

Statistica II - Esercitazione 0

Valore atteso e varianza

Sia $(\Omega, \mathcal{A}, \mathbb{P})$ uno spazio di probabilità e X una variabile aleatoria definita su tale spazio. Il valore atteso della variabile casuale X , indicato con $\mathbb{E}[X]$ (o speranza matematica) è un indice utilizzato per descrivere la posizione di un fenomeno aleatorio. Sia $f(x)$ la funzione di densità della variabile casuale X , funzione di probabilità $p(x)$ nel caso discreto, il valore atteso è definito come:

$$\mathbb{E}[X] := \int_{-\infty}^{+\infty} xf(x)dx$$

$$\mathbb{E}[X] := \sum_{x=0}^{+\infty} xp(x).$$

La varianza della variabile casuale X , indicata con $Var[X]$, è un indicatore atto a descrivere la dispersione di un fenomeno aleatorio. La varianza è definita come:

$$Var[X] := \int_{-\infty}^{+\infty} (x - \mathbb{E}[X])^2 f(x)dx \quad \text{caso continuo } (\Omega = \mathbb{R})$$

$$Var[X] := \sum_{x=0}^{+\infty} (x - \mathbb{E}[X])^2 p(x) \quad \text{caso discreto } (\Omega = \mathbb{N}).$$

La varianza può inoltre essere vista (utile per rendere agevoli i calcoli) come differenza tra il momento secondo ed il quadrato del momento primo della distribuzione considerata:

$$Var[X] = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2.$$

Trasformazioni lineari

Sia $(\Omega, \mathcal{A}, \mathbb{P})$ uno spazio di probabilità e X una variabile aleatoria (continua o discreta) definita su tale spazio. Sia Y una variabile casuale definita come trasformazione lineare della variabile X :

$$Y = aX + b.$$

Valgono i seguenti risultati:

$$\mathbb{E}[Y] = \mathbb{E}[aX + b] = a\mathbb{E}[X] + b$$

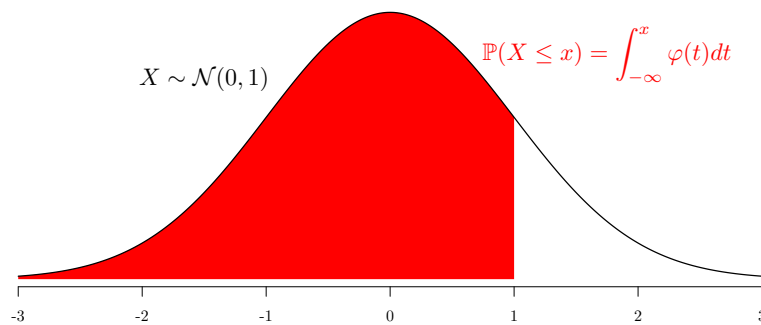
$$Var[Y] = Var[aX + b] = a^2 Var[X].$$

Disuguaglianza di Markov

Sia $(\Omega, \mathcal{A}, \mathbb{P})$ uno spazio di probabilità e X una variabile aleatoria (continua o discreta) definita su tale spazio. Sia $c > 0$, allora la probabilità dell'evento $(X \geq c)$ ha limite superiore:

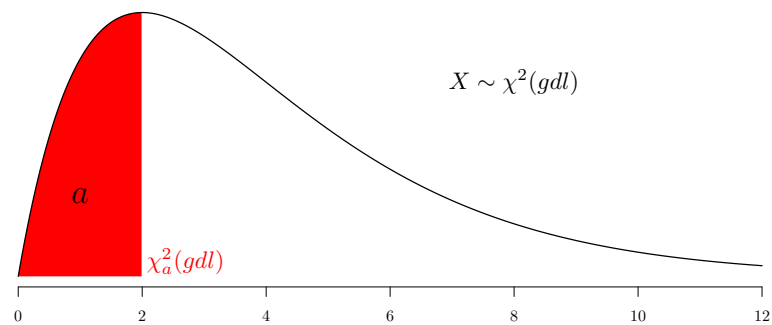
$$\mathbb{P}(X \geq c) \leq \frac{\mathbb{E}[X]}{c}$$

Tavola Normale



	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

Tavola Chi-quadro



	0.001	0.01	0.025	0.05	0.1	0.9	0.95	0.975	0.99
1	0.0000	0.0002	0.0010	0.0039	0.0158	2.7055	3.8415	5.0239	6.6349
2	0.0020	0.0201	0.0506	0.1026	0.2107	4.6052	5.9915	7.3778	9.2103
3	0.0243	0.1148	0.2158	0.3518	0.5844	6.2514	7.8147	9.3484	11.3449
4	0.0908	0.2971	0.4844	0.7107	1.0636	7.7794	9.4877	11.1433	13.2767
5	0.2102	0.5543	0.8312	1.1455	1.6103	9.2364	11.0705	12.8325	15.0863
6	0.3811	0.8721	1.2373	1.6354	2.2041	10.6446	12.5916	14.4494	16.8119
7	0.5985	1.2390	1.6899	2.1673	2.8331	12.0170	14.0671	16.0128	18.4753
8	0.8571	1.6465	2.1797	2.7326	3.4895	13.3616	15.5073	17.5345	20.0902
9	1.1519	2.0879	2.7004	3.3251	4.1682	14.6837	16.9190	19.0228	21.6660
10	1.4787	2.5582	3.2470	3.9403	4.8652	15.9872	18.3070	20.4832	23.2093
11	1.8339	3.0535	3.8157	4.5748	5.5778	17.2750	19.6751	21.9200	24.7250
12	2.2142	3.5706	4.4038	5.2260	6.3038	18.5493	21.0261	23.3367	26.2170
13	2.6172	4.1069	5.0088	5.8919	7.0415	19.8119	22.3620	24.7356	27.6882
14	3.0407	4.6604	5.6287	6.5706	7.7895	21.0641	23.6848	26.1189	29.1412
15	3.4827	5.2293	6.2621	7.2609	8.5468	22.3071	24.9958	27.4884	30.5779
16	3.9416	5.8122	6.9077	7.9616	9.3122	23.5418	26.2962	28.8454	31.9999
17	4.4161	6.4078	7.5642	8.6718	10.0852	24.7690	27.5871	30.1910	33.4087
18	4.9048	7.0149	8.2307	9.3905	10.8649	25.9894	28.8693	31.5264	34.8053
19	5.4068	7.6327	8.9065	10.1170	11.6509	27.2036	30.1435	32.8523	36.1909
20	5.9210	8.2604	9.5908	10.8508	12.4426	28.4120	31.4104	34.1696	37.5662
21	6.4467	8.8972	10.2829	11.5913	13.2396	29.6151	32.6706	35.4789	38.9322
22	6.9830	9.5425	10.9823	12.3380	14.0415	30.8133	33.9244	36.7807	40.2894
23	7.5292	10.1957	11.6886	13.0905	14.8480	32.0069	35.1725	38.0756	41.6384
24	8.0849	10.8564	12.4012	13.8484	15.6587	33.1962	36.4150	39.3641	42.9798
25	8.6493	11.5240	13.1197	14.6114	16.4734	34.3816	37.6525	40.6465	44.3141
26	9.2221	12.1981	13.8439	15.3792	17.2919	35.5632	38.8851	41.9232	45.6417
27	9.8028	12.8785	14.5734	16.1514	18.1139	36.7412	40.1133	43.1945	46.9629
28	10.3909	13.5647	15.3079	16.9279	18.9392	37.9159	41.3371	44.4608	48.2782
29	10.9861	14.2565	16.0471	17.7084	19.7677	39.0875	42.5570	45.7223	49.5879
30	11.5880	14.9535	16.7908	18.4927	20.5992	40.2560	43.7730	46.9792	50.8922

Esercizio 1

Sia $(\Omega, \mathcal{A}, \mathbb{P})$ uno spazio di probabilità e X una variabile aleatoria continua definita su tale spazio. Mostrare che valgono i seguenti risultati.

$$\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$$

$$\text{Var}[aX + b] = a^2 \text{Var}[X].$$

Soluzione

Posto che:

$$\mathbb{E}[X] = \int_{-\infty}^{+\infty} xf(x)dx \quad \mathbb{E}[X^2] = \int_{-\infty}^{+\infty} x^2 f(x)dx$$

sono, rispettivamente, il valore atteso (o momento primo) ed il momento secondo della variabile casuale X , allora per il momento primo vale:

$$\mathbb{E}[aX + b] = \int_{-\infty}^{+\infty} (ax + b)f(x)dx = a \int_{-\infty}^{+\infty} xf(x)dx + b = a\mathbb{E}[X] + b$$

quindi il primo risultato è valido. Per il momento secondo vale:

$$\begin{aligned} \mathbb{E}[(aX + b)^2] &= \int_{-\infty}^{+\infty} (ax + b)^2 f(x)dx = \int_{-\infty}^{+\infty} (a^2x^2 + 2axb + b^2)f(x)dx = \\ &= \int_{-\infty}^{+\infty} a^2x^2 f(x)dx + \int_{-\infty}^{+\infty} 2axbf(x)dx + \int_{-\infty}^{+\infty} b^2 f(x)dx = \\ &= a^2\mathbb{E}[X^2] + 2ab\mathbb{E}[X] + b^2 \end{aligned}$$

e per la varianza abbiamo:

$$\begin{aligned} \text{Var}[X] &= a^2\mathbb{E}[X^2] + 2ab\mathbb{E}[X] + b^2 - (a\mathbb{E}[X] + b)^2 = \\ &= a^2\mathbb{E}[X^2] + 2ab\mathbb{E}[X] + b^2 - a^2\mathbb{E}[X]^2 - 2ab\mathbb{E}[X] - b^2 = \\ &= a^2\mathbb{E}[X^2] - a^2\mathbb{E}[X]^2 = a^2(\mathbb{E}[X^2] - \mathbb{E}[X]^2) = a^2\text{Var}[X] \end{aligned}$$

che conferma il secondo risultato.

Esercizio 2

Svolgere l'Esercizio 1 considerando X variabile casuale discreta

Soluzione

Posto che:

$$\mathbb{E}[X] = \sum_{x=0}^{+\infty} xp(x) \quad \mathbb{E}[X^2] = \sum_{x=0}^{+\infty} x^2p(x)$$

sono, rispettivamente, il valore atteso (o momento primo) ed il momento secondo della variabile casuale X , allora per il momento primo vale:

$$\mathbb{E}[aX + b] = \sum_{x=0}^{+\infty} (ax + b)p(x) = \sum_{x=0}^{+\infty} xp(x) + b = a\mathbb{E}[X] + b$$

quindi il primo risultato è valido. Per il momento secondo vale:

$$\begin{aligned} \mathbb{E}[(aX + b)^2] &= \sum_{x=0}^{+\infty} (ax + b)^2 p(x) = \sum_{x=0}^{+\infty} (a^2x^2 + 2abx + b^2)p(x) = \\ &= \sum_{x=0}^{+\infty} a^2x^2p(x) + \sum_{x=0}^{+\infty} 2abxp(x) + \sum_{x=0}^{+\infty} b^2p(x) = \\ &= a^2\mathbb{E}[X^2] + 2ab\mathbb{E}[X] + b^2 \end{aligned}$$

e, quindi, per la varianza abbiamo:

$$\begin{aligned} Var[X] &= a^2\mathbb{E}[X^2] + 2ab\mathbb{E}[X] + b^2 - (a\mathbb{E}[X] + b)^2 = \\ &= a^2\mathbb{E}[X^2] + 2ab\mathbb{E}[X] + b^2 - a^2\mathbb{E}[X]^2 - 2ab\mathbb{E}[X] - b^2 = \\ &= a^2\mathbb{E}[X^2] - a^2\mathbb{E}[X]^2 = a^2(\mathbb{E}[X^2] - \mathbb{E}[X]^2) = a^2Var[X] \end{aligned}$$

che conferma il secondo risultato.



Esercizio 3

Paperino, Qui, Quo, Qua ed Archimede decidono di fare una gita al mare. Durante la sera, camminando lungo la spiaggia, si imbattono in uno spettacolo pirotecnico. I fuochi sono sparati da una piattaforma in mezzo al mare. Iniziano a discutere su quanto sia distante la piattaforma dalla spiaggia, finché Archimede non propone una sfida: ognuno deve misurare il tempo che trascorre tra il lampo prodotto da un fuoco d'artificio ed il suono prodotto, così da poter ricavare successivamente la distanza tra la spiaggia e la piattaforma. I tempi misurati, espressi in secondi, sono:

$$x_1 = 1.15$$

$$x_2 = 0.98$$

$$x_3 = 1.03$$

$$x_4 = 1.14$$

Archimede, dopo qualche calcolo, esclama "La piattaforma è distante 369.681 metri con una varianza pari a 617.906!".

1. Calcolare la mediana dei tempi misurati e lo scostamento medio dalla mediana.
2. Calcolare la media dei tempi misurati e la varianza.
3. Mostrare come Archimede sia riuscito ad ottenere il suo risultato (può essere utile considerare una trasformazione lineare - il suono viaggia a 1238 Km/h).

Soluzione

1. Essendo $n = 4$ pari, la mediana è data dal valore medio dei due valori centrali misurati:

$$Me(\underline{x}) = \frac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2} = \frac{1.03 + 1.14}{2} = 1.085$$

Lo scostamento medio dalla mediana è dato dalla media degli scarti in valore assoluto dalla mediana, ovvero:

$$\frac{1}{n} \sum_{i=1}^n |x_i - Me| = \frac{1}{4} [0.105 + 0.055 + 0.055 + 0.065] = 0.28$$

2. La media dei tempi misurati è uguale a:

$$M_1(\underline{x}) = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{4} [1.15 + 0.98 + 1.03 + 1.14] = 1.075$$

La media quadratica di $x_1 \dots x_4$ è pari a:

$$M_2(\underline{x}) = \frac{1}{n} \sum_{i=1}^n x_i^2 = \frac{1}{n} [1.323 + 0.960 + 1.061 + 1.300] = 1.161$$

quindi la varianza può essere calcolata come:

$$Var(\underline{x}) = M_2(\underline{x}) - [M_1(\underline{x})]^2 = 1.161 - 1.075^2 = 0.00538$$

3. Considerando come trascurabile il fatto che anche il lampo prodotto da un fuoco d'artificio abbia una sua velocità, e trattandolo come istantaneo, Archimede ha ben pensato che avendo a disposizione la media dei tempi trascorsi, definendo a come:

$$a = \frac{\text{velocità suono (m/h)}}{\text{secondi in un'ora}} = \frac{1238000}{3600}$$

e moltiplicando la media trovata precedentemente si può ottenere la distanza media della piattaforma dalla spiaggia e la varianza. Definendo una trasformazione lineare della variabile x come $y = ax$, dove y rappresenta la distanza in metri, associata ad un ritardo x tra lampo e percezione sonora, ed utilizzando una proprietà della media, che è un operatore lineare:

$$\bar{y} = a \bar{x} = \frac{1238000}{3600} 1.075 = 369.681$$

e

$$Var(\underline{y}) = a^2 Var(\underline{x}) = \left(\frac{1238000}{3600} \right)^2 0.00538 = 636.237$$

Esercizio 4

Sono stati rilevati i seguenti pesi \underline{x} , espressi in Kg, per due gruppi di atleti differenti, rispettivamente femmine e maschi:

$$\underline{x}^F = (51, 56, 58, 59, 62)$$

$$\underline{x}^M = (75, 77, 82, 84)$$

Verificare la proprietà di scomposizione della varianza.

Soluzione

Bisogna calcolare valore medio e varianza per l'intero campione e nei due gruppi considerati, per poi verificare l'uguaglianza secondo cui (posto K pari al numero di gruppi considerati):

$$\sigma^2 = \left[\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_i)^2 \right] + \left[\frac{1}{n} \sum_{i=1}^n (\bar{x}_i - \bar{x})^2 \right] = \left[\frac{1}{n} \sum_{j=1}^K n_j \sigma_j^2 \right] + \left[\frac{1}{n} \sum_{j=1}^K n_j (\bar{x}_j - \bar{x})^2 \right]$$

con \bar{x}_j media nel gruppo j-esimo, \bar{x}_i media di gruppo corrispondente all'osservazione i-esima e σ_j^2 varianza del gruppo j-esimo. La varianza può quindi essere scomposta in varianza nei gruppi (primo termine) e varianza tra i gruppi (secondo termine). Per il primo gruppo media e varianza risultano essere:

$$\bar{x}^F = \frac{1}{5} \sum_{i=1}^5 x_i^F = 57.2$$

$$\sigma_F^2 = \frac{1}{5} \sum_{i=1}^5 (x_i^F - \bar{x}^F)^2 = \frac{66.8}{5} = 13.36$$

mentre per il secondo gruppo:

$$\bar{x}^M = \frac{1}{4} \sum_{i=1}^4 x_i^M = 79.5$$

$$\sigma_M^2 = \frac{1}{4} \sum_{i=1}^4 (x_i^M - \bar{x}^M)^2 = \frac{53}{4} = 13.25$$

Considerando l'intero campione, ovvero i due gruppi congiuntamente, si ottiene:

$$\bar{x} = \frac{1}{9} \sum_{i=1}^9 x_i = 67.11$$

$$\sigma^2 = \frac{1}{9} \sum_{i=1}^9 (x_i - \bar{x})^2 = \frac{1224.89}{9} = 136.10$$

Ora è possibile mostrare che vale la scomposizione della varianza:

$$\begin{aligned} \sigma^2 &= \frac{1}{n} [n^F \sigma_F^2 + n^M \sigma_M^2] + \frac{1}{n} [n^F (\bar{x}^F - \bar{x})^2 + n^M (\bar{x}^M - \bar{x})^2] \\ &= \frac{1}{9} [119.8] + \frac{1}{9} [1105.09] = 136.10 \end{aligned}$$

che coincide con quanto trovato precedentemente.

Esercizio 5

Sia X una variabile casuale continua Uniforme definita sull'intervallo $[0, 2]$, $X \sim Unif(0, 2)$, con funzione di densità:

$$f(x; 0, 2) = \frac{1}{2} \mathbb{I}_{[0,2]} = \begin{cases} 0 & \text{se } x < 0 \\ \frac{1}{2} & \text{se } 0 \leq x < 2 \\ 0 & \text{se } x \geq 2 \end{cases}$$

1. Mostrare che l'integrale della variabile aleatoria definita dalla precedente funzione di densità assume valore pari a 1.
2. Calcolare il valore atteso della variabile aleatoria definita precedentemente.
3. Calcolare la varianza della variabile aleatoria definita precedentemente.

Soluzione

1. X è una variabile aleatoria con funzione di densità pari a $f(x; 0, 2) = \frac{1}{2} \mathbb{I}_{[0,2]}$. L'integrale di tale funzione di densità assume valore pari a:

$$\int_{-\infty}^{+\infty} f(x; 0, 2) dx = \int_{-\infty}^{+\infty} \frac{1}{2} \mathbb{I}_{[0,2]} dx = \frac{1}{2} \int_0^2 dx = \frac{1}{2} x \Big|_0^2 = \frac{1}{2} [2 - 0] = 1$$

2. Il valore atteso della variabile casuale $X \sim Unif(0, 2)$ è dato da:

$$\mathbb{E}[X] = \int_{-\infty}^{+\infty} x f(x; 0, 2) dx = \int_{-\infty}^{+\infty} x \frac{1}{2} \mathbb{I}_{[0,2]} dx = \frac{1}{2} \int_0^2 x dx = \frac{1}{2} \frac{x^2}{2} \Big|_0^2 = \frac{1}{2} \left[\frac{4}{2} - 0 \right] = 1$$

3. La varianza può essere calcolata, tramite il metodo indiretto, utilizzando il momento primo ed il momento secondo. Il momento secondo della variabile casuale $X \sim Unif(0, 2)$ risulta essere:

$$\mathbb{E}[X^2] = \int_{-\infty}^{+\infty} x^2 f(x; 0, 2) dx = \int_{-\infty}^{+\infty} x^2 \frac{1}{2} \mathbb{I}_{[0,2]} dx = \frac{1}{2} \int_0^2 x^2 dx = \frac{1}{2} \frac{x^3}{3} \Big|_0^2 = \frac{1}{2} \left[\frac{8}{3} - 0 \right] = \frac{4}{3}$$

La varianza è quindi uguale a:

$$Var[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{4}{3} - 1 = \frac{1}{3}$$

NB Sia $X \sim Unif(a, b)$, allora valgono i seguenti risultati:

$$\mathbb{E}[X] = \frac{1}{2}(a + b) \quad Var[X] = \frac{1}{12}(b - a)^2$$

Quindi, applicando i risultati alla variabile casuale $X \sim Unif(0, 2)$ si ottiene che

$$\mathbb{E}[X] = \frac{1}{2}(0 + 2) = 1$$

e

$$Var[X] = \frac{1}{12}(b - a)^2 = \frac{1}{12}2^2 = \frac{1}{3}$$

concordi con quanto trovato precedentemente.

Esercizio 6

Sia X una variabile casuale con funzione di densità $f(x)$ proporzionale ad una funzione $g(x) \geq 0$:

$$f(x) \propto g(x) = (x-1)\mathbb{I}_{[1,2]} + \left(2 - \frac{1}{2}x\right)\mathbb{I}_{[2,4]} = \begin{cases} 0 & \text{se } x < 1 \\ x-1 & \text{se } 1 \leq x < 2 \\ 2 - \frac{1}{2}x & \text{se } 2 \leq x < 4 \\ 0 & \text{se } x \geq 4 \end{cases}$$

1. Trovare la costante di normalizzazione c tale per cui $f(x) = c g(x)$ è una funzione di densità.
2. Calcolare il valore atteso della variabile aleatoria X .
3. Calcolare la varianza della variabile aleatoria X .

Soluzione

1. Per poter trovare la costante di normalizzazione, bisogna calcolare il valore dell'integrale della funzione $g(x)$:

$$\begin{aligned} \int_{-\infty}^{+\infty} g(x) dx &= \int_{-\infty}^{+\infty} \left[(x-1)\mathbb{I}_{[1,2]} + \left(2 - \frac{1}{2}x\right)\mathbb{I}_{[2,4]} \right] dx = \int_{-\infty}^{+\infty} (x-1)\mathbb{I}_{[1,2]} dx + \int_{-\infty}^{+\infty} \left(2 - \frac{1}{2}x\right)\mathbb{I}_{[2,4]} dx = \\ &= \int_1^2 (x-1) dx + \int_2^4 \left(2 - \frac{1}{2}x\right) dx = \left(\frac{x^2}{2} - x\right)\Big|_1^2 + \left(2x - \frac{x^2}{4}\right)\Big|_2^4 = \\ &= \frac{4}{2} - 2 - \frac{1}{2} + 1 + 8 - 4 - 4 + 1 = \frac{3}{2} \end{aligned}$$

Affinché $f(x)$ risulti essere una funzione di densità, deve valere che $\int f(x) dx = 1$, quindi:

$$1 = \int_{-\infty}^{+\infty} f(x) dx = \int_{-\infty}^{+\infty} c g(x) dx \implies c = \left[\int_{-\infty}^{+\infty} g(x) dx \right]^{-1} = \frac{2}{3}$$

2. Usando la costante di normalizzazione trovata al punto precedente è ora possibile scrivere la funzione di densità della variabile aleatoria X , ovvero $f(x) = \frac{2}{3}g(x)$, e calcolare il suo valore atteso:

$$\begin{aligned} \mathbb{E}[X] &= \int_{-\infty}^{+\infty} x f(x) dx = \frac{2}{3} \int_{-\infty}^{+\infty} x g(x) dx = \frac{2}{3} \int_{-\infty}^{+\infty} x \left[(x-1)\mathbb{I}_{[1,2]} + \left(2 - \frac{1}{2}x\right)\mathbb{I}_{[2,4]} \right] dx = \\ &= \frac{2}{3} \left[\left(\frac{x^3}{3} - \frac{x^2}{2}\right)\Big|_1^2 + \left(x^2 - \frac{x^3}{6}\right)\Big|_2^4 \right] = \frac{2}{3} \frac{7}{2} = \frac{7}{3} \end{aligned}$$

3. La varianza può essere calcolata tramite il metodo indiretto. Il momento secondo risulta essere:

$$\begin{aligned} \mathbb{E}[X^2] &= \int_{-\infty}^{+\infty} x^2 f(x) dx = \frac{2}{3} \int_{-\infty}^{+\infty} x^2 g(x) dx = \frac{2}{3} \int_{-\infty}^{+\infty} x^2 \left[(x-1)\mathbb{I}_{[1,2]} + \left(2 - \frac{1}{2}x\right)\mathbb{I}_{[2,4]} \right] dx = \\ &= \frac{2}{3} \left[\left(\frac{x^4}{4} - \frac{x^3}{3}\right)\Big|_1^2 + \left(\frac{2x^3}{3} - \frac{x^4}{8}\right)\Big|_2^4 \right] = \frac{2}{3} \frac{115}{12} = \frac{115}{18} \end{aligned}$$

La varianza è quindi uguale a:

$$\text{Var}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{115}{18} - \left(\frac{7}{3}\right)^2 = \frac{115 - 98}{18} = \frac{17}{18}$$

Esercizio 7

Siano X una variabile aleatoria Uniforme generica, $X \sim Unif(a, b)$, ed Y una variabile aleatoria Gamma, $Y \sim Gamma(\alpha, \beta)$. Sapendo che X ed Y sono indipendenti, calcolare:

$$\mathbb{E}[X + Y]$$

Soluzione

Il valore atteso di $(X + Y)$ può essere calcolato come:

$$\begin{aligned} \mathbb{E}[X + Y] &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x + y) f_{XY}(x, y) dx dy = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x + y) f_X(x) f_Y(y) dx dy = \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x f_X(x) f_Y(y) dx dy + \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} y f_X(x) f_Y(y) dx dy = \\ &= \int_{-\infty}^{+\infty} x f_X(x) \int_{-\infty}^{+\infty} f_Y(y) dy dx + \int_{-\infty}^{+\infty} y f_Y(y) \int_{-\infty}^{+\infty} f_X(x) dx dy = \\ &= \int_{-\infty}^{+\infty} x f_X(x) dx + \int_{-\infty}^{+\infty} y f_Y(y) dy = \int_{-\infty}^{+\infty} x \frac{1}{b-a} \mathbb{I}_{[a, b]}(x) dx + \int_{-\infty}^{+\infty} y \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\beta y} dy = \\ &= \frac{x^2}{2(b-a)} \Big|_a^b + \int_{-\infty}^{+\infty} \frac{\beta^\alpha}{\Gamma(\alpha)} y^\alpha e^{-\beta y} dy = \frac{b^2 - a^2}{2(b-a)} + \frac{\alpha}{\beta} \int_{-\infty}^{+\infty} \frac{\beta^{\alpha+1}}{\Gamma(\alpha+1)} y^\alpha e^{-\beta y} dy = \\ &= \frac{b+a}{2} + \frac{\alpha}{\beta} = \mathbb{E}[X] + \mathbb{E}[Y] \end{aligned}$$

Esercizio 8

Siano X ed Y due variabili casuali indipendenti e distribuite come una variabile casuale Gaussiana standard, ovvero $X \sim \mathcal{N}(0, 1)$ ed $Y \sim \mathcal{N}(0, 1)$. Sapendo che X è indipendente da Y :

1. Calcolare:

$$\mathbb{E}[X^2 + Y^2]$$

2. Trovare un estremo superiore per $\mathbb{P}[(X^2 + Y^2) \geq 4]$

Soluzione

1. Assumendo indipendenza tra le variabili aleatorie X e Y , possiamo calcolare $\mathbb{E}[X^2 + Y^2]$ come:

$$\begin{aligned} \mathbb{E}[X^2 + Y^2] &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x^2 + y^2) f_{XY}(x, y) dx dy = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x^2 + y^2) f_X(x) f_Y(y) dx dy = \\ &= \int_{-\infty}^{+\infty} x^2 f_X(x) dx + \int_{-\infty}^{+\infty} y^2 f_Y(y) dy = [\text{Var}[X] + \mathbb{E}[X]^2] + [\text{Var}[Y] + \mathbb{E}[Y]^2] = \\ &= [1 + 0] + [1 + 0] = 2 \end{aligned}$$

2. Sia $Z = X^2 + Y^2$. Essendo noto il suo valore atteso, è possibile applicare la disuguaglianza di Markov ed identificare un estremo superiore per l'evento $(Z \geq 4)$ ovvero:

$$\mathbb{P}[(X^2 + Y^2) \geq 4] = \mathbb{P}[Z \geq 4] \leq \frac{\mathbb{E}[Z]}{4} = \frac{2}{4} = \frac{1}{2}$$

NB Siano X e Y due variabili casuali indipendenti e distribuite secondo una legge Gaussiana standard, $X \sim \mathcal{N}(0, 1)$ e $Y \sim \mathcal{N}(0, 1)$, sia inoltre $Z = X^2 + Y^2$. La distribuzione di Z è nota e corrisponde ad una Chi-quadrato con 2 gradi di libertà, $Z = X^2 + Y^2 \sim \chi_2^2$. Il valore atteso di una distribuzione Chi-quadrato è pari al numero di gradi di libertà:

$$\mathbb{E}[Z] = 2$$

concorde con quanto trovato precedentemente. Il valore esatto dell'evento $Z \geq 4$ è pari a $P[Z \geq 4] = 0.135$. La disuguaglianza di Markov fornisce un limite superiore per la probabilità di un evento che può risultare, come in questo caso, essere un'approssimazione non adeguata.

Esercizio 9

Sia $\{X_i\}_{i=1}^n$ un insieme di variabili aleatorie.

1. Calcolare

$$Var \left[\sum_{i=1}^n X_i \right]$$

2. Che forma assume il risultato precedente in caso di indipendenza tra le variabili $\{X_i\}$?

Soluzione

1. La varianza può essere scomposta come:

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

Sviluppando il primo termine si ottiene:

$$\mathbb{E} \left[\left(\sum_{i=1}^n X_i \right)^2 \right] = \mathbb{E} \left[\sum_{i=1}^n X_i^2 + \sum_{i=1}^n \sum_{\substack{j=1 \\ i \neq j}}^n X_i X_j \right] = \sum_{i=1}^n \mathbb{E}[X_i^2] + \sum_{i=1}^n \sum_{\substack{j=1 \\ i \neq j}}^n \mathbb{E}[X_i X_j]$$

Sviluppando il secondo termine:

$$\left(\mathbb{E} \left[\sum_{i=1}^n X_i \right] \right)^2 = \left(\sum_{i=1}^n \mathbb{E}[X_i] \right)^2 = \sum_{i=1}^n \mathbb{E}[X_i]^2 + \sum_{i=1}^n \sum_{\substack{j=1 \\ i \neq j}}^n \mathbb{E}[X_i] \mathbb{E}[X_j]$$

Combinando gli sviluppi dei due termini:

$$\begin{aligned} Var \left[\sum_{i=1}^n X_i \right] &= \sum_{i=1}^n \mathbb{E}[X_i^2] + \sum_{i=1}^n \sum_{\substack{j=1 \\ i \neq j}}^n \mathbb{E}[X_i X_j] - \sum_{i=1}^n \mathbb{E}[X_i]^2 - \sum_{i=1}^n \sum_{\substack{j=1 \\ i \neq j}}^n \mathbb{E}[X_i] \mathbb{E}[X_j] \\ &= \sum_{i=1}^n (\mathbb{E}[X_i^2] - \mathbb{E}[X_i]^2) + \sum_{i=1}^n \sum_{\substack{j=1 \\ i \neq j}}^n (\mathbb{E}[X_i X_j] - \mathbb{E}[X_i] \mathbb{E}[X_j]) = \\ &= \sum_{i=1}^n Var[X_i] + \sum_{i=1}^n \sum_{\substack{j=1 \\ i \neq j}}^n Cov[X_i, X_j] \end{aligned}$$

2. Sotto ipotesi di indipendenza tra le variabili $\{X_i\}$ tutte le covarianze si annullano:

$$Var \left[\sum_{i=1}^n X_i \right] = \sum_{i=1}^n Var[X_i]$$

Esercizio 10

Sia X una variabile casuale Normale con media pari a 5 e varianza pari a 3, $X \sim \mathcal{N}(5, 3)$. Calcolare:

1. La probabilità $\mathbb{P}[X > 6]$.
2. La probabilità $\mathbb{P}[4 < X < 7]$.
3. Il valore di c tale per cui $\mathbb{P}[X > c] = 0.3$.
4. Il valore atteso della trasformazione lineare $\mathbb{E}[2X - 5]$.
5. La varianza della trasformazione lineare $\text{Var}[2X - 5]$.

Soluzione

1. La probabilità $P[X > 6]$ può essere riportata ad un problema riguardante la Normale standardizzata e valutata utilizzando la tavola della variabile casuale Normale (inizio file), infatti:

$$P[X > 6] = P\left[\frac{X - \mu}{\sigma} > \frac{6 - \mu}{\sigma}\right] = P[Z > 0.58] = 1 - P[Z \leq 0.58] = 1 - \Phi(0.58) = 0.281$$

dove $\Phi(\cdot)$ è la funzione di ripartizione della variabile casuale Normale standard.

2. Seguendo un ragionamento analogo a quanto fatto nel punto precedente, abbiamo:

$$\begin{aligned} P[4 < X < 7] &= P\left[\frac{4 - \mu}{\sigma} < \frac{X - \mu}{\sigma} < \frac{7 - \mu}{\sigma}\right] = P\left[\frac{4 - 5}{\sqrt{3}} < \frac{X - 5}{\sqrt{3}} < \frac{7 - 5}{\sqrt{3}}\right] = \\ &= P[-0.58 < Z < 1.16] = \Phi(1.16) - \Phi(-0.58) = \Phi(1.16) - [1 - \Phi(0.58)] = \\ &= 0.877 - (1 - 0.719) = 0.596 \end{aligned}$$

3. La costante c può essere trovata riconducendosi ad una variabile casuale Normale standard:

$$P[X > c] = 0.3 \implies P\left[Z > \frac{(c - 5)}{\sqrt{3}}\right] = 0.3 \implies \Phi(z) = \Phi\left(\frac{c - 5}{\sqrt{3}}\right) = 0.7$$

che è soddisfatta per $z = 0.52$, quindi:

$$z = \frac{c - 5}{\sqrt{3}} = 0.52 \iff c = 5.9$$

4. Il valore atteso della trasformazione lineare può essere calcolato sfruttando la proprietà di linearità:

$$\mathbb{E}[2X - 5] = 2\mathbb{E}[X] - 5 = 10 - 5 = 5$$

5. Analogamente la Varianza della trasformazione lineare diventa:

$$\text{Var}[2X - 5] = \text{Var}[2X] = 2^2 \text{Var}[X] = 12$$

NB Sia X una variabile aleatoria con legge di distribuzione Normale, $X \sim \mathcal{N}(\mu, \sigma^2)$, allora una variabile casuale Y trasformazione lineare della variabile casuale X , $Y = aX + b$ continua ad avere distribuzione Normale di parametri $Y \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$.

Esercizio 11

Un impianto di produzione inserisce automaticamente lo zucchero nelle bustine di zucchero. Sia X la variabile casuale che descrive il peso dello zucchero nelle bustine, espresso in grammi. Assumendo che X segua una distribuzione Normale, $X \sim \mathcal{N}(\mu, \sigma^2)$, che la deviazione standard è pari a $\sigma = 1.3$ e che il 5% delle bustine pesa più di 10.132g:

1. Calcolare il peso medio delle bustine.
2. Calcolare la probabilità che il peso di una bustina sia compreso tra 7g e 9g.
3. Specificare i parametri della distribuzione Y del peso delle bustine di zucchero, espresso in milligrammi.

Soluzione

1. Essendo noto il quantile di ordine 0.95 è possibile ricavare la media, infatti:

$$0.05 = \mathbb{P}[X > 10.132] = \mathbb{P}\left[Z > \frac{10.132 - \mu}{1.3}\right]$$

che risulta essere vera se, posto $1 - 0.05 = 0.95 = \Phi(z_{0.95}) = \Phi(1.64)$. Ora è sufficiente uguagliare:

$$1.64 = z_{0.95} = \frac{10.132 - \mu}{1.3} \quad \Longleftrightarrow \quad \mu = 8.$$

2. Essendo nota la media della distribuzione, è semplice calcolare la probabilità dell'evento $(7 < X < 9)$.

$$\begin{aligned}\mathbb{P}[7 < X < 9] &= \mathbb{P}[X < 9] - \mathbb{P}[X < 7] = \mathbb{P}\left[Z < \frac{9 - 8}{1.3}\right] - \mathbb{P}\left[Z < \frac{7 - 8}{1.3}\right] = \\ &= \Phi\left(\frac{9 - 8}{1.3}\right) - \Phi\left(\frac{7 - 8}{1.3}\right) = \Phi(0.77) - [1 - \Phi(0.77)] = \\ &= 2\Phi(0.77) - 1 = 2(0.7794) - 1 = 0.5588.\end{aligned}$$

3. La variabile Y rappresenta il peso, espresso in milligrammi. Essendo X la legge di distribuzione del peso espresso in grammi, vale $Y = aX$, con $a = 1000 = 10^3$. Partendo dalla legge di distribuzione di X , e ricordando che $Y \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$, è possibile ricavare i parametri per la distribuzione di Y :

$$Y \sim \mathcal{N}(10^3(8), 10^6(1.3)) = \mathcal{N}(8000, 1300000)$$

Esercizio 12

Siano X_1, \dots, X_n n variabili casuali indipendenti di media 50 e varianza 100. Posto:

$$Y_n = \sum_{i=1}^n X_i$$

1. Fornire un valore approssimato della probabilità che $Y_8 \geq 700$.
2. Determinare la probabilità che $Y_{50} > 2700$.
3. Indicare il valore massimo di n tale per cui $P[Y_n > 2700] \leq 0.03$.

Soluzione

Essendo Y_n definita come:

$$Y_n = \sum_{i=1}^n X_i$$

ed essendo le variabili indipendenti, si ha che $\mathbb{E}[Y_n] = 50n$ e $\text{Var}[Y_n] = 100n$.

1. Applicando la disuguaglianza di Markov:

$$P[Y_8 \geq 700] \leq \frac{\mathbb{E}[Y_8]}{700} = \frac{400}{700} = 0.571$$

2. Per il Teorema Centrale del Limite:

$$\frac{Y_n - 50n}{10\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, 1)$$

Quindi $Y_{50} \approx \mathcal{N}(2500, 5000)$ e la probabilità dell'evento $Y_{50} > 2700$ è pari a:

$$P[Y_{50} > 2700] = P\left[\frac{Y_{50} - 2500}{\sqrt{5000}} > \frac{200}{\sqrt{5000}}\right] = P[Z > 2.828] = 1 - P[Z \leq 2.828] = 0.0025$$

- 3.

$$P[Y_n > 2700] = P\left[\frac{Y_n - 50n}{\sqrt{100n}} > \frac{2700 - (50n)}{\sqrt{100n}}\right]$$

si vuole quindi trovare quel valore n tale per cui:

$$\frac{2700 - (50n)}{\sqrt{100n}} > 1.88$$

dove il valore 1.88 corrisponde al quantile di ordine 0.97 ricavato dalle tavole della Normale standard, ovvero il valore di n tale per cui:

$$2700 - (50n) > 1.88\sqrt{100n}$$

Posto $x^2 = n$:

$$-50x^2 - 18.8x + 2700 > 0$$

I valori che supportano la precedente disequazione sono $-7.54 \leq x \leq 7.16$, quindi si ha che $0 \leq n \leq 51.3$ ed $n = 51$ è l'intero più grande che soddisfa $P[Y_n > 2700] \leq 0.03$.

Esercizio 13

Siano X_1, \dots, X_{10} un insieme di 10 variabili casuali indipendenti ed identicamente distribuite secondo una legge di distribuzione esponenziale di parametro λ .

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

1. Specificare la distribuzione Gamma identificata dalla somma:

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

2. Specificare la distribuzione Chi-quadrato identificata dalla somma:

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

3. Usando il risultato trovato nel punto precedente, stimare la probabilità:

$$P \left[\sum_{i=1}^n X_i \geq 35 \right]$$

Soluzione

1. Se $X \sim \text{Exp}(\lambda)$ allora $X \sim \text{Gamma}(1, \lambda)$. Stando all'ipotesi di indipendenza ed identica distribuzione dell'insieme X_1, \dots, X_{10} , si ha che:

$$\sum_{i=1}^n X_i \sim \text{Gamma}(n, \lambda)$$

ovvero la somma di 10 variabili casuali Esponenziali, indipendenti e identicamente distribuite, si distribuisce come una variabile casuale Gamma con $\alpha = 10$ e $\beta = \lambda$.

2. Stando al punto precedente, la somma si distribuisce come una variabile casuale $\text{Gamma}(10, \lambda)$. Posto che una variabile casuale $\text{Gamma}(\frac{k}{2}, \frac{1}{2}) \sim \chi_k^2$, allora la somma è riconducibile ad una variabile casuale Chi-quadrato:

$$\sum_{i=1}^n X_i \sim \text{Gamma}(10, \lambda) = \frac{2}{\lambda} \text{Gamma}\left(\frac{20}{2}, \frac{1}{2}\right) \sim \frac{2}{\lambda} \chi_{20}^2$$

3. Applicando la disuguaglianza di Markov:

$$P \left[\sum_{i=1}^n X_i \geq 35 \right] = P \left[\frac{2}{\lambda} \chi_{20}^2 \geq 35 \right] = P \left[\chi_{20}^2 \geq \frac{35}{2} \lambda \right] \leq \frac{\mathbb{E}[\chi_{20}^2]}{\frac{35}{2} \lambda} = \frac{40}{35\lambda}$$

Esercizio 14

Si assuma che tra i non diabetici il livello di glucosio nel sangue a digiuno sia distribuito in maniera approssimativamente Normale con una media di 105mg/100ml ed una deviazione standard di 9mg/100ml.

1. Quale percentuale di non diabetici hanno livelli compresi tra 90 e 125mg/100ml?
2. Quale livello lascia il 10% dei non diabetici nella coda sinistra?
3. Quali livelli comprendono il 95% dei non diabetici?

Soluzione

1. Sia X il livello di glucosio nel sangue a digiuno.

$$P[X > 90] = P\left[Z > \frac{90 - 105}{9}\right] = P[Z > -1.67] = 0.952$$

$$P[X > 125] = P\left[Z > \frac{125 - 105}{9}\right] = P[Z > 2.22] = 0.0131$$

$$P[90 \leq X \leq 125] = P[-1.67 \leq Z \leq 2.22] = 0.952 - 0.0131 = 0.939$$

2. Bisogna trovare quel valore x tale per cui $P[X \leq x] = 10\%$, ovvero:

$$0.1 = P[Z \leq z] = P[Z \leq -1.28]$$

Posto che $z = \frac{x-105}{9}$, allora:

$$-1.28 = \frac{x - 105}{9} \implies x = -1.28 \times 9 + 105 = 93.5$$

3. Bisogna trovare i valori x_L e x_U tali per cui $P[x_L \leq X \leq x_U] = 95\%$, ovvero:

$$0.95 = P[x_L \leq X \leq x_U] = P[z_L \leq Z \leq z_U] = P[-1.96 \leq Z \leq 1.96]$$

Quindi

$$x_L = -1.96 \times 9 + 105 = 87.4$$

$$x_U = 1.96 \times 9 + 105 = 122.6$$