00
	TOTALE	20	1,00	

La rappresentazione grafica mostra la distribuzione delle frequenze relative e cumulate (per le variabili discrete è utilizzato il metodo dell'istogramma):

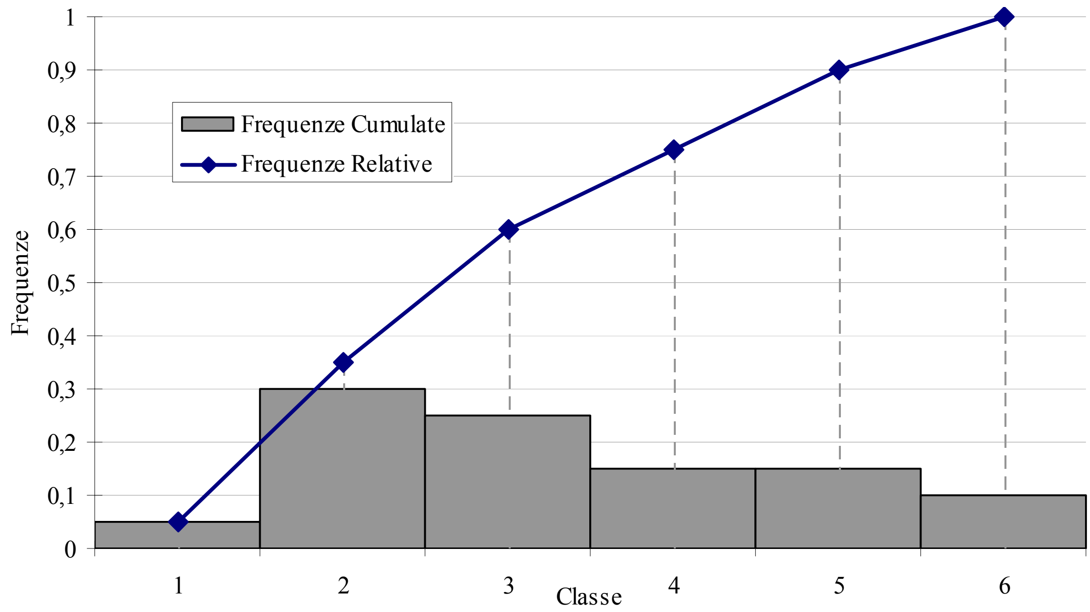


Figura 3 - *Istogramma delle frequenze relative e cumulate*

Il modo di classificare il dato influisce pesantemente sull'esito dell'analisi statistica in quanto ciascuna classificazione mette in evidenza tendenze singolari delle serie di dati.

Una classificazione troppo segmentata rischia di rendere poco sintetica e quindi illeggibile la rappresentazione della fenomeno analizzato mentre una suddivisione in classi molto ampie assorbe le singolarità evidenziando comportamento "medio" del fenomeno stesso.

Classificazione alternativa (2): intervallo più ampio del range, con 4 classi di ampiezza 0,10 cm. Si ottengono i valori della tabella seguente:

Classe n°	Intervallo	Freq. assoluta	Freq. relativa	Freq.
1	$1,70 < x < 1,80$	3	0,15	0,15
2	$1,81 < x < 1,90$	9	0,45	0,60
3	$1,91 < x < 2,00$	5	0,25	0,85
4	$2,01 < x < 2,10$	3	0,15	1,00
	TOTALE	20	1,00	

La rappresentazione fornisce informazioni meno dettagliate sulla serie di dati

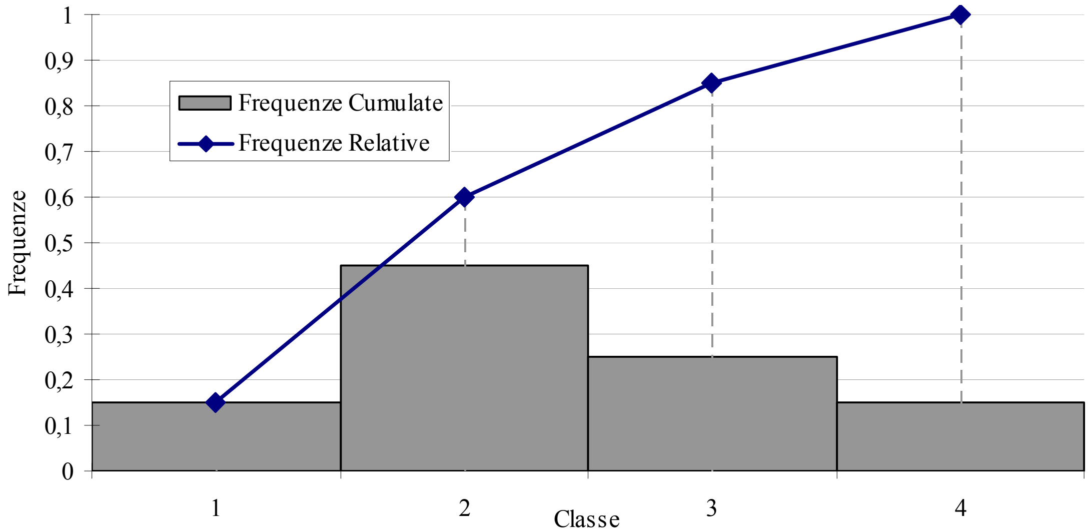


Figura 4 – *Istogramma delle frequenze relative e cumulata per la seconda classificazione*

Classificazione alternativa (3): intervallo più ampio del range, con 10 classi di ampiezza 0,04 cm. Si ottengono i valori della tabella seguente:

Classe n°	Intervallo	Freq. assoluta	Freq. relativa	Freq.
1	1,70 < x < 1,74	1	0,05	0,05
2	1,75 < x < 1,78	0	0	0,05
3	1,79 < x < 1,82	4	0,20	0,25
4	1,83 < x < 1,86	5	0,25	0,50
5	1,87 < x < 1,90	2	0,10	0,60
6	1,91 < x < 1,94	2	0,10	0,70
7	1,95 < x < 1,98	1	0,05	0,75
8	1,99 < x < 2,02	3	0,15	0,90
9	2,03 < x < 2,06	0	0	0,90
10	2,07 < x < 2,10	2	0,10	1,00
	TOTALE	20	1,00	

La rappresentazione è poco significativa:

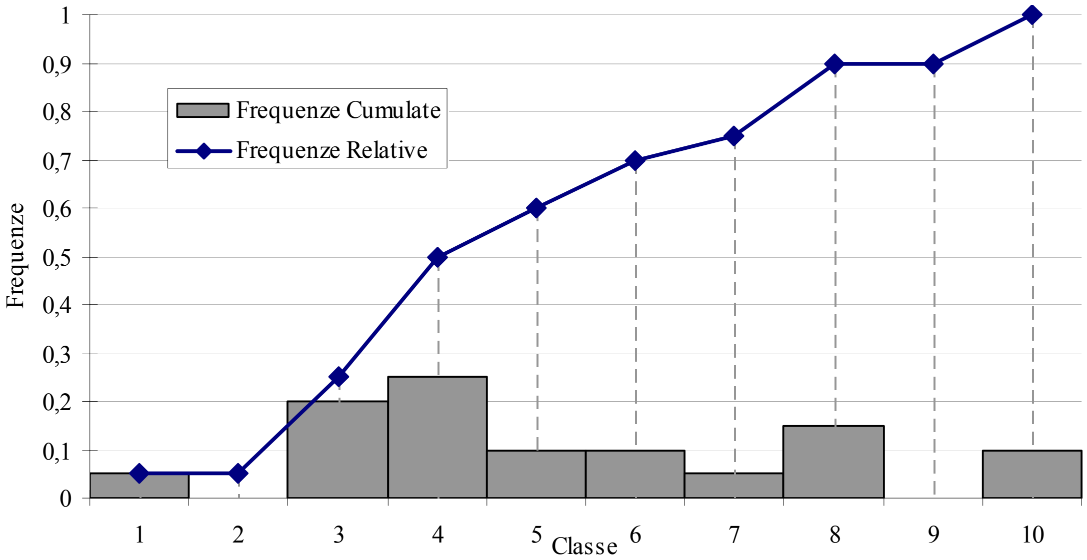


Figura 5 - *Istogramma delle frequenze relative e cumulata per la classificazione (3)*

Esercizio 4

(Tratto da "Lezioni di topografia", A. Manzino)

Di un punto P si sono misurate la distanza dall'origine $\underline{\rho}$ e l'anomalia $\underline{\vartheta}$, rappresentate da variabili casuali ρ e ϑ, con media e sqm seguenti:

$\underline{\rho}$= 1 km $\qquad \underline{\sigma_\rho}$= ±1 mm $\qquad$ ($\underline{\rho}$= 1000 mm)

$\underline{\vartheta}$= π/6 $\qquad \underline{\sigma_\vartheta}$= ±2 • 10^{-6} rad

Calcolare media e covarianza delle coordinate (x,y) del punto P e media e varianza dell'area A del rettangolo che ha <u>OP</u> per diagonale.

____

La trasformazione **g** permette di ricavare (x,y) in funzione delle misure dirette ($\underline{\rho}, \underline{\vartheta}$).

Formalizzando le scritture in forma matriciale:

$$\eta = \begin{bmatrix} x \\ y \end{bmatrix} \qquad \xi = \begin{bmatrix} \rho \\ \theta \end{bmatrix} \qquad C_{\xi\xi} = \begin{bmatrix} \sigma_\rho^2 & 0 \\ 0 & \sigma_\theta^2 \end{bmatrix} = \begin{bmatrix} 1mm^2 & 0 \\ 0 & 4\cdot 10^{-12} \end{bmatrix}$$

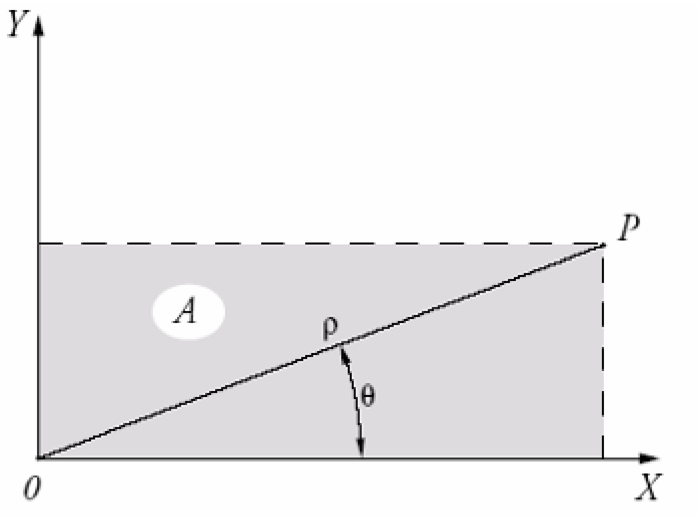


Figura 6 - *Rappresentazione dello schema di misura e del SDR*

Dai valori medi delle variabili, noti, sfruttando il teorema della media si ricavano i valori medi delle incognite:

$$\eta = g\begin{bmatrix} \underline{\rho} \\ \underline{\xi} \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \underline{\rho}\cdot\cos\underline{\theta} \\ \underline{\rho}\cdot\sin\underline{\theta} \end{bmatrix} = \begin{bmatrix} 866{,}025m \\ 500{,}00m \end{bmatrix}$$

Si ricava poi la matrice disegno, calcolando i valori numerici delle derivate parziali di **g** nei valori medi:

$$A = \left[\frac{\partial g}{\partial \xi}\right]_{(\underline{\rho},\underline{\theta})} = \begin{bmatrix} \cos\underline{\theta} & -\underline{\rho}\cdot\sin\underline{\theta} \\ \sin\underline{\theta} & \underline{\rho}\cdot\cos\underline{\theta} \end{bmatrix} = \begin{bmatrix} \sqrt{3}/2 & -10^6/2\ mm \\ 1/2 & 10^6\cdot\sqrt{3}/2\ mm \end{bmatrix}$$

E applicando il teorema di propagazione della varianza:

$$C_{\eta\eta} = A\cdot C_{\xi\xi}\cdot A^T = \begin{bmatrix} \sqrt{3}/2 & -10^6/2\ mm \\ 1/2 & 10^6\cdot\sqrt{3}/2\ mm \end{bmatrix}\cdot\begin{bmatrix} 1mm^2 & 0 \\ 0 & 4\cdot10^{-12} \end{bmatrix}\cdot\begin{bmatrix} \sqrt{3}/2 & 1/2 \\ -10^6/2\ mm & 10^6\cdot\sqrt{3}/2\ mm \end{bmatrix} =$$

$$= \begin{bmatrix} 1{,}75mm^2 & -1{,}30mm^2 \\ -1{,}30mm^2 & 3{,}25mm^2 \end{bmatrix} = \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{bmatrix}$$

Per rispondere alla seconda domanda si calcola il valore media della superficie A, funzione delle misure $(\underline{\rho},\underline{\vartheta})$ (si semplifica la scrittura con le formule di duplicazione) :

$$A = \underline{\rho}^2\cdot\sin\underline{\theta}\cdot\cos\underline{\theta} = \frac{\underline{\rho}^2}{2}\cdot\sin 2\underline{\theta} = 0{,}433\cdot10^6 m^2$$

Ed applicando il principio della propagazione delle varianza:

$$\sigma_A^2 = \left(\frac{\partial A}{\partial\rho}\right)^2\cdot\sigma_\rho^2 + \left(\frac{\partial A}{\partial\theta}\right)^2\cdot\sigma_\theta^2 = (\underline{\rho}\cdot\sin 2\underline{\theta})^2\cdot\sigma_\rho^2 + (\underline{\rho}^2\cdot\cos 2\underline{\theta})^2\cdot\sigma_\theta^2 = ... = 1{,}750m^2$$

$$\sigma_A = \pm 1{,}323m$$

Esercizio da svolgere autonomamente:

. Calcolare media e varianza del perimetro del rettangolo che ha <u>OP</u> come diagonale.

. Calcolare il coefficiente di correlazione tra Area e Perimetro.