

Statistica e Analisi dei Dati

Indice

I	Statistica Descrittiva	6
1	Definizioni Generali	7
1.1	Carattere	7
1.1.1	Qualitativo	7
1.1.2	Quantitativo	7
1.2	Popolazione	8
1.3	Campione	8
1.3.1	Campione bivariato	8
1.3.2	Bias	8
1.4	Frequenza	8
1.4.1	Frequenza assoluta	8
1.4.2	Frequenza relativa	8
1.5	Classificatori	8
1.5.1	Curva ROC	10
1.5.2	AUC	10
2	Indici	12
2.1	Indici di centralità	12
2.1.1	Media Campionaria	12
2.1.2	Mediana Campionaria	12
2.1.3	Moda Campionaria	12
2.2	Indici di dispersione	13
2.2.1	Quantile campionario	13
2.2.2	Varianza campionaria	13
2.2.3	Deviazione standard campionaria	14
2.2.4	Coefficiente di variazione	14
2.2.5	Covarianza campionaria	14
2.3	Indici di correlazione	15
2.3.1	Coefficiente di correlazione campionaria	15
2.3.2	Coefficiente di correlazione di Pearson	15
2.4	Indici di eterogeneità	16
2.4.1	Eterogeneità	16
2.4.2	Indice di Gini	16
2.4.3	Indice di Gini normalizzato	17
2.4.4	Entropia	17
2.5	Indici di concentrazione	17
2.5.1	Indice di Gini per la concentrazione	18
2.6	Trasformazioni lineari	18
2.6.1	Traslazione	18
2.6.2	Contrazione e dilatazione	18
2.6.3	Standardizzazione	19

2.6.4	Trasformazione logaritmica	19
2.7	ANOVA	19
II Calcolo Combinatorio e Probabilità		21
3	Calcolo Combinatorio	22
3.1	Principio di enumerazione	22
3.2	Disposizioni	22
3.2.1	Disposizioni con ripetizione	22
3.2.2	Disposizioni senza ripetizione	23
3.2.3	Permutazioni	23
3.3	Combinazioni	23
3.4	Esercizi	24
4	Teoria della probabilità	27
4.1	Teoria degli insiemi	27
4.1.1	Insieme universo	27
4.1.2	Evento elementare	27
4.1.3	Evento	27
4.1.4	Algebra degli eventi	27
4.1.5	Funzione di probabilità	27
4.1.6	Operazioni	28
4.1.7	Proprietà di intersezione e unione	28
4.1.8	Spazio delle probabilità	28
4.1.9	Spazi equiprobabili	28
4.2	Teoremi fondamentali	29
4.2.1	Probabilità del complementare di un evento	29
4.2.2	Probabilità dell'unione di due eventi	30
4.2.3	Probabilità di un evento impossibile	30
4.3	Probabilità condizionata	31
4.3.1	Regola di fattorizzazione	31
4.3.2	Teorema delle probabilità totali	32
4.3.3	Formula generalizzata del teorema delle probabilità totali	32
4.3.4	Teorema di Bayes	33
4.3.5	Teorema Naive-Bayes	34
4.3.6	Indipendenza	34
4.4	Variabile aleatoria	36
4.4.1	Esempio	36
4.5	Funzione indicatrice	37
4.5.1	Esempio 1	37
4.5.2	Esempio 2	37
4.6	Funzione di ripartizione di distribuzione cumulativa	37
4.7	Variabili aleatorie discrete	37
4.7.1	Funzione di massa di probabilità	38
4.7.2	Funzione di ripartizione	38
4.8	Valore atteso	39
4.8.1	Considerazione	41
4.9	Varianza	41
4.9.1	Esempio 1	42
4.9.2	Deviazione standard	43
4.10	Funzione di massa di probabilità congiunta	43

4.10.1	Considerazione	43
4.11	Funzione di ripartizione congiunta	43
4.11.1	Esempio	43
4.12	Indipendenza	44
4.12.1	Considerazione	44
4.13	Variabile aleatoria multivariata	44
4.13.1	Considerazione	47
4.14	Covarianza	47
4.14.1	Considerazione 1	48
4.14.2	Considerazione 2	48
4.14.3	Considerazione 3	49
4.15	Coefficiente di correlazione	50
4.15.1	Formula	50
4.16	Funzione di densità di probabilità	50
4.17	Sommario delle proprietà essenziali	52
4.18	Disuguaglianza di Markov	52
4.19	Disuguaglianza di Chebyshev	53
4.20	Esercizi	54
5	Modelli Discreti	57
5.1	Modello di Bernoulli	57
5.2	Modello Binomiale	59
5.2.1	Esempio	60
5.3	Modello Uniforme Discreto	61
5.4	Modello Geometrico	63
5.4.1	Considerazione	66
5.4.2	Esempio	67
5.5	Modello di Poisson	67
5.5.1	Esempio 1	71
5.5.2	Esempio 2	72
5.6	Modello Ipergeometrico	72
5.6.1	Esempio 1	76
5.6.2	Esempio 2	76
6	Modelli Continui	77
6.1	Modello Uniforme Continuo	77
6.1.1	Considerazione 1	79
6.1.2	Considerazione 2	80
6.1.3	Esempio	80
6.2	Modello Esponenziale	80
6.2.1	Considerazione 1	83
6.2.2	Considerazione 2	83
6.2.3	Considerazione 3	84
6.2.4	Considerazione 4	85
6.2.5	Esempio 1	86
6.2.6	Esempio 2	86
6.3	Modello Gaussiano	86
6.3.1	Considerazione 1	88
6.3.2	Considerazione 2	88
6.3.3	Considerazione 3	88
6.3.4	Considerazione 4	89
6.3.5	Considerazione 5	89

6.3.6	Legge empirica	90
6.3.7	Considerazione 6	90
6.3.8	Considerazione 7	90
6.3.9	Quantili	91
6.3.10	Teorema Centrale del Limite	92
6.3.11	Teorema di De Moivre-Laplace	93
6.3.12	Esempio 1	93
6.3.13	Esempio 2	94
6.3.14	Esempio 3	94
6.3.15	Esempio 4	95
6.3.16	Esempio 5	95
6.3.17	Esempio 6	96
6.4	Esercizi	96

III Statistica Inferenziale 98

7 Statistiche e Stimatori 99

7.1	Statistica	99
7.1.1	Esempio	100
7.2	Distorsione	100
7.2.1	Esempio 1	100
7.2.2	Esempio 2	101
7.2.3	Esempio 3	101
7.2.4	Considerazione 1	102
7.2.5	Considerazione 2	102
7.3	Mean Square Error	102
7.4	Bias	103
7.5	Consistenza in media quadratica	103
7.5.1	Considerazione	104
7.5.2	Esempio 1	104
7.5.3	Esempio 2	105
7.5.4	Esempio 3	108
7.5.5	Esempio 4	108
7.5.6	Esempio 5	109
7.5.7	Esempio 6	109
7.5.8	Esempio 7	110
7.6	Legge dei Grandi numeri	110
7.6.1	Forma forte	110
7.6.2	Forma debole	110
7.6.3	Esempio 1	110
7.6.4	Esempio 2	111
7.7	Stima dell'errore	113
7.8	Riepilogo stimatori	114
7.8.1	Distorsione	114
7.8.2	Stimatore media campionaria	114
7.8.3	Stimatore varianza e deviazione standard	114

8 Processo di Poisson 115

8.1	Il processo di Poisson	115
8.1.1	Esempio	117

IV	Appendice	118
9	Cheatsheet Matematica	119
9.1	Dai dati alla distribuzione	119
9.2	Derivate	120
9.3	Integrali	120
10	Cheatsheet Python	121
10.1	Alberi di decisione	121
10.2	Calcolo combinatorio	122
10.3	Errore nella stima	122
10.4	Frequenze	123
10.5	Grafici	124
10.6	Indici	125
10.7	Miscellanea	126
10.8	Ordinamento	127

Parte I

Statistica Descrittiva

Capitolo 1

Definizioni Generali

Carattere

1.1

1.1.1 Qualitativo

Si parla di dati qualitativi quando la misurazione è fatta scegliendo un'etichetta a partire da un insieme di etichette disponibili. Le tipologie di dato qualitativo sono:

Nominale

- Sono dati che non posso confrontare e che non possiedono un ordine o categorie intrinseche.
- Colore dei capelli, categoria di prodotti e simili.

Binario o Booleano

- Sono un caso particolare dei dati a carattere nominale.
- Genere, vero o falso, assenza o presenza.

Ordinale

- Rappresentano caratteri che possono essere messi in relazione con operatori di $<$, $=$, $>$.
- Livelli di soddisfazione (basso, medio, alto), classifiche (primo, secondo, terzo) o risultati (triste, neutro, felice).

1.1.2 Quantitativo

Si parla di dati quantitativi se l'esito della misurazione è una quantità numerica. Inoltre:

- Sono scalari e sono caratteri quantitativi con un ordine definito e spesso hanno unità di misura.
- Altezza, peso, temperatura in gradi Celsius.

Infine, un carattere quantitativo può essere diviso in:

Discreto

- Tutti quei caratteri che ha senso rappresentare singolarmente.
- Data di nascita, oggetti venduti da un negozio e così via.

Continuo

- Tutti quei caratteri che ha senso rappresentare tramite un intervallo.
- Altezza, temperatura corporea, forza di un supereroe.

Popolazione

1.2

Insieme completo di oggetti. Esso è spesso troppo grande perché sia possibile un esame esaustivo. Per esempio i residenti di una regione oppure i televisori prodotti da un'azienda.

Campione

1.3

Un sottogruppo della popolazione, utile per apprendere informazioni dalla popolazione. Esso deve essere rappresentativo della popolazione.

$$C = \{x_1, \dots, x_n\} \quad (1)$$

1.3.1 Campione bivariato

Talvolta non abbiamo a che fare con sequenze di dati singoli, ma con sequenze di coppie di numeri, tra i quali esiste qualche relazione. In questi casi ogni coppia è da considerarsi un'osservazione. Possiamo denotare con (x_i, y_i) la coppia di valori ed essi prendono il nome di campione bivariato. Un esempio di campione bivariato potrebbe essere le ore di studio di uno studente e il suo punteggio all'esame.

1.3.2 Bias

Quando sei convinto di campionare bene la popolazione ma in realtà non lo stai facendo. Il bias è una tendenza a deviare dal valore atteso, introducendo distorsioni nei risultati di un esperimento.

Frequenza

1.4

1.4.1 Frequenza assoluta

Frequenza f con cui uno dei valori compare nel campione.

1.4.2 Frequenza relativa

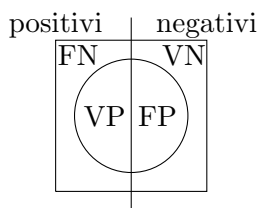
Rapporto $\frac{f}{n}$ con cui la frequenza compare nel campione rispetto a l'insieme dei dati. Possiamo poi definire la media pesata come

$$\begin{aligned} \bar{x} &= \frac{1}{n} \sum_{j=1}^m v_j f_j \\ &= v_1 \frac{f_1}{n} + v_2 \frac{f_2}{n} + \dots + v_m \frac{f_m}{n} \end{aligned} \quad (2)$$

Classificatori

1.5

Considero un classificatore binario



Dentro al cerchio mettiamo le predizioni positive mentre il quadrato è il "mondo intero". In sostanza: se sto a sinistra o a destra del quadrato, sono sicuro. Se sono dentro al cerchio, posso essere VP o se sono nel quadrato a destra sono VN, quindi il predittore ha avuto ragione. FN è quando mi dice che è negativo ma è positivo nella realtà, FP è quando mi dice positivo ma è negativo. Qui sotto la *Matrice di Confusione*, dove con E è indicata l'etichetta e con P la predizione:

		+	E	-
+	VP			FP
-	FN			VN
		TP		TN

L'**accuratezza**, compresa tra 0 e 1 è data da

$$\frac{VP + VN}{TP + TN}$$

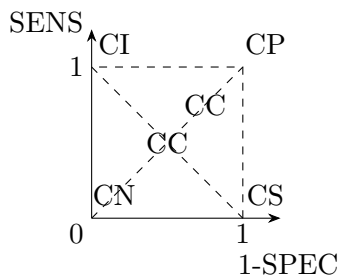
Per la collettività, a volte, i FN hanno un peso maggiore sull'accuratezza (per esempio, se c'è una malattia, i FN sono liberi di uscire di casa come i VN ma i FN dovrebbero essere come i VP e FP). Dunque si usano la **sensibilità** e la **specificità**:

$$SENS = \frac{VP}{TP} = \frac{VP}{VP + FN}$$

$$SPEC = \frac{VN}{TN} = \frac{VN}{VN + FP}$$

Per essere più chiari usando l'esempio di tamponi per la rivelazione di un patogeno, il termine *sensibilità*, quando elevata, si riferisce alla capacità di trovare il patogeno anche se presente in piccole quantità sul tampone. Invece un tampone con elevata *specificità* è in grado di rilevare solo e soltanto quel patogeno, evitando di dare un falso positivo in caso un patogeno con RNA simile sia presente sul tampone. In particolare si può costruire un grafico a partire da

$$(1 - SPEC, SENS) = \left(1 - \frac{VN}{TN}, \frac{VP}{TP}\right) = \left(\frac{FP}{TN}, \frac{VP}{TP}\right) = \left(\frac{FP}{FP + VN}, \frac{VP}{VP + FN}\right)$$



- **Classificatore Costante Positivo:** CP becca correttamente i casi positivi ma non azzecca alcun caso negativo

$$SENS = \frac{VP}{TP} = 1, SPEC = \frac{0}{TN} = 0$$

	+	E	-
+	TP		TN
-	0		0

- **Classificatore Costante Negativo:** CN è l'opposto di CP

$$SENS = \frac{0}{TP} = 0, SPEC = \frac{VN}{TN} = 1$$

	+	E	-
+	0		0
P			
-	TP		TN

- **Classificatore Ideale:** CI non sbaglia mai

$$\text{SENS} = \frac{VP}{TP} = 1, \text{ SPEC} = \frac{VN}{TN} = 1$$

	+	E	-
+	TP		0
P			
-	0		TN

- **Classificatore Sbagliato:** CS sbaglia sempre

$$\text{SENS} = \frac{0}{TP} = 0, \text{ SPEC} = \frac{0}{TN} = 0$$

	+	E	-
+	0		TN
P			
-	TP		0

- **Classificatore Casuale:** $CC_{1/2}$ lancio della moneta e $CC_{4/5}$ dado a 5 facce

$$\text{SENS} = \frac{1}{2}, \text{ SPEC} = \frac{1}{2}$$

	+	E	-
+	$\frac{1}{2}$ TP		$\frac{1}{2}$ TN
P			
-	$\frac{1}{2}$ TP		$\frac{1}{2}$ TN

$$\text{SENS} = \frac{4}{5}, \text{ SPEC} = \frac{1}{5}$$

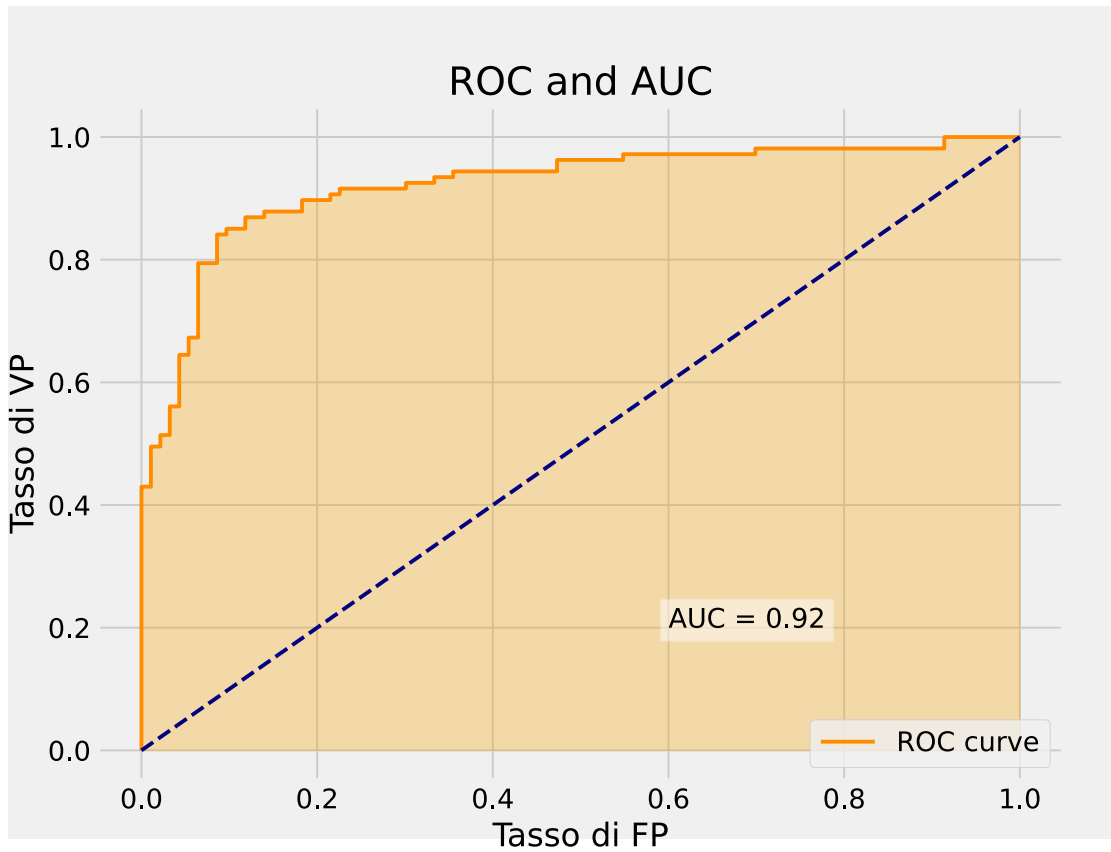
	+	E	-
+	$\frac{4}{5}$ TP		$\frac{4}{5}$ TN
P			
-	$\frac{1}{5}$ TP		$\frac{1}{5}$ TN

1.5.1 Curva ROC

Più ci si avvicina a CI e meglio è, dunque si osserva la curva ROC (*Receiver Operating Characteristic*). Tuttavia, guardare la curva ROC da sola è un giudizio qualitativo, dunque osservo l'area sottesa a ROC ossia la AUC.

1.5.2 AUC

L'*Area Under the roc Curve* è l'area sottesa alla curva ROC. Essa è $0 \leq \text{AUC} \leq 1$ e raggiunge 1, area del quadrato ($1^2 = 1$) nel caso in cui la ROC vada subito a CI e non si muova da quella posizione.



Capitolo 2

Indici

Indici di centralità

2.1

2.1.1 Media Campionaria

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (1)$$

La media campionaria è un indice di centralità ed è definita come il "baricentro dei dati". Dalla media campionaria si può arrivare al concetto di **scarto** dimostrando che

$$\begin{aligned} \sum_{i=1}^n (x_i - \bar{x}) &= \sum_{i=1}^n x_i - \sum_{i=1}^n \bar{x} \\ &= n\bar{x} - n\bar{x} \\ &= 0 \end{aligned} \quad (2)$$

La media campionaria rimane poco robusta rispetto agli outliers.

2.1.2 Mediana Campionaria

- Ordino il campione in ordine crescente
- Selezione il valore m tale che:

$$m = \begin{cases} x_{(\frac{n+1}{2})} & \text{se } |C| \text{ è dispari} \\ \frac{[x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}]}{2} & \text{se } |C| \text{ è pari} \end{cases} \quad (3)$$

Quindi nel caso dispari seleziono il valore che si trova al centro seguendo l'ordine crescente, mentre nel caso pari faccio la media aritmetica tra i due valori centrali. Se mi capita di avere elementi non scalari e dunque non divisibili, posso riportarli entrambi o sceglierne uno dei due.

2.1.3 Moda Campionaria

Definito come il dato che si presenta più spesso nel campione (caso unimodale). Se mi capita di avere due o più valori con la stessa frequenza (caso multimodale), posso riportare entrambi i valori modali o sceglierne uno dei due, come nella mediana campionaria.

2.2.1 Quantile campionario

Possiamo definire la mediana campionaria come il valore del campione che è \geq almeno il 50% dei dati e \leq almeno il 50% dei dati. Generalizzando questo concetto si può introdurre il quantile campionario di livello $q \in [0, 1]$, ossia il valore del campione \geq almeno il $100 \cdot q\%$ dei dati e \leq almeno il $100 \cdot (1 - q)\%$ dei dati.

Per esempio, immagino di avere 12 osservazioni e voglio trovare il quantile $q = 0.9$

$$x_1 \ x_2 \ x_3 \ x_4 \ x_5 \ x_6 \ x_7 \ x_8 \ x_9 \ x_{10} \ x_{11} \ x_{12} \quad (4)$$

Evidenziamo il valore $\geq 90\%$ di $12 = 10.8 = 11$

$$x_1 \ x_2 \ x_3 \ x_4 \ x_5 \ x_6 \ x_7 \ x_8 \ x_9 \ x_{10} \ \underline{x_{11}} \ x_{12} \quad (5)$$

Evidenziamo il valore $\leq 10\%$ di $12 = 1.2 = 2$

$$\underline{x_1 \ x_2 \ x_3 \ x_4 \ x_5 \ x_6 \ x_7 \ x_8 \ x_9 \ x_{10} \ x_{11} \ x_{12}} \quad (6)$$

Sovrapponiamoli

$$x_1 \ x_2 \ x_3 \ x_4 \ x_5 \ x_6 \ x_7 \ x_8 \ x_9 \ x_{10} \ \underline{x_{11}} \ x_{12} \quad (7)$$

x_{11} è il quantile che cercavo. Se invece io mi trovo ad avere dei quantili con valori interi, come per esempio nel caso di $n = 20$ e $q = 0.95$, dove x_{19} e x_{20} sono entrambi risultati, allora eseguo la media aritmetica tra i due

$$\frac{x_{19} + x_{20}}{2} \quad (8)$$

Se cambio la scala, il procedimento resta lo stesso e i quantili assumono nomi specifici. Per scale $[0 - 100]$ prendono il nome di *percentili*, per scale $[0 - 10]$ invece *decili* e infine i *quartili* per scale $[0 - 4]$. A proposito di quartili, definiamo con *range interquartile (IQR)* la differenza tra il terzo e il primo quartile.

Riassumendo, nel caso specifico dei quartili, la mediana risulta essere il secondo quartile (Q2), mentre il valore minimo e il valore massimo sono rispettivamente $Q1 - (1.5 \cdot IQR)$ e $Q3 + (1.5 \cdot IQR)$.

2.2.2 Varianza campionaria

Vogliamo osservare quanto i dati siano dispersi o concentrati, per farlo potremmo giudicare dalla loro distanza in modulo $|x_i - \bar{x}|$. Tuttavia non vogliamo usare il valore assoluto per togliere i valori negativi perché è scomodo lavorare con il valore assoluto, dunque eleviamo al quadrato $(x_i - \bar{x})^2$. Facciamo questo procedimento per ogni elemento del campione e dividiamo così da ottenere

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (9)$$

In particolare possiamo osservare che

$$\begin{aligned} \sum_i (x_i - \bar{x})^2 &= \sum_i (x_i^2 - 2\bar{x}x_i + \bar{x}^2) \\ &= \sum_i x_i^2 - \sum_i 2\bar{x}x_i + \sum_i \bar{x}^2 \\ &= \sum_i x_i^2 - 2n\bar{x}^2 + n\bar{x}^2 \\ &= \sum_i x_i^2 - n\bar{x}^2 \end{aligned} \quad (10)$$

E dunque possiamo riscrivere la varianza campionaria come

$$\frac{1}{n-1} \left(\sum_i x_i^2 - n\bar{x}^2 \right) \quad (11)$$

Alcune proprietà della varianza rispetto alle trasformazioni:

- **Traslazione:** $\bar{y} = \bar{x} + b$

$$\begin{aligned} s_y^2 &= \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2 \\ &= \frac{1}{n-1} \sum_{i=1}^n (x_i + b - \bar{x} - b)^2 \\ &= s^2 \end{aligned} \quad (12)$$

- **Scalatura:** $y_i = \alpha x_i$

$$\begin{aligned} s_y^2 &= \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2 \\ &= \frac{1}{n-1} \sum_{i=1}^n (\alpha x_i - \alpha \bar{x})^2 \\ &= \alpha^2 s^2 \end{aligned} \quad (13)$$

2.2.3 Deviazione standard campionaria

Quando usiamo la varianza campionaria abbiamo problemi con le unità di misura che sono tutte elevate al quadrato rispetto al normale, dunque definiamo la deviazione standard campionaria come

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (14)$$

N.B. rispetto alle operazioni di scalatura vale che $s_y = |\alpha| s_x$

2.2.4 Coefficiente di variazione

Siccome la varianza assume importanza inversamente proporzionale rispetto all'ordine di grandezza può aver senso "standardizzarla" rispetto alla media campionaria:

$$s^* = \frac{s}{|\bar{x}|} \quad (15)$$

Nel caso in cui ci siano sia valori positivi che negativi, basta effettuare una opportuna traslazione del campione per far sì che tutte le osservazioni abbiano segno concorde. Il coefficiente di variazione risulta particolarmente utile in situazioni in cui vogliamo confrontare la variabilità di distribuzioni che presentano scale di grandezza distinte.

2.2.5 Covarianza campionaria

Preso un campione bivariato, definiamo

$$Cov(x, y) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \quad (16)$$

Se il risultato è ≥ 0 allora esiste una correlazione diretta tra i dati, indiretta se < 0 . Quindi se ho due campioni di cardinalità n , voglio determinare se sono correlati; un esempio di correlazione diretta può essere

$$x_i \text{ piccolo} \iff y_i \text{ piccolo} \wedge x_i \text{ grande} \iff y_i \text{ grande}$$

2.3.1 Coefficiente di correlazione campionaria

Dato un campione bivariato si dice coefficiente di correlazione campionaria e si denota con r la quantità

$$r = \frac{1}{n-1} \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{s_x s_y} \quad (17)$$

Questo coefficiente gode delle seguenti proprietà:

- (1) $-1 \leq r \leq 1$, con relazione diretta per $r = 1$, indiretta per $r = -1$ e indefinita per $r = 0$. Questo è dimostrabile tramite i seguenti passaggi: considerando $y_i = a + bx_i$ e dunque $\bar{y} = a + b\bar{x}$, $s_y = |b| s_x$ e $y_i - \bar{y} = a + bx_i - a - b\bar{x} = bx_i - b\bar{x}$. Possiamo quindi dedurre che

$$\begin{aligned} r &= \frac{1}{n-1} \frac{b \sum_{i=1}^n (x_i - \bar{x})^2}{s_x |b| s_x} \\ &= \frac{b}{|b|} \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \frac{1}{s_x^2} \\ &= \frac{b}{|b|} \\ &= \text{sign}(b) \end{aligned} \quad (18)$$

- (2) Se r è il coefficiente di correlazione del campione (x_i, y_i) , allora lo è anche per il campione $(a + bx_i, c + dy_i)$ purché b e d abbiano lo stesso segno. Ossia il coefficiente di correlazione del campione non cambia se sommiamo costanti o moltiplichiamo per costanti: ciò significa che r non dipende dalle unità di misura.

2.3.2 Coefficiente di correlazione di Pearson

Considerando che

$$\begin{aligned} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) &= \sum_{i=1}^n x_i y_i - \sum_{i=1}^n \bar{x} y_i - \sum_{i=1}^n x_i \bar{y} + \sum_{i=1}^n \bar{x} \bar{y} \\ &= \sum_{i=1}^n x_i y_i - n\bar{x}\bar{y} - n\bar{x}\bar{y} + n\bar{x}\bar{y} \\ &= \sum_{i=1}^n x_i y_i - n\bar{x}\bar{y} \end{aligned} \quad (19)$$

allora possiamo riscrivere r come

$$\begin{aligned} r &= \frac{1}{n-1} \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{\frac{1}{n-1} \sqrt{\sum_{i=1}^n x_i^2 - n\bar{x}^2} \sqrt{\sum_{i=1}^n y_i^2 - n\bar{y}^2}} \\ &= \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{\sqrt{\sum_{i=1}^n x_i^2 - n\bar{x}^2} \sqrt{\sum_{i=1}^n y_i^2 - n\bar{y}^2}} \end{aligned} \quad (20)$$

2.4.1 Eterogeneità

Un campione omogeneo è tale quando ogni elemento è uguale, mentre un campione eterogeneo è chiamato così quando vi sono le stesse frequenze per ogni valore del campione. Per esempio il seguente campione è omogeneo



il seguente è eterogeneo



mentre quest'ultimo non è né omogeneo né eterogeneo



2.4.2 Indice di Gini

Nel caso in cui stessimo studiando la distribuzione di un attributo di una popolazione (ad esempio la ricchezza), potresti dover discernere tra una distribuzione equa ed una non equa. Posso usare l'indice di Gini che è un indice per l'eterogeneità ed è indicato come

$$I = 1 - \sum_{j=1}^k f_j^2 \quad (21)$$

definito $\forall j \ 0 \leq f_j \leq 1$, ma che valori può assumere I ?

Analizziamo prima per $\forall j \ f_j \geq 0$

$$(1) \sum_{j=1}^k f_j = 1$$

$$(2) \exists j' \ f_{j'} : f_{j'}^2 > 0 \implies \sum_{j=1}^k f_j^2 > 0$$

$$(3) I = 1 - \sum_{j=1}^k f_j^2 \implies I < 1$$

Ora analizziamo per $\forall j \ 0 \leq f_j \leq 1$

$$(1) f_j^2 \leq f_j$$

$$(2) \sum_{j=1}^k f_j^2 \leq \sum_{j=1}^k f_j = 1 \implies 1 - \sum_{j=1}^k f_j^2 \geq 0 \implies I \geq 0$$

Dunque possiamo asserire che

$$0 \leq I < 1$$

Facciamo degli esempi:

Eterogeneità Minima: Supponendo di avere  ci ritroviamo  con frequenza relativa $f = 1$, dunque

$$\begin{aligned} I &= 1 - \sum_{j=1}^k f_j^2 \\ &= 1 - 1 \\ &= 0 \end{aligned} \quad (22)$$

Eterogeneità Massima: In questo caso supponiamo $\blacksquare\blacksquare\blacklozenge\blacklozenge\blacktriangle\blacktriangle$, abbiamo $k = 3$ simboli diversi dunque la frequenza relativa sarà $\frac{\alpha}{k \cdot \alpha} = \frac{1}{k}$ e quindi

$$\begin{aligned} I &= 1 - \sum_{j=1}^k \frac{1^2}{k} \\ &= 1 - k \frac{1^2}{k} \\ &= 1 - \frac{1}{k} \\ &= \frac{k-1}{k} \end{aligned} \tag{23}$$

2.4.3 Indice di Gini normalizzato

Torna utile normalizzare l'indice di Gini quindi, dato $0 \leq I < 1$, possiamo considerare

$$I' = \frac{k}{k-1} I \tag{24}$$

2.4.4 Entropia

Denotiamo l'entropia con

$$H = \sum_{j=1}^k f_j \cdot \log \frac{1}{f_j} \geq 0 \tag{25}$$

In caso di eterogeneità minima, ossia con $f_j = 1$ abbiamo $\frac{1}{f_j} = 1$ e dunque $H = 1 \cdot \log 1 = 0$. In caso invece di eterogeneità massima si ha $\forall j f_j = \frac{1}{k} \implies H = \sum_{j=1}^k \frac{1}{k} \cdot \log k = \log k$ e dunque $0 \leq H \leq \log k$. Da qui si può normalizzare in

$$H' = \frac{H}{\log k} \tag{26}$$

così da avere $0 \leq H' \leq 1$, utile con il logaritmo in base 2 per misurare in bit.

Indici di concentrazione

2.5

Immaginiamo di avere n osservazioni a_1, \dots, a_n ordinate dalla più piccola alla più grande, ognuna delle quali rappresenta la quantità (o ricchezza) di un elemento del campione. Definiamo quindi la media campionaria e la somma totale come segue:

$$\bar{a} = \frac{1}{n} \sum_{i=1}^n a_i \tag{27}$$

$$\text{TOT} = n\bar{a} = \sum_{i=1}^n a_i \tag{28}$$

Possiamo considerare due situazioni estreme:

- (1) **Caso di concentrazione minima:** tutti gli elementi del campione assumono lo stesso valore $a_1 = a_2 = \dots = a_n$
- (2) **Caso di concentrazione massima:** tutti gli elementi del campione assumono valore 0, tranne uno che vale $n\bar{a}$

Definiamo ora due frequenze relative cumulate:

- **Frequenza relativa cumulata degli individui fino alla i -esima osservazione:**

$$F_i = \frac{i}{n} \quad \forall i = 1, \dots, n \tag{29}$$

- **Quantità relativa cumulata fino alla i -esima osservazione:**

$$Q_i = \frac{\sum_{k=1}^i a_k}{\text{TOT}} \quad (30)$$

Definiamo anche alcune proprietà:

- $0 \leq F_i \leq 1$ e $0 \leq Q_i \leq 1$
- $Q_i \leq F_i$ poiché le osservazioni sono ordinate in maniera crescente
- $Q_i = F_i$ nel caso di concentrazione minima
- $Q_n = F_n$

Pertanto, per $i = 1, \dots, n$, le coppie (F_i, Q_i) indicano che il $100F_i\%$ della popolazione detiene il $100Q_i\%$ della quantità considerata.

2.5.1 Indice di Gini per la concentrazione

L'indice di Gini è definito come:

$$G = \frac{\sum_{i=1}^{n-1} (F_i - Q_i)}{\sum_{i=1}^{n-1} F_i} \quad (31)$$

Questo indice fornisce una misura della disuguaglianza nella distribuzione della quantità considerata. Maggiore è il valore di G , maggiore è la disuguaglianza. Con l'indice di Gini è possibile generare una curva di punti nota come *curva di Lorenz*. L'area tra la curva di Lorenz e la bisettrice di equidistribuzione rappresenta il valore dell'indice di Gini. Un indice di Gini pari a zero indica una distribuzione perfettamente equa, mentre un indice di Gini pari a uno indica una massima disuguaglianza.

Trasformazioni lineari

2.6

Nel caso serva effettuare una trasformazione lineare per mappare dei valori $\{v_1 \dots v_n\}$ da un intervallo $[a, b]$ a un altro $[c, d]$, partendo da

$$\frac{x' - c}{d - c} = \frac{x - a}{b - a} \quad (32)$$

posso definire una funzione $f : X \mapsto X'$ tale che

$$v' = f(v) = c + \frac{d - c}{b - a} \cdot (v - a) \quad (33)$$

2.6.1 Traslazione

Dato $k > 0$

$$\begin{aligned} v \mapsto v' &= v + k \\ v \mapsto v' &= v - k \end{aligned} \quad (34)$$

La traslazione non ha effetto sulle frequenze. Media, mediana e quantili vengono traslati. I range tra dati e i range interquartili restano invariati e non ha nessun effetto sulla varianza (e quindi sulla deviazione standard).

2.6.2 Contrazione e dilatazione

Dato $h \in \mathbb{R}$

$$\begin{aligned} h > 1 &\text{ contrazione} \\ h < 1 &\text{ dilatazione} \end{aligned} \quad (35)$$

Media, mediana e quantili vengono scalati. Range e distanza interquartile vengono scalati. La varianza viene invece scalata di $\frac{1}{h^2}$.

2.6.3 Standardizzazione

Viene impiegata per permettere di utilizzare una scala comune tra due variabili aleatorie.

$$g(v) = \frac{v - \bar{x}}{s_x} \quad (36)$$

2.6.4 Trasformazione logaritmica

La trasformazione logaritmica è un utile strumento quando si hanno dati che presentano variazioni significative. Immagina di avere due case: una costa 100.000€ e l'altra 1.000.000€. La differenza percentuale tra questi due prezzi è del 900%, ma può risultare più intuitiva se applichiamo la trasformazione logaritmica. Dopo la trasformazione, la differenza tra i due prezzi è circa 2.3.

$$v \mapsto v' = \log v \quad (37)$$

ANOVA

2.7

Utilizziamo le tecniche dell'*ANalysis Of VAriance* per confrontare due o più gruppi tramite la loro variabilità interna. Con x_i^g definiremo l' i -esima osservazione nel g -esimo gruppo.

$$\begin{aligned} \bar{x}^g &= \frac{1}{n^g} \sum_{i=1}^{n^g} x_i^g \\ \implies n^g \bar{x}^g &= \sum_{i=1}^{n^g} x_i^g \\ \implies \sum_{g=1}^G n^g \bar{x}^g &= \sum_{g=1}^G \sum_{i=1}^{n^g} x_i^g \\ \implies \frac{1}{n} \sum_{g=1}^G n^g \bar{x}^g &= \frac{1}{n} \cdot \text{somma delle osservazioni} \\ \implies \frac{1}{n} \sum_{g=1}^G n^g \bar{x}^g &= \bar{x} \end{aligned} \quad (38)$$

Da qui possiamo quindi dichiarare SS_T (*Sum of Squares Total*), SS_W (*Sum of Squares Within*) e SS_B (*Sum of Squares Between*).

$$\begin{aligned} SS_T &= \sum_{g=1}^G \sum_{i=1}^{n^g} (x_i^g - \bar{x})^2 \\ SS_W &= \sum_{g=1}^G \sum_{i=1}^{n^g} (x_i^g - \bar{x}^g)^2 \\ SS_B &= \sum_{g=1}^G (\bar{x}^g - \bar{x})^2 \end{aligned} \quad (39)$$

Possiamo quindi ottenere:

$$\text{Varianza: } s^2 = \frac{SS_T}{n-1}$$

$$\text{Varianza Entro i Gruppi: } s^2 = \frac{SS_W}{n-G}$$

$$\text{Varianza Tra i Gruppi: } s^2 = \frac{SS_B}{G-1}$$

Inoltre $SS_T = SS_W + SS_B$ e $\frac{SS_T}{n-1} = \frac{n-G}{n-1} \frac{SS_W}{n-G} + \frac{G-1}{n-1} \frac{SS_B}{G-1}$. L'ANOVA viene spesso utilizzata per testare l'ipotesi che non ci siano differenze significative tra i gruppi. Questo viene fatto confrontando il rapporto delle varianze tra i gruppi (SS_B) e all'interno dei gruppi (SS_W). Se il rapporto è sufficientemente grande da suggerire che la variabilità tra i gruppi è significativa rispetto alla variabilità all'interno dei gruppi, possiamo respingere l'ipotesi e concludere che almeno uno dei gruppi è significativamente diverso dagli altri.

Parte II

Calcolo Combinatorio e Probabilità

Capitolo 3

Calcolo Combinatorio

Principio di enumerazione

3.1

Avendo un esperimento con m esiti e un secondo esperimento con n esiti, abbiamo $m \cdot n$ coppie ordinate possibili di esiti. Più in generale:

$$S = \prod_{i=1}^t s_i \quad (1)$$

Per esempio se produco penne di 4 modelli e 3 colori diversi ciascuno, avrò $3 \cdot 4 = 12$ possibili combinazioni.

Disposizioni

3.2

Disposizioni di n oggetti in k posti:

- L'ordine conta
- Con ripetizione e senza ripetizione

3.2.1 Disposizioni con ripetizione

Grazie al principio di enumerazione, sappiamo che il conteggio totale è n^k , dunque

$$D_{n,k} = n^k \quad (2)$$

Ipotizzando di avere $\{a, b, c, d\}$ oggetti e 2 posti, otterremo:

aa	ba	ca	da
ab	bb	cb	db
ac	bc	cc	dc
ad	bd	cd	dd

ossia $n^k = 4^2 = 16$.

Le disposizioni con ripetizione possono risolvere il problema: *Quanti sono i numeri composti da 6 cifre tutte dispari?* Sappiamo che ci sono 5 numeri dispari, ossia $\{1, 3, 5, 7, 9\}$, i nostri oggetti, e dobbiamo inserirli in un numero da 6 cifre, le nostre posizioni, dunque avremo $5^6 = 15625$ numeri da 6 cifre con tutte cifre dispari.

3.2.2 Disposizioni senza ripetizione

In questo caso, se scelgo un oggetto, non posso sceglierlo nuovamente.

$$d_{n,k} = \frac{n!}{(n-k)!} \quad (3)$$

In una gara di 10 cavalli indovinare i primi 3 tenendo conto della posizione. Ovviamente un cavallo non può arrivare secondo e contemporaneamente quinto, quindi usiamo disposizioni senza ripetizione impostando 10 cavalli e 3 posizioni. Il risultato sarà $\frac{10!}{(10-3)!} = \frac{10 \cdot 9 \cdot 8 \cdot 7!}{7!} = 720$.

3.2.3 Permutazioni

Le permutazioni di n oggetti, identificate con P_n , non sono altro che disposizioni senza ripetizione in cui il numero delle posizioni equivale al numero degli oggetti. Vogliamo quindi riposizionare gli stessi oggetti, solamente tramite configurazioni diverse. Dunque

$$\begin{aligned} P_n &= d_{n,n} \\ &= \frac{n!}{(n-n)!} \\ &= n! \end{aligned} \quad (4)$$

Un esempio di applicazione sono gli anagrammi, per esempio quanti anagrammi ha la parola "giornale"? Esso è composto da 8 lettere da posizionare in 8 posizioni, quindi $8! = 40320$. Quanti anagrammi di "giornale" invece contengono la sillaba "le"? Si considera $\{g, i, o, r, n, a, le\}$, quindi $7! = 5040$.

In realtà, se vogliamo costruire anagrammi "più corretti", se usiamo le permutazioni ci aspettiamo che ogni lettera sia diversa dall'altra, ma se abbiamo lettere ripetute come nel caso di "tappo" e "mamma" non dobbiamo riconsiderare lo stesso anagramma, infatti per "tappo" avremo 5 lettere ma "p" si ripete 2 volte quindi $\frac{5!}{2!} = 60$, mentre per "mamma" abbiamo 5 lettere con "a" ripetuto 2 volte e "m" ripetuto 3 volte, quindi $\frac{5!}{2!3!}$.

Infine possiamo vedere come da $n! = n \cdot (n-1)!$ si riesca a provare $1 = 1! = 1 \cdot 0! = 0!$.

Combinazioni

3.3

In una gara di 10 cavalli indovinare i primi 3, senza considerare l'ordine, dunque non ci interessa se Fulmine è primo, Saetta secondo e Fiamma terzo, ci interessa invece se Fulmine, Saetta e Fiamma salgono sul podio. Quello che bisogna fare è pensare ad una disposizione senza ripetizione in cui le coppie che contengono gli stessi elementi, semplicemente mescolati in maniera differente, identificano in realtà la stessa coppia. Per esempio prendendo $d_{3,2}$ su $\{a, b, c\}$ avremo:

$$\begin{array}{ccc} ab & & ac \\ ba & & bc \\ & ca & cb \\ \hline \{a,b\} & \{a,c\} & \{b,c\} \end{array}$$

Dunque $\{a, b\}$, $\{a, c\}$, $\{b, c\}$ sono le combinazioni che ci interessano. Per ottenerle utilizzeremo la formula

$$\begin{aligned} C_{n,k} &= \frac{d_{n,k}}{P_k} \\ &= \frac{n!}{k!(n-k)!} \end{aligned} \quad (5)$$

essa può essere riscritta sfruttando il coefficiente binomiale

$$C_{n,k} = \binom{n}{k} \quad (6)$$

Riprendendo l'esempio di prima vediamo infatti come $C_{3,2} = \frac{6}{2} = 3$ e tornando all'esercizio dei cavalli basta svolgere $C_{10,3} = \frac{10!}{3!7!} = \frac{10 \cdot 9 \cdot 8 \cdot 7!}{3!7!} = \frac{10 \cdot 9 \cdot 8}{3!} = \frac{720}{6} = 120$.

Nel caso in cui ci trovassimo davanti ad una combinazione in cui però avviene una ripetizione, la formula sarà la seguente

$$C_{n,k} = \binom{n+k-1}{k} \quad (7)$$

Un esempio di applicazione di combinazione con ripetizione potrebbe essere il seguente:

Un'urna contiene 20 palline numerate. In quanti modi si possono estrarre 3 palline, supponendo che dopo ogni singola estrazione la pallina venga reimpressa nell'urna e senza tenere conto dell'ordine con cui le palline sono state estratte?

$$C_{n,k} = \binom{20+3-1}{3} = \binom{22}{3} = 1540 \quad (8)$$

Esercizi

3.4

- In quanti modi diversi possono sedersi 6 persone nei 6 posti di uno scompartimento ferroviario?
 $6!$
- Quanti anagrammi per la parola "matematica"?
 $\frac{10!}{2!3!2!}$
- Quante password da 5 cifre esistono?
 $D_{10,5} = 10^5 = 100000$
- Di queste password, se le prime 3 cifre sono uguali, quante ce ne sono?
Le prime tre cifre uguali possono essere 000, 111, ..., 999, dunque 10 possibili combinazioni. Le restanti due sono $D_{10,2}$, dunque: $10 \cdot 10^2 = 1000$
- Qual è il numero di password distinte da 5 cifre che hanno 3 cifre uguali in qualsiasi posizione consecutiva?
 $10 \cdot 1 \cdot 1 \cdot 10 \cdot 10 + 10 \cdot 10 \cdot 1 \cdot 1 \cdot 10 + 10 \cdot 10 \cdot 10 \cdot 1 \cdot 1 = 10^3 + 10^3 + 10^3 = 3 \cdot 10^3$
- Tra tutti i numeri di 9 cifre diverse tra loro e diverse da zero, quanti sono quelli le cui prime due sono, nell'ordine, 5 e 2?
 $7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 \cdot 1 \cdot 1$
- Un sistemista del Totocalcio vuole giocare tutte le possibili colonne in cui figurano sei segni 1, cinque segni X e due segni 2.
 $\frac{13!}{6!5!2!}$
- Scrivi tutti i numeri di tre cifre, le cui cifre sono tutte dispari e diverse tra loro
 $\frac{5!}{(5-3)!}$
- Tre coppie di amici vanno a cercare un ristorante. Viene riservato a loro un tavolo con 10 posti. In quanti modi si possono disporre se si vuole che le donne siedano al lato del tavolo più vicino al muro e gli uomini al lato opposto?
(donne · uomini) $\rightarrow \binom{5!}{2!} \cdot \binom{5!}{2!}$
- Quante diverse squadre di pallacanestro (che hanno 5 giocatori) è possibile selezionare da 16 ragazzi?
 $\frac{16!}{5!(16-5)!}$
- Si abbia una popolazione di 10 oggetti. Si estraggano senza ripetizione tutti i possibili campioni non ordinati di dimensione 4. Quanti campioni fanno parte dello spazio campionario?
 $\frac{10!}{4!(10-4)!}$

- **Ho 3 libri, 2 di matematica (M) e 1 di italiano (I), allora:**

- $3!$ ordina gli oggetti tra loro (es. {MIM, MMI, IMM, ...})
- $2! \cdot 1!$ ordina gli oggetti tra loro ma col primo gruppo posto davanti (ordinamento intra-gruppo e con precedenza di posizione, in questo contesto risulta {MMI, MMI})
- $(2! \cdot 1!) \cdot 2!$ ordina in tutte le combinazioni possibili di raggruppamenti (in questo contesto risulta {MMI, MMI, IMM, IMM})

- **Tra tutti i numeri di 10 cifre diverse tra loro quanti sono multipli di 100?**

Zero, un multiplo di 100 termina per forza con 00, due cifre uguali.

- **Tra tutti i numeri di 10 cifre tutte diverse tra loro, quanti sono quelli le cui 5 prime cifre sono dispari?**

Abbiamo 5 numeri dispari tutti diversi tra loro alla fine ($5 \cdot 4 \cdot 3 \cdot 2 \cdot 1$) e quindi prima mettiamo tutti i numeri rimanenti, diversi tra loro, ossia quelli pari che sono anch'essi 5 ($5 \cdot 4 \cdot 3 \cdot 2 \cdot 1$), dunque:

$$(5 \cdot 4 \cdot 3 \cdot 2 \cdot 1) * (5 \cdot 4 \cdot 3 \cdot 2 \cdot 1) = 14400$$

- **Tra tutti i numeri che possiamo formare con le cifre del numero 4550444, quanti sono i multipli di 10? E quanti sono i numeri pari?**

I multipli di 10 saranno tutti quei numeri che termineranno per 0 (quindi imponiamo un $\cdot 1$ alla fine). Le cifre a nostra disposizione sono 4 numeri 4 e 2 numeri 5, ossia abbiamo $6! \cdot 1 = 6!$ possibili scelte. Tuttavia scrivere 4550444 e 4550444, invertendo gli ultimi due numeri, è la stessa cosa, quindi ci si comporta come con gli anagrammi:

$$\frac{6!}{4!2!} = 15$$

Quanti sono invece i numeri pari componibili? Esattamente come prima il ragionamento, però questa volta consideriamo che l'ultimo numero può essere uno 0 oppure uno dei 4 possibili 4, dunque 5 possibilità tenendo conto che 4 di esse sono ripetute, quindi:

$$15 \cdot 5! / 4! = 75$$

- **Tra tutti i numeri di 3 cifre, tutte dispari e diverse tra loro, quanti sono i multipli di 5?**

I multipli di 5 sono tutti quei numeri che terminano con 0 o con 5, tuttavia se termina con 0 allora il numero di 3 cifre non è composto da tutte cifre dispari. Quindi consideriamo solo il caso in cui termini per 5 ($\cdot 1$) mentre le altre cifre possono essere qualsiasi numero dispari, escluso il 5 poiché già usato (dato che viene richiesto che le cifre siano tutte diverse tra loro). Infine avremo:

$$4 \cdot 3 \cdot 1 = 12$$

- **Quanti sono i possibili numeri di 5 cifre?**

La cifra più a sinistra di un numero di 5 cifre deve per forza essere un numero compreso da 1 e 9 (9 valor), gli altri invece sono numeri compresi tra 0 e 9 (10 valori), dunque:

$$9 \cdot 10 \cdot 10 \cdot 10 \cdot 10 = 90000$$

- **Quanti sono i numeri di 7 cifre, contenenti solo cifre pari escluso lo zero?**

Le cifre pari escluso lo zero sono 4 ({2, 4, 6, 8}). Dunque:

$$4 \cdot 4 \cdot 4 \cdot 4 \cdot 4 \cdot 4 \cdot 4 = 16384$$

- **Dodici amici dopo aver partecipato a una cena si salutano e ognuno stringe la mano a tutti gli altri, quante sono le strette di mano?**

$$\binom{12}{2} = 66$$

- **In una classe di 20 studenti si devono formare una squadra di calcio e una di pallacanestro. In quanti modi diversi si possono formare le due squadre se nessuno studente può appartenere a entrambe?**

Una squadra di calcio ha 11 giocatori mentre una squadra di pallacanestro ne ha 5. Calcolo per esempio prima quanti potrebbero giocare a calcio:

$$\binom{20}{11}$$

Ora ho scelto 11 persone, ne ho altre 9 da far giocare a basket:

$$\binom{9}{5}$$

Quindi in totale:

$$\binom{20}{11} \cdot \binom{9}{5} = 167960 \cdot 126 = 21162960$$

- **Si mescolano 12 carte, 3 vengono date a A, 3 a B, 3 a C e 3 a D, in quanti modi diversi può avvenire la distribuzione?**

$$\binom{12}{3} \cdot \binom{9}{3} \cdot \binom{6}{3} \cdot \binom{3}{3} = 369600$$

- **In quanti modi possiamo distribuire 7 cioccolatini identici indistinguibili a 4 bambini diversi?**

Possiamo vedere i bambini come delle urne e ciascuna urna riceve almeno un cioccolatino. Siccome i cioccolatini sono 7 e i bambini 4, significa che qualche bambino riceve più di un cioccolatino, dunque l'urna può ricevere una re-immissione. Dunque:

$$\binom{12 - 3 + 1}{3} = \binom{10}{3} = 120$$

Capitolo 4

Teoria della probabilità

Teoria degli insiemi

4.1

La frase "ho il 60% di probabilità di trovare il petrolio" può essere interpretata in tre chiavi differenti: quella soggettiva che ci dice che "il 60% delle volte ho trovato petrolio qui", quella frequentistica che ci suggerisce che "al 60% trovo il petrolio" e infine la chiave assiomatica che è quella della teoria degli insiemi.

4.1.1 Insieme universo

L'insieme universo Ω non è altro che l'insieme degli esiti possibili, lo spazio degli eventi. Per determinare il genere del nascituro $\Omega = \{M, F\}$, per una corsa di 5 cavalli $\Omega = \{\text{permutazioni di 7 elementi}\}$, ossia la cardinalità di Ω risulta $|\Omega| = 7!$, oppure per determinare il dosaggio minimo di un farmaco $\Omega = \mathbb{R}^+$.

4.1.2 Evento elementare

Definiamo un esito w come una delle $x \in \Omega$ tale che $\forall x \in \Omega$. Nel caso della nascita $w = \{F\}$, nel caso di una corsa di cavalli $w = \{(1, 2, 3, 4, 5)\}$, mentre nel caso delle medicine $\{42\}$.

4.1.3 Evento

Qualcosa di incerto, tale che $E \subseteq \Omega$, per esempio $\{F\}$ e $\{M\}$ nel caso del sesso, oppure $(3, 7]$ per il dosaggio di un farmaco, ma non $\{M, F\}$ in quanto evento certo e non \emptyset in quanto evento impossibile.

4.1.4 Algebra degli eventi

Definiamo $\mathcal{A} = \{\Omega, \emptyset\}$ e diciamo che $\bigcup_{i=1}^n E_i \in \mathcal{A}$ e che 2^Ω , anche noto come $P(\Omega)$ è l'insieme delle parti di Ω .

4.1.5 Funzione di probabilità

Data $P : \mathcal{A} \rightarrow \mathbb{R}$, definiamo tre assiomi principali (detti *Assiomi di Kolmogorov*):

(1) $\forall E \in \mathcal{A} \quad 0 \leq P(E) \leq 1$

(2) $P(\Omega) = 1$

(3) *Assioma di Kolmogorov*

$$\forall E_1, E_2, \dots, E_n \quad \forall_i E_i \in \mathcal{A}$$

$$\forall_{i \neq j} E_i \cap E_j = \emptyset \rightarrow P\left(\bigcup_{i=1}^n E_i\right) = \sum_{i=1}^n P(E_i)$$

4.1.6 Operazioni

Intersezione: $E \cap F$

Unione: $E \cup F$

Esclusione: $E \setminus F$

Complementare: $\bar{E} = \Omega \setminus E$ (oppure E^c)

Sottoinsieme: $E \subseteq F : \forall x \in \Omega \ x \in E \rightarrow x \in F$

AND: $E \subseteq F \wedge F \subseteq E \rightarrow E = F$

4.1.7 Proprietà di intersezione e unione

Commutativa: $E \cap F = F \cap E$ e $E \cup F = F \cup E$

Associativa: $(E \cap F) \cap G = E \cap (F \cap G)$ e $(E \cup F) \cup G = E \cup (F \cup G)$

Distributiva: $E \cup (F \cap G) = (E \cup F) \cap (E \cup G)$

Leggi di De Morgan: $\overline{E \cap F} = \bar{E} \cup \bar{F}$ e $\overline{E \cup F} = \bar{E} \cap \bar{F}$

4.1.8 Spazio delle probabilità

Definito come (Ω, \mathcal{A}, P) .

4.1.9 Spazi equiprobabili

Tutti gli eventi elementari hanno probabilità costante $p = \frac{1}{N}$.

Con Ω infinito non esistono spazi equiprobabili, con Ω con soli eventi elementari con probabilità nulla non ha senso dato che la loro somma deve fare 1. Invece per Ω finito abbiamo $\Omega = \{e_1, e_2, \dots, e_n\}$, dunque $P(\{e_1\}) = P(\{e_2\}) = \dots = P(\{e_n\}) = P$. In questo ultimo caso $\Omega = \bigcup_{i=1}^N \{e_i\}$ e dunque

$$\begin{aligned} 1 &= P(\Omega) \\ &= P\left(\bigcup_{i=1}^N \{e_i\}\right) \\ &= \sum_{i=1}^N P(\{e_i\}) \\ &= \sum_{i=1}^N p \\ &= Np \end{aligned} \tag{1}$$

Dunque, in un contesto equiprobabile

$$\begin{aligned}
 P(E) &= \sum_{j=1}^k P(\{e_j\}) \\
 &= \sum_{j=1}^k \frac{1}{N} \\
 &= \frac{k}{N} \\
 &= \frac{|E|}{N} \\
 &= \frac{\# \text{ casi favorevoli}}{\# \text{ casi possibili}}
 \end{aligned} \tag{2}$$

Esempio 1

Consideriamo un'urna con 6 palle bianche e 5 palle nere. L'urna è stata scossa e dunque si trova in un contesto equiprobabile e io devo estrarre 2 palle senza reimmissione: qual è la probabilità che io estragga due colori diversi? Per trovare il numero dei casi possibili consideriamo che ho una sequenza (quindi l'ordine conta) senza reimmissione, dunque si tratta di una disposizione senza ripetizione data da $d_{11,2} = 11 \cdot 10 = 110$.

Il numero dei casi favorevoli è dato dalla possibilità che io estragga due colori differenti: sapendo che ho 6 palle bianche e 5 nere, la probabilità è data da $6 \cdot 5 = 30$. Tuttavia l'ordine conta e non ho vincoli sull'estrarre prima una pallina bianca, quindi considero anche i casi in cui prima estraggo una pallina nera e poi una bianca, portandoci ad un totale di 60.

Dunque $P(\text{estraggo due colori diversi}) = \frac{|E|}{N} = \frac{60}{110} = \frac{6}{11}$.

Esempio 2

In una commissione di 5 persone ci sono 6 uomini e 9 donne. Se vengono tutti scelti con la stessa probabilità, qual è la probabilità di avere 3 uomini e 2 donne?

$$\begin{aligned}
 P(3U + 2D) &= \frac{|E|}{N} \\
 &= \frac{\binom{6}{3} \cdot \binom{9}{2}}{\binom{15}{5}} \\
 &= \frac{240}{1001}
 \end{aligned} \tag{3}$$

Teoremi fondamentali

4.2

4.2.1 Probabilità del complementare di un evento

Teorema

$$\forall E \in \mathcal{A} \quad P(\bar{E}) = 1 - P(E)$$

Dimostrazione

- (1) Sappiamo che $P(\Omega) = 1$ e quindi $1 = P(\Omega)$
- (2) Sappiamo che $E \cap \bar{E} = \emptyset$ e che $E \cup \bar{E} = \Omega$
- (3) $1 = P(\Omega)$ diventa $1 = P(E \cup \bar{E})$
- (4) Per l'assioma di Kolmogorov ottengo $1 = P(E \cup \bar{E}) = P(E) + P(\bar{E})$
- (5) Dunque riscriviamo $1 = P(E) + P(\bar{E})$ come $1 - P(E) = P(\bar{E})$

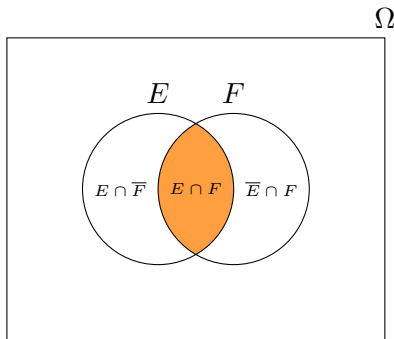
4.2.2 Probabilità dell'unione di due eventi

Teorema

$$\forall E, F \in \mathcal{A} \quad P(E \cup F) = P(E) + P(F) - P(E \cap F)$$

Dimostrazione

(1) Consideriamo il seguente



(2) $(E \cap \bar{F}) \cap (E \cap F) = (E \cap E) \cap (F \cap \bar{F})$

(3) $(E \cap E) \cap (F \cap \bar{F}) = E \cap \emptyset$

(4) $E \cap \emptyset = \emptyset \implies (E \cap \bar{F}) \cap (E \cap F) = \emptyset$

(5) $E \cup F = (E \cap \bar{F}) \cup (E \cap F) \cup (\bar{E} \cap F)$

(6) $P(E \cup F) = P((E \cap \bar{F}) \cup (E \cap F) \cup (\bar{E} \cap F))$

(7) Per l'assioma di Kolmogorov ottengo $P(E \cup F) = P(E \cap \bar{F}) + P(E \cap F) + P(\bar{E} \cap F)$

(8) $P(E \cup F) = P(E) + P(\bar{E} \cap F)$

(9) $P(E \cup F) = P(E) + P(\bar{E} \cap F) + P(E \cap F) - P(E \cap F)$

(10) $P(E \cup F) = P(E) + P(F) - P(E \cap F)$

Esempio

Sappiamo che il 28% degli americani fuma le sigarette (E) e che il 7% fuma i sigari (F). Il 5% degli americani invece, fuma sia sigarette che sigari ($E \cap F$). Qual è la percentuale di non fumatori?

$$\begin{aligned}
 P(\text{non fuma}) &= 1 - P(\overline{\text{non fuma}}) \\
 &= 1 - P(\text{fuma}) \\
 &= 1 - P(E \cup F) \\
 &= 1 - (P(E) + P(F) - P(E \cap F)) \\
 &= 1 - (0.28 + 0.07 - 0.05) \\
 &= 0.7 \implies 70\% \text{ non fuma}
 \end{aligned}
 \tag{4}$$

4.2.3 Probabilità di un evento impossibile

Teorema

$$P(\emptyset) = 0$$

Dimostrazione

(1) $P(\Omega) = 1$

(2) $P(\bar{E}) = 1 - P(E)$

$$(3) P(\emptyset) = P(\overline{\Omega})$$

$$(4) P(\emptyset) = 1 - P(\Omega) \text{ e dato che } P(\Omega) = 1 \text{ allora } P(\emptyset) = 0$$

Probabilità condizionata

4.3

Lanciando due dadi bilanciati, la probabilità che lanciando due dadi io ottenga 8, è data dal numero di casi favorevoli, ossia $\{(6, 2), (5, 3), (4, 4), (3, 5), (2, 6)\}$, quindi 5, diviso il numero di casi possibili, ossia $6 \cdot 6 = 36$. Arrivando così ad avere $P(l_1 + l_2 = 8) = \frac{5}{36}$.

Prendendo sempre come esperimento il lancio di due dadi bilanciati, so che il primo dado vale 3 ma qual è la probabilità $P(\text{somma} = 8)$? I casi possibili sono $\{(3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6)\}$ ma solo $(3, 5)$ è favorevole, quindi posso dire che

$$P(E | F) = \frac{P(E \cap F)}{P(F)} \quad (5)$$

Il che significa che, la probabilità tirato 3 al primo dado (F), allora (|) la somma dei dadi sia 8 (E), ossia equivale a dire che tutti gli eventi al di fuori di F non possono verificarsi. Dunque supponendo che $P(F) \neq 0$ e che infatti $P(F) = \frac{6}{36} = \frac{1}{6}$, e sapendo che $P(E \cap F) = P(\{3, 5\}) = \frac{1}{36}$, possiamo dire che $P(E | F) = \frac{P(E \cap F)}{P(F)} = \frac{\frac{1}{36}}{\frac{1}{6}} = \frac{1}{6}$.

Esempio

Avendo una confezione di 40 pennarelli di cui 5 guasti, 10 difettosi e 25 accettabili, dopo averli mescolati, mi chiedo qual è la probabilità che ne abbia pescato uno accettabile?

Definisco con E quelli accettabili e con F quelli non guasti (difettosi o accettabili)

$$\begin{aligned} P(E | F) &= \frac{P(\text{accettabile})}{1 - P(\text{guasto})} \\ &= \frac{\frac{25}{40}}{1 - \frac{5}{40}} \\ &= \frac{\frac{25}{40}}{\frac{35}{40}} \\ &= \frac{5}{7} \end{aligned} \quad (6)$$

4.3.1 Regola di fattorizzazione

Essa asserisce che

$$P(A \cap B) = P(A | B) \cdot P(B) \quad (7)$$

Esempio

Sappiamo che la probabilità che apra un nuovo ufficio è data da $P(U) = 0.3$, inoltre sappiamo che la probabilità che io diventi direttore dell'ipotetico nuovo ufficio è $P(M | U) = 0.6$, dunque la probabilità che apra un nuovo ufficio e che io diventi direttore è data da $P(U \cap M) = P(M | U) \cdot P(U) = 0.3 \cdot 0.6 = 0.18$

4.3.2 Teorema delle probabilità totali

Ipotizzo di avere due eventi $E, F \in \mathcal{A}$, allora

$$\begin{aligned}
 P(E) &= P((E \cap \bar{F}) \cup (E \cap F)) \\
 &= P(E \cap \bar{F}) + P(E \cap F) \\
 &= P(E | \bar{F}) \cdot P(\bar{F}) + P(E | F) \cdot P(F) \\
 &= P(E | F) \cdot P(F) + P(E | \bar{F}) \cdot P(\bar{F}) \\
 &= P(E | F) \cdot P(F) + P(E | \bar{F}) \cdot (1 - P(F))
 \end{aligned} \tag{8}$$

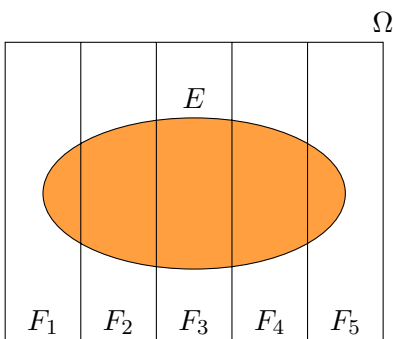
Esempio

Un cliente può essere incline ad un incidente nel prossimo anno con una probabilità $P(\text{incline}) = 0.4$ oppure non incline $P(\text{incline}) = 0.2$. I primi compongono il 30% dei clienti mentre i secondi il 70% dei clienti. Calcolare la probabilità che il prossimo anno avverrà un incidente.

Innanzitutto definiamo con A l'incidente nel prossimo anno e con H il cliente incline agli incidenti. Dati $P(H) = 0.3$, $P(\bar{H}) = 0.7$, $P(A | H) = 0.4$ e $P(A | \bar{H})$, possiamo dire che $P(A) = P(A | H) \cdot P(H) + P(A | \bar{H}) \cdot P(\bar{H}) = 0.4 \cdot 0.3 + 0.2 \cdot 0.7 = 0.26$.

4.3.3 Formula generalizzata del teorema delle probabilità totali

Ipotizziamo di avere una partizione di Ω con $F_1, F_2, \dots, F_n \subseteq \Omega$ le sue parti costituenti tali che $\bigcup_{i=1}^n F_i = \Omega$: $\forall i = j$ $F_i \cap F_j = \emptyset$.



$$\begin{aligned}
 P(E) &= P\left(\bigcup_{i=1}^n (E \cap F_i)\right) \\
 &= \sum_{i=1}^n P(E \cap F_i) \\
 &= \sum_{i=1}^m P(E | F_i) \cdot P(F_i)
 \end{aligned} \tag{9}$$

Esempio

	pezzi difettosi	% pezzi prodotti sul totale
A	2	60
B	3	30
C	4	10

Calcolo la probabilità di un pezzo difettoso, assegnando a D il pezzo difettoso e ad A il pezzo prodotto dal macchinario A , B dal macchinario B e C dal macchinario C .

$$P(D) = P(D | A) \cdot P(A) + P(D | B) \cdot P(B) + P(D | C) \cdot P(C) \text{ ossia } P(D) = 0.02 \cdot 0.6 + 0.03 \cdot 0.3 + 0.04 \cdot 0.1 = 0.025.$$

4.3.4 Teorema di Bayes

Ipotizzo un test diagnostico tramite analisi del sangue che funziona il 99% dei casi mentre l'1% dei casi ritorna falsi positivi. La malattia che si vuole analizzare affligge lo 0.5% della popolazione. L'esito positivo E , e il caso di malattia M , sono legati da $P(E | M) = 0.99$ (vero-positivo), $P(E | \bar{M}) = 0.01$ (falso-positivo) e $P(\bar{E} | \bar{M})$ (vero-negativo). Qual è la probabilità che io sia malato dato che l'esito è positivo? No, la risposta non è 99%, quella è la probabilità che l'esito sia positivo dato che sono malato. Se lo 0.5% = $\frac{1}{200}$ della popolazione soffre di questo male, in media su 200 persone vi sarà un solo malato. Se egli si sottopone alle analisi, verrà trovato positivo quasi certamente (con probabilità 0.99). D'altro canto le 199 persone sane hanno una probabilità di 0.01 di risultare positive e dunque $199 \cdot 0.01 = 1.99$ falsi positivi su 200. Dunque si ricava che la frazione di malati reali tra i soggetti positivi alle analisi è di $\frac{0.99}{0.99+1.99} \approx 0.3322$ e, per essere estensivi

$$\begin{aligned}
 P(M | E) &= \frac{P(M \cap E)}{P(E)} \\
 &= \frac{P(E \cap M)}{P(E)} \\
 &= \frac{P(E | M)P(M)}{P(E)} \\
 &= \frac{P(E | M)P(M)}{P(E | M)P(M) + P(E | \bar{M})P(\bar{M})} \\
 &= \frac{0.99 \cdot 0.005}{0.99 \cdot 0.005 + 0.01 \cdot 0.995} \approx 0.3322
 \end{aligned} \tag{10}$$

Trovandoci quindi nelle stesse ipotesi del teorema delle probabilità totali, possiamo definire il teorema di Bayes come

$$\begin{aligned}
 P(F_j | E) &= \frac{P(F_j \cap E)}{P(E)} \\
 &= \frac{P(E | F_j)P(F_j)}{\sum_{i=1}^n P(E | F_i)P(F_i)}
 \end{aligned} \tag{11}$$

Dunque non c'è necessariamente una relazione causa-effetto tra evento condizionato ed evento condizionante e con dovute manipolazioni si possono invertire.

Esempio 1

Ho un sospetto di colpevolezza $P(C) = 0.6$ e so che il sospettato è mancino, questa cosa influenza? So che il colpevole è mancino (al 100%) allora $P(M | C) = 1$, mentre la probabilità di essere mancini è $P(M) = 0.2$. Dunque $P(C | M) = \frac{P(M|C)P(C)}{P(M|\bar{C})P(\bar{C})+P(M|C)P(C)}$, ma possiamo vedere $P(M | \bar{C})$ come semplicemente $P(M)$ e quindi $P(C | M) \approx 0.87$.

Esempio 2

Un aereo è scomparso e si suppone che possa essere caduto in una qualsiasi di tre regioni, con uguale probabilità. Per $i = 1, 2, 3$, siano α_i le costanti che rappresentano la probabilità di non rinvenire il velivolo nella regione i -esima e $1 - \alpha_i$ la probabilità di rintracciare un velivolo che cada nella regione i -esima. Qual è la probabilità che l'aereo si trovi in ciascuna delle tre regioni se una ricerca della regione 1 ha dato esito negativo?

Denotiamo con R_i l'evento "il velivolo si trova nella regione i -esima" e sia E l'evento "la ricerca nella regione 1 non ha successo", dunque

$$\begin{aligned}
 P(R_1 | E) &= \frac{P(E | R_1)P(R_1)}{\sum_{i=1}^3 P(E | R_i)P(R_i)} = \frac{\frac{\alpha_1}{3}}{\frac{\alpha_1}{3} + \frac{1}{3} + \frac{1}{3}} = \frac{\alpha_1}{\alpha_1 + 2} \\
 P(R_2 | E) &= \frac{P(E | R_2)P(R_2)}{\sum_{i=1}^3 P(E | R_i)P(R_i)} = \frac{\frac{1}{3}}{\frac{\alpha_1}{3} + \frac{1}{3} + \frac{1}{3}} = \frac{1}{\alpha_1 + 2} \\
 P(R_3 | E) &= \frac{P(E | R_3)P(R_3)}{\sum_{i=1}^3 P(E | R_i)P(R_i)} = \frac{\frac{1}{3}}{\frac{\alpha_1}{3} + \frac{1}{3} + \frac{1}{3}} = \frac{1}{\alpha_1 + 2}
 \end{aligned} \tag{12}$$

Quindi se ad esempio fosse $\alpha_1 = 0.4$, la probabilità che il velivolo sia nella prima regione nonostante cercandolo lì non sia stato trovato, sarebbe di $\frac{1}{6}$.

4.3.5 Teorema Naive-Bayes

Ci piacciono i supereroi Marvel M e gli occhi neri N , vogliamo capire se $P(M | N) = \frac{P(N|M)P(M)}{P(N)}$. Guardo la frequenza con cui un supereroe Marvel ha gli occhi neri, se si avvicina a 1 allora tendenzialmente, se il supereroe ha gli occhi neri, allora è della Marvel, altrimenti no. Se però vogliamo prendere in considerazione il colore degli occhi O (neri o no) e il colore dei capelli C (neri o no), allora diventa

$$P(M | C = c \cap O = o) = \frac{P(C = c \cap O = o | M)P(M)}{P(C = c \cap O = o)} \quad (13)$$

questo può essere approssimato (e per questo naive), prendendo spunto dal terzo assioma di Kolmogorov, come

$$P(M | C = c \cap O = o) \approx \frac{P(C = c | M) \cdot P(O = o | M) \cdot P(M)}{P(C = c \cap O = o)} \quad (14)$$

Questo non accade sempre, però se ho m colori degli occhi e n colori dei capelli, allora ho $m \cdot n$ stime e mi conviene approssimare, perché almeno mi ritrovo con solo $m + n$ stime.

Dunque, formalizzando, date m possibili classi $Y = y_k$ (si verifica quando il supereroe appartiene alla classe k -esima) e n attributi (features) $X_i = x_i$ (si verifica quando l' i -esimo degli attributi assume un determinato valore), ottengo

$$\begin{aligned} P(Y = y_k | X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) &= \frac{P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n | Y = y_k)P(Y = y_k)}{P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)} \\ &\approx \frac{\prod_{i=1}^n P(X_i = x_i | Y = y_k)P(Y = y_k)}{P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)} \\ &\rightarrow \text{se } P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) \text{ non dipende da } k \\ &= \operatorname{argmax}_k \prod_{i=1}^n P(X_i = x_i | Y = y_k)P(Y = y_k) \end{aligned} \quad (15)$$

4.3.6 Indipendenza

Nel caso avessimo $P(E | F) = P(E)$ significa che, sapere che si è verificato F , non ci dice nulla di più su E . Definiamo questo concetto come *indipendenza* tra E e F , scritto $E \perp F$. Dato che $\frac{P(E \cap F)}{P(F)} = P(E)$ allora $P(E \cap F) = P(E) \cdot P(F)$. Di seguito alcune considerazioni:

(1) **Teorema**

E, F indipendenti $\implies E, \bar{F}$ indipendenti

Dimostrazione

(a) $E = (E \cap F) \cup (E \cap \bar{F})$

(b) $P(E) = P(E \cap F) + P(E \cap \bar{F})$

(c) $P(E \cap \bar{F}) = P(E) - P(E \cap F) = P(E) - P(E) \cdot P(F) = P(E) \cdot (1 - P(F)) = P(E) \cdot P(\bar{F})$

(2) Ipotizzo un esperimento in cui lancio due dadi. Sappiamo che la somma vale 7 (E) e che il primo vale 4 (F) mentre il secondo vale 3 (G), dunque $P(E) = P(F) = P(G) = \frac{1}{6}$ e $P(E \cap F) = P(E \cap G) = \frac{1}{36}$. Però cosa posso dire di $P(E | F \cap G)$? Beh, $P(E | F \cap G) = 1$, dato che $4 + 3 = 7$, ma come estendo questo concetto?

Dati E, F, G indipendenti, allora $P(E \cap F \cap G) = P(E) \cdot P(F) \cdot P(G)$.

(3) Ipotizzo di avere n eventi E_1, \dots, E_n che sono indipendenti se e solo se $\forall r = 2, \dots, n \forall 1 \leq \alpha_1 < \dots < \alpha_r \leq n \quad P(\bigcap_{i=1}^r E_{\alpha_i}) = \prod_{i=1}^r P(E_{\alpha_i})$. Questo concetto lo posso applicare per esempio ai sistemi in serie e in parallelo.

In un sistema in serie $\forall i = 1, \dots, n \quad p_i = P(\text{l}'i\text{-esimo componente funziona})$, dunque $P(\text{il sistema funziona}) =$

P (tutti i componenti funzionano) e quindi $P(\bigcap_{i=1}^n \text{componente } i\text{-esimo funziona}) = \prod_{i=1}^n p_i$.
 Nel caso del sistema in parallelo invece P (il sistema funziona) = $1 - P$ (il sistema non funziona) (ossia il complementare). Siccome è in parallelo, anche se un componente non funziona non importa, tutti i componenti non devono andare per far sì che il sistema non funzioni, dunque

$$\begin{aligned}
 P(\text{il sistema funziona}) &= 1 - P(\text{il sistema non funziona}) \\
 &= 1 - P(\text{tutti i componenti non funzionano}) \\
 &= 1 - P\left(\bigcap_{i=1}^n \text{componente } i\text{-esimo non funziona}\right) \\
 &= 1 - \prod_{i=1}^n P(\text{componente } i\text{-esimo non funziona}) \\
 &= 1 - \prod_{i=1}^n (1 - p_i)
 \end{aligned} \tag{16}$$

Esempio 1

Pesco una carta da un mazzo di 52 carte mischiato in maniera equiprobabile. La probabilità che io peschi un asso è data da $P(A) = \frac{4}{52}$ mentre la probabilità che io peschi cuori è data da $P(C) = \frac{13}{52}$. Siccome $A \perp C$, la probabilità che io peschi un asso di cuori è $P(A \cap C) = P(A) \cdot P(C) = \frac{4}{52} \cdot \frac{13}{52} = \frac{1}{52}$.

Esempio 2

Considero E e $F \cup G$ dunque

$$\begin{aligned}
 P(E \cap (F \cup G)) &= P((E \cap F) \cup (E \cap G)) \\
 &= P(E \cap F) + P(E \cap G) - P((E \cap F) \cap (E \cap G)) \\
 &= P(E)P(F) + P(E)P(G) - P(E \cap F \cap G) \\
 &= P(E)P(F) + P(E)P(G) - P(E)P(F)P(G) \\
 &= P(E)(P(F) + P(G) - P(F)P(G)) \\
 &= P(E)(P(F) + P(G) - P(F \cap G)) \\
 &= P(E)P(F \cup G)
 \end{aligned} \tag{17}$$

Esempio 3

Considero una raccolta di figurine numerate da 1 a n . Per $j = 1, 2, \dots, n$ sia p_j la probabilità di ottenere la figurina j , con $\sum_{j=1}^n p_j = 1$. Determinare la probabilità che una collezione di k figurine ottenute indipendentemente contenga almeno un esemplare di quella numero j , sapendo che ne contiene almeno uno di quella numero i .

Iniziamo col denotare A_r l'evento che la collezione contenga almeno una figurina numero r , allora per definizione $P(A_j | A_i) = \frac{P(A_j \cap A_i)}{P(A_i)}$. Da questo possiamo calcolare $P(A_i)$ come

$$\begin{aligned}
 P(A_i) &= 1 - P(\overline{A_i}) \\
 &= 1 - P(\text{nessuna figurina ha numero } i) \\
 &= 1 - P\left(\bigcap_{r=1}^k \text{all}'r\text{-esimo acquisto nessuna figurina ha numero } i\right) \\
 &= 1 - \prod_{i=1}^n P(\text{all}'r\text{-esimo acquisto nessuna figurina ha numero } i) \\
 &= 1 - \prod_{i=1}^n (1 - p_i) \\
 &= 1 - (1 - p_i)^k
 \end{aligned} \tag{18}$$

e possiamo calcolare $P(A_j \cap A_i)$ come

$$\begin{aligned}
P(A_j \cap A_i) &= 1 - P(\overline{A_j \cap A_i}) \\
&= 1 - P(\overline{A_j} \cup \overline{A_i}) \\
&= 1 - (P(\overline{A_j}) + P(\overline{A_i}) - P(\overline{A_j} \cap \overline{A_i})) \\
&= 1 - ((1 - p_j)^k + (1 - p_i)^k - P(\overline{A_j} \cap \overline{A_i})) \\
&= 1 - ((1 - p_j)^k + (1 - p_i)^k - P(\text{mai né i né j})) \\
&= 1 - \left((1 - p_j)^k + (1 - p_i)^k - P\left(\bigcap_{r=1}^k \text{all'r-esimo acquisto né i né j}\right) \right) \\
&= 1 - \left((1 - p_j)^k + (1 - p_i)^k - \prod_{r=1}^k (1 - p_i - p_j) \right) \\
&= 1 - ((1 - p_j)^k + (1 - p_i)^k - (1 - p_i - p_j)^k) \\
&= 1 - (1 - p_j)^k - (1 - p_i)^k + (1 - p_i - p_j)^k
\end{aligned} \tag{19}$$

e per concludere possiamo quindi dire che

$$P(A_j | A_i) = \frac{1 - (1 - p_j)^k - (1 - p_i)^k + (1 - p_i - p_j)^k}{1 - (1 - p_i)^k} \tag{20}$$

Variabile aleatoria

4.4

Una variabile aleatoria è definita in uno spazio di probabilità (Ω, \mathcal{A}, P) con $X : \Omega \rightarrow \mathbb{R}$, ossia le immagini di X sono una codifica numerica dei possibili esiti di un esperimento aleatorio. L'evento $X = \alpha$ si verificherà quando avrò estratto un evento di Ω tale che $X(\text{evento}) = \alpha$, per l'esattezza

$$\{X = \alpha\} \equiv \{\omega \in \Omega : X(\omega) = \alpha\} \tag{21}$$

o meglio

$$P(X = \alpha) \tag{22}$$

In particolare definiamo con **specificazione** il valore dell'evento permettendoci di dire che essa determina l'occorrenza di un evento X .

4.4.1 Esempio

Considero l'esperimento del lancio di due dadi: Ω è il lancio dei due dadi, X è la somma dei due esiti (tutti i numeri da 2 a 12), ma bisogna chiedersi se X assumerà i suoi valori con la stessa probabilità. La risposta è no. Prendo per esempio $X = 3$

$$\{X = 3\} \equiv \{\omega \in \Omega : X(\omega) = 3\} = \{(1, 2), (2, 1)\} \implies P(X = 3) = \frac{2}{36} \tag{23}$$

Ora osservo $X = 9$

$$\{X = 9\} \equiv \{\omega \in \Omega : X(\omega) = 9\} = \{(3, 6), (4, 5), (5, 4), (6, 3)\} \implies P(X = 9) = \frac{4}{36} \tag{24}$$

Si può notare come il picco sia a $P(X = 7) = \frac{6}{36}$ e andando verso il 2 o verso il 12 i numeri diminuiscano fino a raggiungere $\frac{1}{36}$ da entrambi i lati. Inoltre evidenziamo come

$$\begin{aligned}
\sum_{i=2}^{12} P(X = i) &= P\left(\bigcup_{i=2}^{12} \{X = i\}\right) \\
&= P(\Omega) \\
&= 1
\end{aligned} \tag{25}$$

Definiamo un'altra variabile aleatoria a partire dallo stesso esperimento del lancio dei due dadi e chiamiamola I , ossia l'esito del primo lancio. E notiamo che $P(\mathbb{I} = i) = \frac{1}{6} \quad \forall i = 1, \dots, 6$. Dunque chiamiamo la funzione indicatrice (o *caratteristica*), la funzione tale che, dato un evento $A \subseteq \mathbb{R}$ e $I_A : \mathbb{R} \rightarrow \{0, 1\}$ allora

$$I_A(x) = \begin{cases} 1 & \text{se } x \in A \\ 0 & \text{se } x \notin A \end{cases} \quad (26)$$

e quindi $P(\mathbb{I} = i) = \frac{1}{6} \mathbb{I}_{\{1, \dots, 6\}}(i)$.

4.5.1 Esempio 1

Acquisto 2 componenti elettronici: la probabilità che uno sia difettoso è $P(d) = 0.3$ e la probabilità che uno sia funzionante è $P(f) = 0.7$. Nell'ipotesi in cui ci sia indipendenza tra gli eventi, allora

$$\Omega = \{(d, d), (d, f), (f, d), (f, f)\} \quad (27)$$

e dunque, identificando con X il numero di componenti funzionanti (X può assumere valori da 0 a 2 ovviamente)

$$\begin{aligned} P(X = 0) &= \{(d, d)\} = 0.3 \cdot 0.3 = 0.09 \\ P(X = 1) &= \{(d, f), (f, d)\} = (0.3 \cdot 0.7) + (0.7 \cdot 0.3) = 0.42 \\ P(X = 2) &= \{(f, f)\} = 0.7 \cdot 0.7 = 0.49 \end{aligned} \quad (28)$$

o più precisamente possiamo costruire la funzione indicatrice di un evento che ci dice se almeno un componente funziona

$$\mathbb{I} = \begin{cases} 1 & \text{se almeno un componente funziona} & \rightarrow P(\mathbb{I} = 1) = 0.42 + 0.49 = 0.91 \\ 0 & \text{altrimenti} & \rightarrow P(\mathbb{I} = 0) = 0.09 \end{cases} \quad (29)$$

4.5.2 Esempio 2

Immaginiamo di avere tre eventi A , B e C , e voglio definire una funzione indicatrice che restituisca 1 se e solo se almeno due degli eventi si verificano contemporaneamente. Posso scrivere questa funzione indicatrice come $I_{(A \cap B) \cup (A \cap C) \cup (B \cap C)}(x) = 1$ se x appartiene a uno degli insiemi dati.

Anche nota come *f.d.c.* (o *cdf* su Python) essa è definita per $F_x : \mathbb{R} \rightarrow [0, 1]$, ci dice che $\forall x \in \mathbb{R} \quad F_X(x) = P(X \leq x)$. Una particolare variabile aleatoria è distribuita seconda una particolare funzione di ripartizione, così da avere $X \sim F_X$, nel dettaglio

$$\{X \leq b\} = \{X \leq a\} \cup \{a < X \leq b\} \quad (30)$$

e applicando il terzo assioma di Kolmogorov

$$\begin{aligned} P(X \leq b) &= P(X \leq a) + P(a < X \leq b) \\ F_X(b) &= F_X(a) + P(a < X \leq b) \\ P(a < X \leq b) &= F_X(b) - F_X(a) \end{aligned} \quad (31)$$

Variabili aleatorie con un numero finito o numerabile di valori possibili sono dette discrete.

4.7.1 Funzione di massa di probabilità

$$p_X : \mathbb{R} \rightarrow [0, 1] \quad \forall x \in \mathbb{R} \quad p_X(x) = P(X = x)$$

Proprietà:

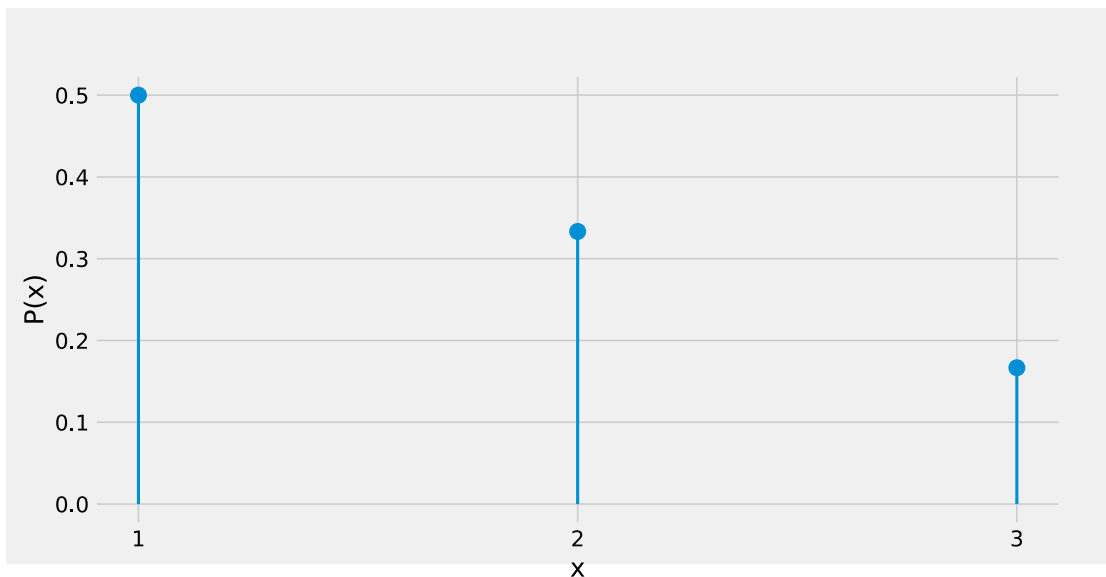
$$(1) \quad p_X(x) \geq 0 \quad \forall x \in \mathbb{R}$$

$$(2) \quad \sum_{i=1}^n p_X(x_i) = 1$$

Esempio

Ipotizzo di avere una variabile aleatoria X discreta tale che $D_X = \{1, 2, 3\}$, $P(X = 1) = \frac{1}{2}$ e $P(X = 2) = \frac{1}{3}$. Mi domando quindi, quanto vale $P(X = 3)$?

$$\begin{aligned} p_X(1) + p_X(2) + p_X(3) &= 1 \\ p_X(3) &= 1 - \frac{1}{2} - \frac{1}{3} = \frac{1}{6} \end{aligned} \tag{32}$$



4.7.2 Funzione di ripartizione

Sappiamo che

$$\begin{aligned} F_X(x) &= P(X \leq x) \\ &= P\left(\bigcup_{a \leq x} \{X = a\}\right) \\ &= \sum_{a \leq x} P(X = a) \\ &= \sum_{a \leq x} p_X(a) \end{aligned} \tag{33}$$

Inoltre

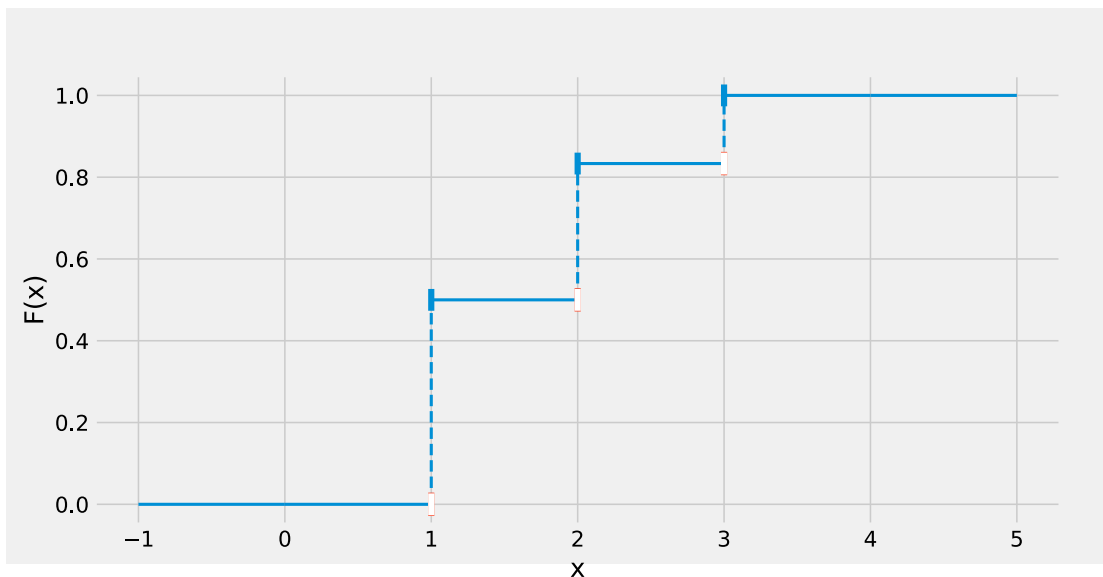
$$\begin{aligned} \forall x \in \mathbb{R} \quad F_X(x) &\geq 0 \\ \lim_{x \rightarrow \infty} F_X(x) &= 1 \end{aligned} \tag{34}$$

Quindi F_X continua da destra, ma non continua da sinistra.

Esempio

Prendo l'esempio di prima in cui ipotizzo di avere una variabile aleatoria X discreta tale che $D_X = \{1, 2, 3\}$, $P(X = 1) = \frac{1}{2}$ e $P(X = 2) = \frac{1}{3}$.

$$F_X(x) = \begin{cases} 0 & \text{se } x \leq 1 \\ \frac{1}{2} & \text{se } 1 \leq x \leq 2 \\ \frac{5}{6} & \text{se } 2 \leq x \leq 3 \\ 1 & \text{se } x \geq 3 \end{cases} \quad (35)$$



Attenzione

Nel capitolo dei modelli discreti, per semplicità pratica, uso degli step-plot. Tuttavia è questo qui sopra il grafico preciso da utilizzare quando si ha a che fare con funzioni di ripartizione discrete.

Valore atteso

4.8

Il valore atteso di una variabile aleatoria X è definito come

$$\mathbb{E}(X) = \sum_{i=1}^n x_i P(X = x_i) \quad (36)$$

Esempio 1

Dato X il punteggio di un dado bilanciato e $D_X = \{1, \dots, 6\}$, $p_X(x) = \frac{1}{6}I_{\{1, \dots, 6\}}(x)$, allora

$$\begin{aligned} \mathbb{E}(X) &= \sum_{x=1}^6 x \cdot p_X(x) \\ &= \sum_{x=1}^6 x \frac{1}{6} \\ &= \frac{1}{6} \sum_{x=1}^6 x \end{aligned} \quad (37)$$

e sapendo che $\sum_{x=1}^n x = \frac{n(n+1)}{2}$, allora $\mathbb{E}(X) = \frac{1}{6} \sum_{x=1}^6 x = \frac{1}{6} \cdot \frac{6 \cdot 7}{2} = \frac{7}{2}$.

Esempio 2

Considerando X la mia vincita ad un gioco d'azzardo e sapendo che gioco n volte, la mia vincita media non è altro che $\frac{n_1x_1+\dots+n_kx_k}{n}$ ossia

$$\sum_{i=1}^k \frac{n_i}{n} x_i \approx \sum_{i=1}^k P(X = x_i) \cdot x_i \quad (38)$$

Esempio 3

Avendo un evento A e sapendo che $D_{I_A} = \{0, 1\}$, con I_A che vale 0 se A non si verifica o 1 se A si verifica, allora $\mathbb{E}(I_A) = P(A)$, dato che la probabilità che la variabile aleatoria sia uguale a 1 avviene quando A si verifica.

Esempio 4

Consideriamo $g(x) = y$ ossia $g(x) = x^2$ e dunque $x = \sqrt{y}$

x	$P(X = x)$	y
0	0.2	0
1	0.5	1
2	0.3	4

Possiamo quindi dire che

$$\begin{aligned} \mathbb{E}(y) &= 0 \cdot 0.2 + 1 \cdot 0.5 + 4 \cdot 0.3 = 1.7 \\ \mathbb{E}(x^2) &= 0^2 \cdot 0.2 + 1^2 \cdot 0.5 + 2^2 \cdot 0.3 = 1.7 \end{aligned} \quad (39)$$

e più in generale

$$\begin{aligned} \mathbb{E}(g(x)) &= \sum_i g(x_i) \cdot P(X = x_i) \\ &= \sum_i g(x_i) p_X(x_i) \end{aligned} \quad (40)$$

Esempio 5

Un'urna contiene 3 palline contrassegnate dai numeri $\{1, 2, 3\}$. Si estrae con riposizione un campione di ampiezza 2. Sia Y la variabile casuale che esprime la media aritmetica dei numeri riportati sulle palline estratte. Calcolare il valore atteso di Y .

Le possibili combinazioni sono

$$\{(1, 1), (1, 2), (1, 3), (2, 1), (2, 2), (2, 3), (3, 1), (3, 2), (3, 3)\}$$

Ora calcoliamo la media aritmetica Y per ciascuna combinazione

$$\begin{aligned}
 Y_1 &= \frac{1+1}{2} = 1 \\
 Y_2 &= \frac{1+2}{2} = 1.5 \\
 Y_3 &= \frac{1+3}{2} = 2 \\
 Y_4 &= \frac{2+1}{2} = 1.5 \\
 Y_5 &= \frac{2+2}{2} = 2 \\
 Y_6 &= \frac{2+3}{2} = 2.5 \\
 Y_7 &= \frac{3+1}{2} = 2 \\
 Y_8 &= \frac{3+2}{2} = 2.5 \\
 Y_9 &= \frac{3+3}{2} = 3
 \end{aligned} \tag{41}$$

Da qui possiamo per esempio vedere che il valore 1 compare 1 volta, il valore 1.5 compare 2 volte e così via, dunque, data la formula del valore atteso, abbiamo

$$\begin{aligned}
 \mathbb{E}(X) &= \sum_{i=1}^n x_i P(X = x_i) \\
 &= 1 \cdot \frac{1}{9} + 1.5 \cdot \frac{2}{9} + 2 \cdot \frac{3}{9} + 2.5 \cdot \frac{2}{9} + 3 \cdot \frac{1}{9} \\
 &= \frac{18}{9} \\
 &= 2
 \end{aligned} \tag{42}$$

4.8.1 Considerazione

Dato $g(x) = ax + b$, allora

$$\begin{aligned}
 \mathbb{E}(g(x)) &= \mathbb{E}(ax + b) \\
 &= \sum_i (ax_i + b) \cdot P(X = x_i) \\
 &= \sum_i (ax_i P(X = x_i) + bP(X = x_i)) \\
 &= a \sum_i x_i P(X = x_i) + b \sum_i P(X = x_i) \\
 &= a\mathbb{E}(X) + b
 \end{aligned} \tag{43}$$

E dunque se $a = 0$ allora $\mathbb{E}(b) = b$, mentre se $b = 0$ allora $\mathbb{E}(aX) = a\mathbb{E}(X)$.

Varianza

4.9

Possiamo vedere quanto una X si discosta dal valore atteso, dunque $g(X) = |X - \mathbb{E}(X)|$. Tuttavia cerchiamo di togliere il valore assoluto, ottenendo quindi la varianza di X come

$$\text{Var}(X) = \mathbb{E}((X - \mathbb{E}(X))^2) \tag{44}$$

Da qui in poi useremo nuove nomenclature quali $\mu = \mathbb{E}X$ e $Var(X) = \mathbb{E}((X - \mu)^2)$. Inoltre

$$\begin{aligned}
 Var(X) &= \mathbb{E}((X - \mu)^2) \\
 &= \mathbb{E}(X^2 - 2\mu X + \mu^2) \\
 &= \mathbb{E}(X^2) - 2\mu\mathbb{E}(X) + \mu^2 \\
 &= \mathbb{E}(X^2) - 2\mu^2 + \mu^2 \\
 &= \mathbb{E}(X^2) - \mu^2 \\
 &= \mathbb{E}(X^2) - \mathbb{E}(X)^2
 \end{aligned} \tag{45}$$

Proseguendo posso anche notare che

$$\begin{aligned}
 Var(ax + b) &= \mathbb{E}(((ax + b) - (\mathbb{E}(ax + b))))^2) \\
 &= \mathbb{E}((ax + b - a\mathbb{E}(x) - b)^2) \\
 &= \mathbb{E}((ax - a\mathbb{E}(x))^2) \\
 &= \mathbb{E}(a^2(x - \mathbb{E}(x))^2) \\
 &= a^2\mathbb{E}((x - \mathbb{E}(x))^2) \\
 &= a^2Var(X)
 \end{aligned} \tag{46}$$

e che, considerando $\mathbb{I}_A = 1$ se A si verifica, 0 altrimenti, allora

$$\begin{aligned}
 Var(\mathbb{I}_A) &= \mathbb{E}(\mathbb{I}_A^2) - \mathbb{E}(\mathbb{I}_A)^2 \\
 &= \mathbb{E}(\mathbb{I}_A) - \mathbb{E}(\mathbb{I}_A)^2 \\
 &= \mathbb{E}(\mathbb{I}_A)(1 - \mathbb{E}(\mathbb{I}_A)) \\
 &= P(A) \cdot P(\bar{A})
 \end{aligned} \tag{47}$$

4.9.1 Esempio 1

Sapendo che $p_X(x) = \frac{1}{6}\mathbb{I}_{D_X}(x)$ e che $\mu = \frac{7}{2}$, allora

$$\begin{aligned}
 Var(X) &= \mathbb{E}\left(\left(X - \frac{7}{2}\right)^2\right) \\
 &= \sum_{x=1}^6 \left(\left(X - \frac{7}{2}\right)^2\right) \\
 &= \dots
 \end{aligned} \tag{48}$$

ma è troppo lungo da fare, quindi... (ricordiamoci anche di essere in un contesto di lancio di dadi equiprobabile)

$$\begin{aligned}
 Var(X) &= \mathbb{E}(X^2) - \mathbb{E}(X)^2 \\
 &= \mathbb{E}(X^2) - \left(\frac{7}{2}\right)^2 \\
 &= \sum_{x=1}^6 x^2 \frac{1}{6} - \left(\frac{7}{2}\right)^2 \\
 &= \frac{1}{6} \sum_{x=1}^6 x^2 - \left(\frac{7}{2}\right)^2 \\
 &= \frac{1}{6} \cdot \frac{6 \cdot 7 \cdot 13}{6} - \left(\frac{7}{2}\right)^2 \\
 &= \frac{91}{6} - \frac{49}{4} = \frac{35}{12} = \frac{36 - 1}{12}
 \end{aligned} \tag{49}$$

4.9.2 Deviazione standard

Non è altro che

$$\sigma_X = \sqrt{\text{Var}(X)} \quad (50)$$

Funzione di massa di probabilità congiunta

4.10

Considero $p_{X,Y}(x, y) = P(X = x, Y = y)$ ossia $\bigcup_j \{X = x_i, Y = y_j\} = \{X = x_i\}$. In egual modo distribuzione marginale risulta

$$\sum_j p_{X,Y}(x_i, y_j) = p_X(x_i) \quad (51)$$

4.10.1 Considerazione

Questa tipologia di funzione viene impiegata per campioni bivariati di variabili aleatorie discrete (una funzione di densità viene impiegata nel caso di variabili aleatorie continue). Risulta necessario sottolineare che un campione bivariato non è un campione bimodale, infatti:

- **Campione bimodale**

- **Scenario:** Misurare l'altezza di una popolazione composta da uomini e donne.
- **Descrizione:** Se la popolazione ha due sottogruppi distinti, ad esempio uomini e donne, la distribuzione delle altezze potrebbe presentare due picchi distinti, uno per ciascun gruppo.
- **Evento:** Le altezze delle donne potrebbero concentrarsi attorno a un certo valore, mentre quelle degli uomini attorno a un altro, generando un campione con due mode.

- **Campione bivariato**

- **Scenario:** Misurare il numero di ore di studio e i punteggi ottenuti in un esame per uno studio di studenti.
- **Descrizione:** Considerando due variabili casuali, come il numero di ore di studio (X) e i punteggi in un esame (Y), il campione bivariato mostrerebbe la relazione tra queste due variabili.
- **Evento:** Ad esempio, potrebbe emergere che uno studente ha una maggiore probabilità di ottenere un punteggio elevato se ha studiato molte ore, creando una distribuzione di probabilità congiunta.

Funzione di ripartizione congiunta

4.11

Data una coppia X, Y , la funzione $F_{X,Y}(x, y) = P(X \leq x, Y \leq y)$. La distribuzione marginale è invece nota come

$$\begin{aligned} \lim_{y \rightarrow \infty} F_{X,Y}(x, y) &= \lim_{y \rightarrow \infty} P(X \leq x, Y \leq y) \\ &= P(X \leq x) \\ &= F_X(x) \end{aligned} \quad (52)$$

Noto che i profili delle funzioni di ripartizione discrete sono a gradini.

4.11.1 Esempio

Se volessimo calcolare la probabilità che il dado X assuma un valore minore o uguale a 3 e contemporaneamente il dado Y assuma un valore minore o uguale a 4, avremmo $F_{X,Y}(3, 4) = P(X \leq 3, Y \leq 4)$. Questa probabilità dipenderà dalla distribuzione congiunta dei due dadi, considerando tutte le combinazioni possibili dei risultati.

X e Y sono indipendenti se e solo se $\forall A, B \in \mathbb{R} \quad P(X \in A, Y \in B) = P(X \in A) \cdot P(Y \in B)$ (il fatto che $4 < 5$ non ha nulla a che fare col fatto che 4 sia pari). Ciò equivale a dimostrare che:

(1) $\forall a, b \in \mathbb{R} \quad F_{X,Y}(a, b) = F_X(a)F_Y(b)$

(2) $\forall a, b \in \mathbb{R} \quad p_{X,Y}(a, b) = p_X(a)p_Y(b)$

quindi:

(1) Dimostro che $P(X \in A, Y \in B) = P(X \in A) \cdot P(Y \in B) \rightarrow p_{X,Y}(a, b) = p_X(a)p_Y(b)$:
 $A = \{a\}$ e $B = \{b\}$

(2) Dimostro che $p_{X,Y}(a, b) = p_X(a)p_Y(b) \rightarrow \forall A, B \in \mathbb{R} \quad P(X \in A, Y \in B) = P(X \in A) \cdot P(Y \in B)$:

$$\begin{aligned} P(X \in A, Y \in B) &= \sum_{a \in A} \sum_{b \in B} p_{X,Y}(a, b) \\ &= \sum_{a \in A} \sum_{b \in B} p_X(a)p_Y(b) \\ &= \sum_{a \in A} p_X(a) \sum_{b \in B} p_Y(b) \\ &= P(X \in A) \cdot P(Y \in B) \end{aligned} \tag{53}$$

Inoltre, se ho un vettore di variabili aleatorie X_1, X_2, \dots, X_n indipendenti, allora ne consegue

$$\begin{aligned} F_{X_1, X_2, \dots, X_n}(X_1, X_2, \dots, X_n) &= P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n) \\ &= \prod_{i=1}^n F_{X_i}(X_i) \end{aligned} \tag{54}$$

e

$$\begin{aligned} p_{X_1, X_2, \dots, X_n}(X_1, X_2, \dots, X_n) &= P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) \\ &= P\left(\bigcap_{i=1}^n X_i \in A_i\right) \\ &= \prod_{i=1}^n P(X_i \in A_i) \end{aligned} \tag{55}$$

4.12.1 Considerazione

Per riassumere, la dipendenza o indipendenza tra variabili aleatorie in una distribuzione congiunta dipende dalla natura specifica della distribuzione. Due variabili aleatorie X e Y si dicono indipendenti nella distribuzione congiunta se la probabilità congiunta è il prodotto delle probabilità marginali.

Cosa succede quando calcolo il valore atteso di una funzione di variabili aleatorie?

Per una variabile aleatoria X e funzione $g : \mathbb{R} \rightarrow \mathbb{R}$

$$\mathbb{E}(X) = \sum_i x_i P(X = x_i) \rightarrow \mathbb{E}(g(x)) = \sum_i g(x_i) P(X = x_i) \tag{56}$$

Per variabili aleatorie X, Y e funzione $f : \mathbb{R}^2 \rightarrow \mathbb{R}$

$$\mathbb{E}(f(X, Y)) = \sum_{i,j} f(x_i, y_j) P(X = x_i, Y = y_j) \tag{57}$$

Esempio 1

Dato $f(x, y) = x + y$, allora

$$\begin{aligned}\mathbb{E}(X + Y) &= \sum_{i=1}^n \sum_{j=1}^m (x_i + y_j) P(X = x_i, Y = y_j) \\ &= \sum_{i=1}^n \sum_{j=1}^m x_i P(X = x_i, Y = y_j) + \sum_{i=1}^n \sum_{j=1}^m y_j P(X = x_i, Y = y_j) \\ &= \sum_{i=1}^n x_i \sum_{j=1}^m P(X = x_i, Y = y_j) + \sum_{j=1}^m y_j \sum_{i=1}^n P(X = x_i, Y = y_j) \\ &\rightarrow \sum_{j=1}^m P(Y = y_j) = \Omega \\ &= \sum_{i=1}^n x_i P(X = x_i) + \sum_{j=1}^m y_j P(Y = y_j) \\ &= \mathbb{E}(X) + \mathbb{E}(Y)\end{aligned}\tag{58}$$

dunque

$$\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)\tag{59}$$

Esempio 2

Il valore atteso della somma dei risultati di due dadi non truccati $\mathbb{E}(X) = \sum_2^{12} iP(X = i)$ con $X = X_1 + X_2$ non è altro che

$$\begin{aligned}\mathbb{E}(X) &= \mathbb{E}(X_1 + X_2) \\ &= \mathbb{E}(X_1) + \mathbb{E}(X_2) \\ &= \frac{7}{2} + \frac{7}{2} \\ &= 7\end{aligned}\tag{60}$$

Esempio 3

Ho n lettere e n buste accoppiate a caso, quanti abbinamenti corretti riusciamo ad avere tale che alla busta n corrisponda l'esatta lettera n ? Stabiliamo che X_i vale 1 se l' i -esima lettera è soddisfatta la nostra richiesta, 0 altrimenti, dunque

$$\begin{aligned}\mathbb{E}\left(\sum_{i=1}^n X_i\right) &= \sum_{i=1}^n \mathbb{E}(X_i) \\ &= \sum_{i=1}^n \frac{1}{n} \\ &= 1\end{aligned}\tag{61}$$

Esempio 4

Ci sono 20 buoni differenti e io ho comprato 10 confezioni, ognuna con un buono al suo interno: quanti buoni diversi avrò? Stabilisco che X_i vale 1 se ho almeno 1 buono i -esimo, 0 altrimenti, dunque

$$\begin{aligned}\mathbb{E}\left(\sum_{i=1}^{20} X_i\right) &= \sum_{i=1}^{20} \mathbb{E}(X_i) \\ &= \sum_{i=1}^{20} P(X_i = 1) \\ &= \sum_{i=1}^{20} P(\text{almeno un esemplare del buono } i\text{-esimo}) \\ &\rightarrow \text{"almeno"} \implies \text{ragiono con l'evento complementare} \\ &= \sum_{i=1}^{20} (1 - P(\text{in nessuna delle 10 confezioni compare il buono } i\text{-esimo})) \\ &= \sum_{i=1}^{20} \left(1 - P\left(\bigcap_{j=1}^{10} \text{nella confezione } j \text{ non compare il buono } i\text{-esimo}\right)\right) \\ &= \sum_{i=1}^{20} \left(1 - \prod_{j=1}^{10} P(\text{nella confezione } j \text{ non compare il buono } i\text{-esimo})\right) \\ &= \sum_{i=1}^{20} \left(1 - \prod_{j=1}^{10} (1 - P(\text{nella confezione } j \text{ compare il buono } i\text{-esimo}))\right) \\ &= \sum_{i=1}^{20} \left(1 - \prod_{j=1}^{10} \left(1 - \frac{1}{20}\right)\right) \\ &= \sum_{i=1}^{20} \left(1 - \prod_{j=1}^{10} \frac{19}{20}\right) \\ &= \sum_{i=1}^{20} \left(1 - \left(\frac{19}{20}\right)^{10}\right) \\ &\approx \sum_{i=1}^{20} \mathbb{E}(X_i) \\ &\approx 20 \cdot \left(1 - \left(\frac{19}{20}\right)^{10}\right) \approx 8.025\end{aligned}\tag{62}$$

4.13.1 Considerazione

Ipotizzo di voler calcolare $\mathbb{E}((X - c)^2)$, non posso fare $\mathbb{E}(X^2 - 2Xc + c^2)$, tuttavia

$$\begin{aligned}
 \mathbb{E}((X - c)^2) &= \mathbb{E}((X - \mu + \mu - c)^2) \\
 &= \mathbb{E}((a + b)^2) \\
 &= \mathbb{E}(a^2 - 2ab + b^2) \\
 &= \mathbb{E}((X - \mu)^2 - 2(X - \mu)(\mu - c) + (\mu - c)^2) \\
 &= \mathbb{E}((X - \mu)^2) - 2(\mu - c)\mathbb{E}(X - \mu) + (\mu - c)^2 \\
 &\rightarrow \mathbb{E}(X - \mu) = \mathbb{E}(X) - \mu = 0 \\
 &= \mathbb{E}((X - \mu)^2) + (\mu - c)^2 \\
 &= \text{Var}(X) + (\mu - c)^2
 \end{aligned} \tag{63}$$

e dunque

$$\mathbb{E}((X - c)^2) = \text{Var}(X) + (\mu - c)^2 \geq \text{Var}(X) \tag{64}$$

Definendo $(X - c)^2$ come **scarto quadratico**, questo implica che io decido che c descriva in modo deterministico la mia quantità aleatoria, togliendo quindi l'aleatorietà. Se scelgo $c = \mu$ e quindi minimizzo lo scarto quadratico medio, allora $\mathbb{E}((X - c)^2) = \text{Var}(X)$, altrimenti c mi serve per rappresentare l'incertezza in relazione al valore dello scarto quadratico medio.

Covarianza

4.14

Sappiamo che $\mathbb{E}(X + X) = \mathbb{E}(X) + \mathbb{E}(X) = 2\mathbb{E}(X)$ e quindi $\mathbb{E}(2X) = 2\mathbb{E}(X)$, mentre la varianza $\text{Var}(X + X) = \text{Var}(2X) = 2^2\text{Var}(X) = 4\text{Var}(X)$. Ma se ho due variabili aleatorie X, Y definisco la covarianza tra X e Y come

$$\text{Cov}(X, Y) = \mathbb{E}((X - \mu_X)(Y - \mu_Y)) \tag{65}$$

e dichiaro le seguenti proprietà

(1) simmetria

(2)

$$\begin{aligned}
 \text{Cov}(X, Y) &= \mathbb{E}(XY - \mu_X Y - \mu_Y X + \mu_X \mu_Y) \\
 &= \mathbb{E}(XY) - \mu_X \mathbb{E}(Y) - \mu_Y \mathbb{E}(X) + \mu_X \mu_Y \\
 &= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)
 \end{aligned} \tag{66}$$

(3)

$$\begin{aligned}
 \text{Cov}(aX, Y) &= \mathbb{E}(aXY) - \mathbb{E}(aX)\mathbb{E}(Y) \\
 &= a\mathbb{E}(XY) - a\mathbb{E}(X)\mathbb{E}(Y) \\
 &= a(\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)) \\
 &= a\text{Cov}(X, Y)
 \end{aligned} \tag{67}$$

(4)

$$\begin{aligned}
 \text{Cov}(X + Y, Z) &= \mathbb{E}((X + Y)Z) - \mathbb{E}(X + Y)\mathbb{E}(Z) \\
 &= \mathbb{E}(XZ + YZ) - (\mathbb{E}(X) + \mathbb{E}(Y))\mathbb{E}(Z) \\
 &= \mathbb{E}(XZ) + \mathbb{E}(YZ) - \mathbb{E}(X)\mathbb{E}(Z) - \mathbb{E}(Y)\mathbb{E}(Z) \\
 &= \text{Cov}(X, Z) + \text{Cov}(Y, Z)
 \end{aligned} \tag{68}$$

(5)

$$\begin{aligned} \text{Cov}\left(\sum_i X_i, \sum_j Y_j\right) &= \sum_i \text{Cov}\left(X_i, \sum_j Y_j\right) \\ &= \sum_i \text{Cov}\left(\sum_j Y_j, X_i\right) \\ &= \sum_i \sum_j \text{Cov}(Y_j, X_i) \\ &= \sum_i \sum_j \text{Cov}(X_i, Y_j) \end{aligned} \tag{69}$$

(6)

$$\begin{aligned} \text{Cov}(X, X) &= \mathbb{E}((X - \mu_X)(X - \mu_X)) \\ &= \mathbb{E}((X - \mu_X)^2) \\ &= \text{Var}(X) \end{aligned} \tag{70}$$

o anche

$$\begin{aligned} \text{Cov}(X, X) &= \mathbb{E}(X \cdot X) - \mathbb{E}(X)\mathbb{E}(X) \\ &= \mathbb{E}(2X) - \mathbb{E}(X)^2 \\ &= \text{Var}(X) \end{aligned} \tag{71}$$

4.14.1 Considerazione 1

$$\begin{aligned} \text{Var}(X + Y) &= \mathbb{E}((X + Y)^2) - \mathbb{E}(X + Y)^2 \\ &= \mathbb{E}(X^2 + 2XY + Y^2) - (\mathbb{E}(X) + \mathbb{E}(Y))^2 \\ &= \mathbb{E}(X^2) + \mathbb{E}(2XY) + \mathbb{E}(Y^2) - \mathbb{E}(X)^2 - 2\mathbb{E}(X)\mathbb{E}(Y) - \mathbb{E}(Y)^2 \\ &= \mathbb{E}(X^2) - \mathbb{E}(X)^2 + \mathbb{E}(Y^2) - \mathbb{E}(Y)^2 + \mathbb{E}(2XY) - 2\mathbb{E}(X)\mathbb{E}(Y) \\ &= \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y) \end{aligned} \tag{72}$$

e dunque

$$\begin{aligned} \text{Var}(X + X) &= \text{Var}(X) + \text{Var}(X) + 2\text{Cov}(X, X) \\ &= \text{Var}(X) + \text{Var}(X) + 2\text{Var}(X) \\ &= 4\text{Var}(X) \end{aligned} \tag{73}$$

e infine

$$\text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}(X_i) + 2 \sum_{i \neq j}^n \text{Cov}(X_i, X_j) \tag{74}$$

4.14.2 Considerazione 2

In caso di differenza, siccome

$$\text{Var}(-Y) = \text{Var}(Y) \tag{75}$$

allora

$$\text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y) - 2\text{Cov}(X, Y) \tag{76}$$

4.14.3 Considerazione 3

Se X, Y indipendenti, allora $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$, infatti

$$\begin{aligned}\mathbb{E}(XY) &= \sum_i \sum_j x_i y_j P(X = x_i, Y = y_j) \\ &= \sum_i \sum_j x_i y_j P(X = x_i) P(Y = y_j) \\ &= \sum_i x_i P(X = x_i) \sum_j y_j P(Y = y_j) \\ &= \mathbb{E}(X)\mathbb{E}(Y)\end{aligned}\tag{77}$$

e sempre con variabili indipendenti abbiamo che

$$\begin{aligned}\text{Cov}(X, Y) &= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) \\ &= \mathbb{E}(X)\mathbb{E}(Y) - \mathbb{E}(X)\mathbb{E}(Y) \\ &= 0\end{aligned}\tag{78}$$

e che

$$\text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}(X_i)\tag{79}$$

Esempio 1

La varianza della somma di 10 lanci indipendenti di un dado bilanciato?

$$\begin{aligned}\text{Var}\left(\sum_{i=1}^{10} X_i\right) &= \sum_{i=1}^{10} \text{Var}(X_i) \\ &= \sum_{i=1}^{10} \frac{35}{12} \\ &= 10 \cdot \frac{35}{12} \\ &= \frac{175}{6}\end{aligned}\tag{80}$$

Esempio 2

La varianza del numero di teste che si verificano lanciando 10 volte una moneta bilanciata?

$$\begin{aligned}\text{Var}(I_A) &= P(A) \cdot P(\bar{A}) \\ &= \frac{1}{4}\end{aligned}\tag{81}$$

e quindi

$$\begin{aligned}\text{Var}\left(\sum_{i=1}^{10} I_A\right) &= \sum_{i=1}^{10} \text{Var}(I_A) \\ &= 10 \cdot \frac{1}{4} \\ &= \frac{5}{2}\end{aligned}\tag{82}$$

Ipotizzo $X = I_A$, $Y = I_B$ e $A, B \in \mathcal{A}$. Allora $\mathbb{E}(X) = P(X = 1)$ e $\mathbb{E}(Y) = P(Y = 1)$ per definizione. XY vale 1 se A e B si verificano, 0 altrimenti, quindi $\mathbb{E}(XY) = P(XY = 1)$, e dunque

$$\begin{aligned} \text{Cov}(X, Y) &= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) \\ &= P(XY = 1) - P(X = 1)P(Y = 1) \\ &= P(X = 1, Y = 1) - P(X = 1)P(Y = 1) \end{aligned} \quad (83)$$

ora pongo

$$\begin{aligned} \text{Cov}(X, Y) > 0 &\rightarrow P(X = 1, Y = 1) > P(X = 1)P(Y = 1) \\ &= \frac{P(X = 1, Y = 1)}{P(Y = 1)} > P(X = 1) \end{aligned} \quad (84)$$

$$\implies P(X = 1 \mid Y = 1) > P(X = 1)$$

ossia la probabilità che X valga 1 conoscendo Y , è maggiore della probabilità che X valga 1. Infine mi annoto che $\text{Cov}(2X, 2Y) = 4\text{Cov}(X, Y)$.

4.15.1 Formula

$$\rho_{X,Y} = \frac{\text{Cov}(X, Y)}{\sigma_X \cdot \sigma_Y} \quad (85)$$

Quindi

$$\begin{aligned} \rho_{2X,2Y} &= \frac{\text{Cov}(2X, 2Y)}{\sigma_{2X} \cdot \sigma_{2Y}} \\ &= 4 \frac{\text{Cov}(X, Y)}{4 \cdot \sigma_X \cdot \sigma_Y} \end{aligned} \quad (86)$$

e da cui si ottiene

$$\text{Var}(2X) = 4\text{Var}(X) \quad (87)$$

e

$$\begin{aligned} \sigma_{2X} &= \sqrt{\text{Var}(2X)} \\ &= 2\sqrt{\text{Var}(X)} \\ &= 2\sigma_X \end{aligned} \quad (88)$$

Tale coefficiente è invariante rispetto alla scalatura, infatti $\rho_{x,y} = \rho_{2x,2y}$.

Funzione di densità di probabilità

Data $f_X : \mathbb{R} \rightarrow \mathbb{R}$, la funzione di densità di probabilità è

$$\forall B \subseteq \mathbb{R}, P(X \in B) = \int_B f_X(x) dx \quad (89)$$

Se volessi calcolare $P(a \leq X \leq b)$?

$$\int_a^b f_X(x) dx \quad (90)$$

I valori a e b sono finiti, come anche $f_X(a)$ e $f_X(b)$, dunque

$$P(a \leq X \leq b) = P(a < X < b) \quad (91)$$

Inoltre

$$\begin{aligned} P(X = a) &= P(a \leq X \leq a) \\ &= \int_a^a f_X(x) dx \\ &= 0 \end{aligned} \quad (92)$$

e

$$P\left(a - \frac{\varepsilon}{2} \leq X \leq a + \frac{\varepsilon}{2}\right) = \int_{a-\frac{\varepsilon}{2}}^{a+\frac{\varepsilon}{2}} f_X(x) dx \approx \varepsilon f(a) \quad (93)$$

Infine posso anche dire che

$$\begin{aligned} \int_{-\infty}^{\infty} f_X(x) dx &= \int_{\mathbb{R}} f_X(x) dx \\ &= P(X \in \mathbb{R}) \\ &= 1 \end{aligned} \quad (94)$$

la funzione di ripartizione viene vista come

$$\int_{-\infty}^a f_X(x) dx = P(X \leq a) = F_X(a) \quad (95)$$

quindi ci sta dicendo che la funzione di ripartizione la ottengo integrando la funzione di densità di probabilità: per il teorema fondamentale del calcolo integrale sappiamo che

$$\frac{d}{dx} F_X(x) = f(x) \quad (96)$$

Se una variabile aleatoria è continua, il suo valore atteso sarà

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f_X(x) dx \quad (97)$$

Se la variabile aleatoria continua X assume solo specificazioni non negative, allora

$$\mathbb{E}(X) = \int_0^{\infty} 1 - F_x(x) dx \quad (98)$$

Riassumendo, una funzione è funzione di densità quando entrambe le seguenti proprietà sono vere:

- $\int_{-\infty}^{\infty} f_X(x) dx = 1$
- $f_X(x)$ è sempre non negativa

Esempio

Considero una v.a. X continua

$$f_X(x) = \begin{cases} c \cdot (4x - 2x^2) & \text{se } 0 < x < 2 \\ 0 & \text{altrimenti} \end{cases} \quad (99)$$

dunque $f_X(x) = c \cdot (4x - 2x^2)I_{(0,2)}(x)$.

- Quanto vale c ?

$$\begin{aligned}
 \int_{-\infty}^{\infty} f_X(x) dx &= 1 \\
 \int_0^2 c \cdot (4x - 2x^2) dx &= 1 \\
 c \cdot \int_0^2 (4x - 2x^2) dx &= 1 \\
 c \cdot \left(4 \cdot \frac{x^2}{2} - 2 \cdot \frac{x^3}{3} \right)_0^2 &= 1 \\
 c \cdot 8 - 8 \cdot \frac{2}{3} &= 1 \\
 c &= \frac{3}{8}
 \end{aligned} \tag{100}$$

- Quanto vale $P(X > 1)$?

$$\begin{aligned}
 P(X > 1) &= \int_1^{\infty} f_X(x) dx \\
 &= \int_1^2 \frac{3}{8} \cdot (4x - 2x^2) dx \\
 &= \frac{3}{8} \cdot \left(4 \cdot \frac{x^2}{2} - 2 \cdot \frac{x^3}{3} \right)_1^2 \\
 &= \frac{1}{2}
 \end{aligned} \tag{101}$$

Sommario delle proprietà essenziali

4.17

- $\forall X, Y$ indipendenti $\iff f_{x,y}(x, y) = f_x(x) \cdot f_y(y)$ (regola di fattorizzazione)
- $\forall X, Y$ indipendenti $\rightarrow Cov(X, Y) = 0$
- $\forall X, Y$ indipendenti $\rightarrow \mathbb{E}(\prod_1^n X_i) = \prod_1^n \mathbb{E}(X_i)$
- $\mathbb{E}(X_1 + \dots + X_n) = \sum_{i=1}^n \mathbb{E}(X_i)$
- $Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$
- $Var(\sum_{i=1}^n X_i) = \sum_{i=1}^n Var(X_i) + \sum_{i \neq j} Cov(X_i, X_j)$
- $Var(\sum_{i=1}^n X_i) = \sum_{i=1}^n Var(X_i)$, se X, Y indipendenti

Disuguaglianza di Markov

4.18

Il teorema vuole che data una variabile aleatoria $X \geq 0$, $\forall a > 0$ allora

$$P(X \geq a) \leq \frac{\mathbb{E}(X)}{a} \tag{102}$$

Lo si dimostra facilmente

$$\begin{aligned}
 \mathbb{E}(X) &= \int_0^{\infty} x f_X(x) dx \\
 &= \int_0^a x f_X(x) dx + \int_a^{\infty} x f_X(x) dx \geq \int_a^{\infty} x f_X(x) dx
 \end{aligned} \tag{103}$$

e dunque, siccome x non è sotto $f_X(x)$

$$\begin{aligned} \int_a^\infty x f_X(x) dx &\geq \int_a^\infty a \cdot f_X(x) dx \\ &= a \cdot \int_a^\infty f_X(x) dx \\ &= a \cdot P(X \geq a) \end{aligned} \quad (104)$$

per concludere

$$\mathbb{E}(X) \geq a \cdot P(X \geq a) = \frac{\mathbb{E}(X)}{a} \geq P(X \geq a) \quad (105)$$

Esempio

Ipotizziamo di avere una variabile aleatoria X che rappresenta il numero di pezzi prodotti in una settimana. Inoltre sappiamo che $\mathbb{E}(X) = 50$: trova $P(X \geq 75)$.

Semplicemente

$$P(X \geq 75) \leq \frac{\mathbb{E}(X)}{75} = \frac{50}{75} = \frac{2}{3} \quad (106)$$

quindi $P(X \geq 75) \leq \frac{2}{3}$, tuttavia in questo caso mi serve a poco, dunque dobbiamo dedurre un'altra disuguaglianza.

Disuguaglianza di Chebyshev

4.19

Siano date una variabile aleatoria, il suo valore atteso e la sua varianza, allora $\forall > 0$

$$P(|X - \mu| \geq r) \leq \frac{\sigma^2}{r^2} \quad (107)$$

Lo si dimostra tramite i seguenti passaggi

$$|X - \mu| \geq r \iff (X - \mu)^2 \geq r^2 \quad (108)$$

quindi

$$\begin{aligned} P(|X - \mu| \geq r) &= P((X - \mu)^2 \geq r^2) \\ &= P(Y \geq r^2) \end{aligned} \quad (109)$$

e per Markov

$$\begin{aligned} P(Y \geq r^2) &\leq \frac{\mathbb{E}(Y)}{r^2} \\ P((X - \mu)^2 \geq r^2) &\leq \frac{\mathbb{E}((X - \mu)^2)}{r^2} \\ P(|X - \mu| \geq r) &\leq \frac{Var(X)}{r^2} \\ P(|X - \mu| \geq r) &\leq \frac{\sigma^2}{r^2} \end{aligned} \quad (110)$$

Esempio 1

Riprendendo l'esempio usato per Markov in cui si ipotizza di avere una variabile aleatoria X che rappresenta il numero di pezzi prodotti in una settimana, con un $\mathbb{E}(X) = 50$, veniamo ora a conoscenza della sua varianza ossia $Var(X) = 25$. A questo punto possiamo calcolare per esempio

$$\begin{aligned} P(40 < X < 60) &= P(40 - \mu < X < 60 - \mu) \\ &= P(-10 < X < 10) \\ &= P(|X - \mu| < 10) \end{aligned} \quad (111)$$

Non posso applicare Chebyshev perché non ho \geq , quindi

$$P(|X - \mu| < 10) = 1 - P(|X - \mu| \geq 10) \quad (112)$$

Risolvendo solo $P(|X - \mu| \geq 10)$ si ottiene

$$P(|X - \mu| \geq 10) \leq \frac{25}{100} = \frac{1}{4} \quad (113)$$

E dunque ritornando a $1 - P(|X - \mu| \geq 10)$ si ha

$$1 - P(|X - \mu| \geq 10) \geq 1 - \frac{1}{4} = \frac{3}{4} \quad (114)$$

Quindi

$$\begin{aligned} 1 - P(|X - \mu| \geq 10) &\geq \frac{3}{4} \\ P(|X - \mu| < 10) &\geq \frac{3}{4} \end{aligned} \quad (115)$$

Esempio 2

(Preso da Wikipedia)

Suppongo che la media di parole in ogni articolo sia 1000 con una σ di 200, considerando 2σ possiamo dire che la probabilità che il numero di parole in un articolo valga $600 < X < 1400$ è $\geq \frac{1}{k^2}$, ossia almeno $\frac{1}{4}$, quindi almeno il 75% dato che consideriamo

$$\begin{aligned} P(-r < X < r) &= 1 - P(|X - \mu| \geq r) \\ \rightarrow 1 - P(|X - \mu| \geq r) &\geq 1 - \frac{\sigma^2}{r^2} \\ \rightarrow P(|X - \mu| < r) &\geq 1 - \frac{\sigma^2}{r^2} \end{aligned} \quad (116)$$

In altre parole, la probabilità di stare entro 2σ è almeno il 75%.

Considerazione

Se $P(|X - \mu| \geq k\sigma)$ allora

$$\begin{aligned} P(|X - \mu| \geq k\sigma) &\leq \frac{\sigma^2}{\sigma^2 k^2} \\ P(|X - \mu| \geq k\sigma) &\leq \frac{1}{k^2} \end{aligned} \quad (117)$$

Dunque la deviazione standard è un indice di dispersione, ma da qui capiamo proprio che essa è una sorta di unità di misura tanto che *"qual è la probabilità che un valore si collochi entro k deviazioni standard?"*.

Esercizi

4.20

- Si lanciano 3 monete, calcolare la probabilità che escano 3 teste

$$\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8} = 0.125$$

In alternativa potevo impostarlo come una distribuzione binomiale $X \sim B(3, 0.5)$ e risolvere $P(X = 3) = 0.125$.

- **Si lanciano 4 monete, calcolare la probabilità che escano 2 teste e 2 croci**

Considero i seguenti eventi favorevoli:

$$\{TTTT, TTCC, CCTT, CTCT, TCCT, TCTC\}$$

Sono 6, mentre gli eventi possibili sono $2^4 = 16$, dunque:

$$\frac{6}{16} = \frac{3}{8} = 0.375$$

In alternativa potevo impostarlo come una distribuzione binomiale $X \sim B(4, 0.5)$ e risolvere $P(X = 2) = 0.375$.

- **Si lanciano due dadi, calcolare la probabilità che la somma dei risultati sia 4**

I casi favorevoli sono 3 $\{(1, 3), (2, 2), (3, 1)\}$ mentre i casi possibili sono 36, dunque:

$$\frac{3}{36} = \frac{1}{12}$$

- **L'urna 1 contiene 4 biglie rosse e 3 biglie blu, e l'urna 2 contiene 2 biglie rosse e 2 blu. Una biglia viene scelta a caso dall'urna 1 e inserita nell'urna 2. Poi viene estratta una biglia dall'urna 2. Qual è la probabilità che la biglia estratta dall'urna 2 sia rossa? Qual è la probabilità che la biglia estratta dall'urna 1 sia rossa se la biglia estratta dall'urna 2 è blu?**

Si noti che dopo la prima estrazione la composizione dell'urna 1 cambia: se si estrae dall'urna 1 una biglia rossa, l'urna 2 conterrà 5 palline, 3 rosse e 2 blu. Se invece la biglia estratta dall'urna 1 è blu, allora l'urna 2 conterrà 5 palline, 2 rosse e 3 blu. Possiamo quindi calcolare la probabilità richiesta come segue:

$$P(R_2 \cap (R_1 \cup B_1)) = P(R_2 \cap R_1) + P(R_2 \cap B_1)$$

Noi non conosciamo l'intersezione, tuttavia:

$$P(R_2 \cap R_1) + P(R_2 \cap B_1) = P(R_2|R_1)P(R_1) + P(R_2|B_1)P(B_1) = \frac{3}{5} \cdot \frac{4}{7} + \frac{2}{5} \cdot \frac{3}{7} = 0.5143$$

Qual è quindi la probabilità che la biglia estratta dall'urna 1 sia rossa se la biglia estratta dall'urna 2 è blu?

$$P(R_1|B_2) = \frac{P(B_2|R_1)P(R_1)}{P(B_2|R_1)P(R_1) + P(B_2|B_1)P(B_1)} = 0.4706$$

- **Calcola la probabilità che estraendo una carta da un mazzo di 40 questa non sia un asso nero**

La probabilità che sia un asso nero è $\frac{2}{40}$, quindi la probabilità che non sia un asso nero è:

$$1 - \frac{2}{40} = \frac{38}{40} = \frac{19}{20}$$

- **Calcola la probabilità che estraendo una carta da un mazzo di 40 carte, questa sia un asso sapendo che non è di fiori**

$$P(\text{asso}|\overline{\text{fiori}}) = \frac{P(A \cap \overline{\text{fiori}})}{P(\overline{\text{fiori}})} = \frac{\frac{3}{40}}{\frac{30}{40}} = \frac{1}{10}$$

- **Si estraggono consecutivamente due carte da un mazzo di 40, calcolare la probabilità che la seconda carta sia un asso dato che la prima è un 2**

$$P(\text{asso}|2) = \frac{\frac{4}{39} \cdot \frac{4}{40}}{\frac{4}{40}} = \frac{4}{39}$$

- **Si estraggono 3 carte di fila da un mazzo di 40, calcolare la probabilità che siano tutte carte di fiori**

$$\frac{10}{40} \cdot \frac{9}{39} \cdot \frac{8}{38} = 0.0121$$

- Un'urna contiene 30 palline di cui 12 rosse, 10 bianche e 8 nere. Se ne estraggono 2, calcola la probabilità che siano dello stesso colore nei due seguenti casi:

A) La prima pallina viene reinserita nell'urna prima di procedere alla seconda estrazione

B) La prima pallina non viene reinserita nell'urna prima di procedere alla seconda estrazione

Partendo dal punto B, la probabilità che una pallina sia bianca è $\frac{10}{30}$, che sia rossa è $\frac{12}{30}$ e che sia nera è $\frac{8}{30}$. La probabilità di inserire due bianche senza reinserimento qual è?

$$\frac{10}{30} \cdot \frac{9}{29}$$

Per le rosse?

$$\frac{12}{30} \cdot \frac{11}{29}$$

E per le nere?

$$\frac{8}{30} \cdot \frac{7}{29}$$

Ne possiamo pescare una qualsiasi, non ci dice quale risultato scegliere, quindi è o bianca o nera o rossa. Il termine "o" si traduce con l'unione, dunque bisogna effettuare una somma, quindi:

$$\left(\frac{10}{30} \cdot \frac{9}{29}\right) + \left(\frac{12}{30} \cdot \frac{11}{29}\right) + \left(\frac{8}{30} \cdot \frac{7}{29}\right) \approx 0.32$$

Analogamente per rispondere al punto A si applica la stessa procedura ma con reinserimento, dunque:

$$\left(\frac{10}{30} \cdot \frac{10}{30}\right) + \left(\frac{12}{30} \cdot \frac{12}{30}\right) + \left(\frac{8}{30} \cdot \frac{8}{30}\right) \approx 0.34$$

Generalizzando, possiamo dire che la formula generale risulta essere:

$$P((B_1 \cap B_2) \cup (R_1 \cap R_2) \cup (N_1 \cap N_2))$$

Dove la probabilità singola muta a seconda della presenza o meno di reinserimento.

Capitolo 5

Modelli Discreti

Modello di Bernoulli

5.1

Descrive un esperimento con due esiti (successo=1 e fallimento=0)

$$X \sim B(p) \quad (1)$$

- **Supporto:** $D_X = \{0, 1\}$
- **Funzione di massa di probabilità:**

$$p_X(x) = P(X = x) = \begin{cases} 1-p & x = 0 \\ p & x = 1 \\ 0 & \text{altrimenti} \end{cases} = p^x(1-p)^{1-x}I_{\{0,1\}}(x) \quad (2)$$

- **Funzione di ripartizione:**

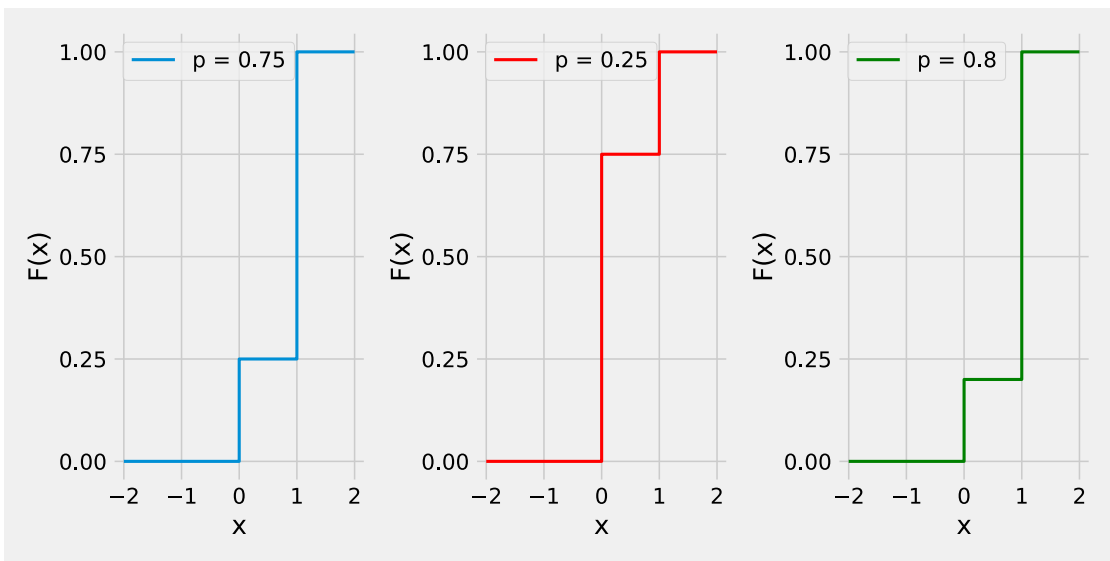
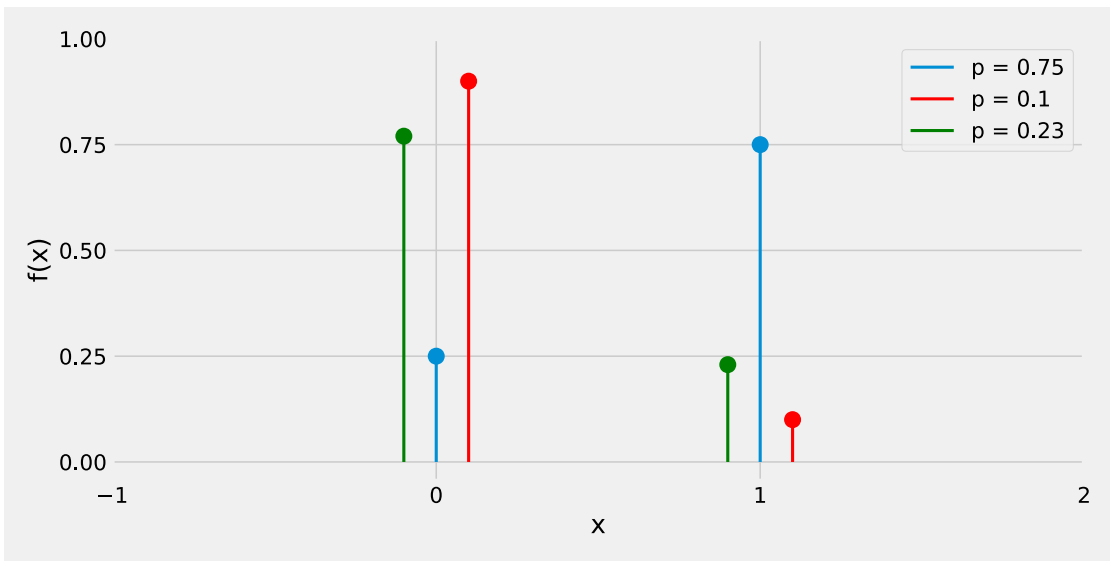
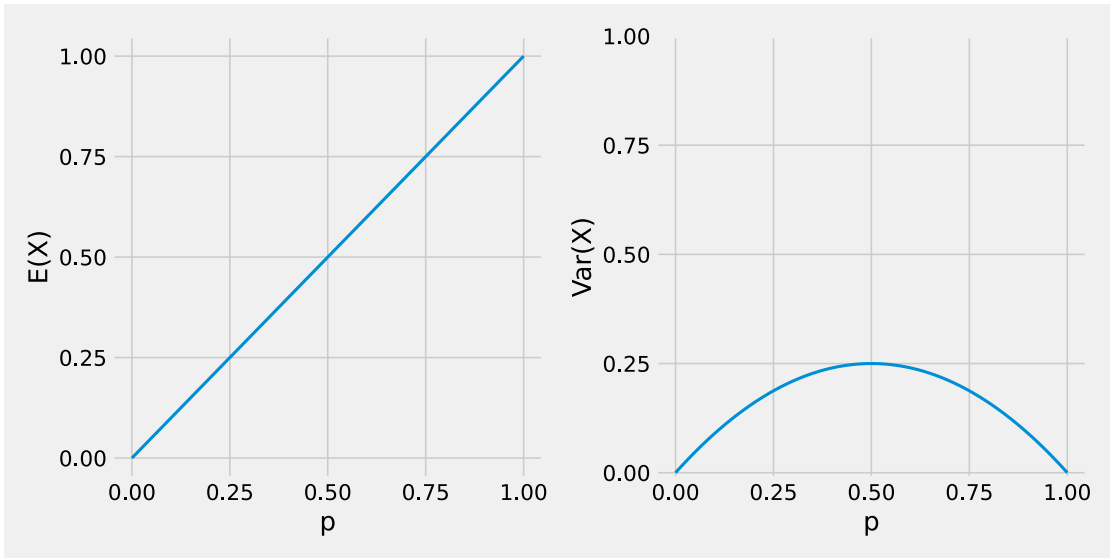
$$F_X(x) = \begin{cases} 0 & x < 0 \\ 1-p & 0 \leq x < 1 \\ 1 & x \geq 1 \end{cases} = (1-p)I_{[0,1)}(x) + I_{[1,\infty)}(x) \quad (3)$$

- **Valore atteso:**

$$\begin{aligned} \mathbb{E}(X) &= \sum_i x_i P(X = x_i) \\ &= 0 \cdot P(X = 0) + 1 \cdot P(X = 1) \\ &= p \end{aligned} \quad (4)$$

- **Varianza:**

$$\begin{aligned} \text{Var}(X) &= \mathbb{E}(X^2) - \mathbb{E}(X)^2 \\ &= \mathbb{E}(X) - \mathbb{E}(X)^2 \\ &= p - p^2 \\ &= p \cdot (1-p) \end{aligned} \quad (5)$$



Se eseguo n ripetizioni indipendenti di un esperimento bernoulliano di parametro p , il conteggio dei successi mi porta ad un modello chiamato *binomiale*.

$$X \sim B(n, p) \quad n \in \mathbb{N} \quad (6)$$

- **Supporto:** $D_X = \{0, 1, \dots, n\}$

- **Funzione di massa di probabilità:**

$$p_X(i) = \binom{n}{i} p^i (1-p)^{n-i} \mathbb{I}_{\{0, \dots, n\}}(i) \quad (7)$$

Dove i è il numero di successi e $n - i$ il numero di insuccessi.

- **Funzione di ripartizione:**

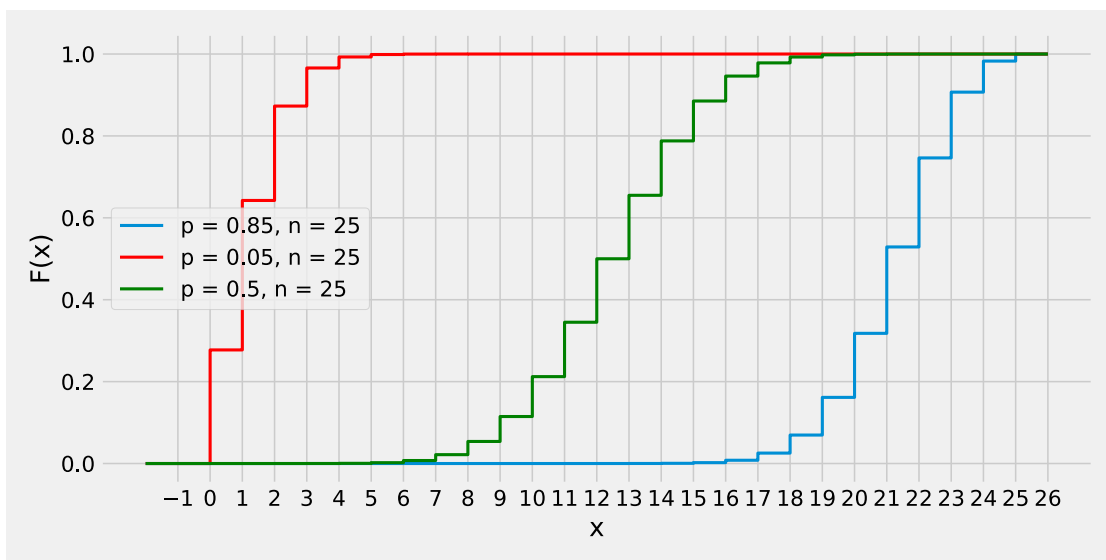
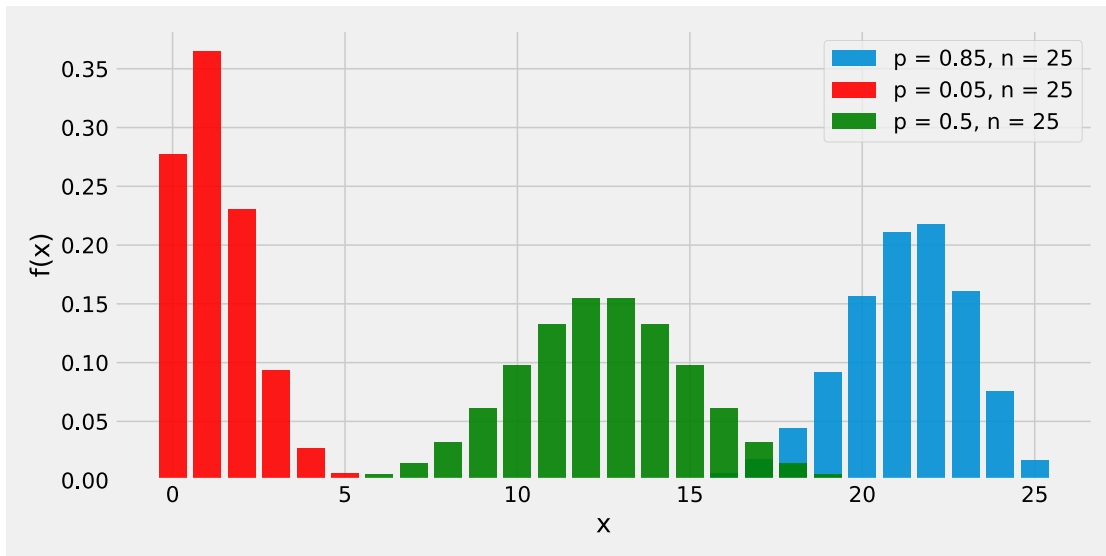
$$\begin{aligned} F_X(x) &= P(X \leq x) \\ &= \sum_{i=0}^{\lfloor x \rfloor} \binom{n}{i} p^i (1-p)^{n-i} I_{[0, n]}(x) + I_{(n, \infty)}(x) \end{aligned} \quad (8)$$

- **Valore atteso:**

$$\begin{aligned} \mathbb{E}(X) &= \mathbb{E} \left(\sum_{i=1}^n X_i \right) \\ &= \sum_{i=1}^n \mathbb{E}(X_i) \\ &= \sum_{i=1}^n p \\ &= np \end{aligned} \quad (9)$$

- **Varianza:**

$$\begin{aligned} Var(X) &= Var \left(\sum_{i=1}^n X_i \right) \\ &\rightarrow \text{indipendenza} \\ &= \sum_{i=1}^n Var(X_i) \\ &= \sum_{i=1}^n p(1-p) \\ &= np \cdot (1-p) \end{aligned} \quad (10)$$



5.2.1 Esempio

La probabilità di avere una penna difettosa è $P(\text{penna difettosa}) = 0.01$ e una confezione contiene 10 penne. Il rimborso avviene se c'è più di una penna difettosa nella confezione: quale percentuale delle scatole verrà restituita?

$$P(\text{rimborso}) = P(\text{più di una penna difettosa su 10}) \quad (11)$$

Identifichiamo con X il numero di penne difettose per confezione, dunque $X \sim B(10, 0.01)$ e quindi

$$P(\text{rimborso}) = P(X > 1) = \sum_{i=2}^{10} \binom{10}{i} 0.01^i (0.99)^{10-i} \quad (12)$$

Ma mi è scomodo lavorare così, quindi posso considerare che:

$$\begin{aligned} P(X > 1) &= 1 - P(X \leq 1) \\ &= 1 - \binom{10}{0} 0.01^0 (0.99)^{10} - \binom{10}{1} 0.01^1 (0.99)^9 \\ &= 1 - 0.99^{10} - 10 \cdot 0.01 \cdot 0.99^9 \\ &\approx 0.0043 \end{aligned} \quad (13)$$

Il che significa che lo 0.43% delle scatole viene rimborsato.

Ma se compro 3 scatole, qual è la probabilità di restituire esattamente 1 scatola?

Definisco Y come il numero di scatole che restituisco su 3 acquistate e dunque $Y \sim B(3, 0.0043)$, quindi

$$P(Y = 1) = \binom{3}{1} 0.0043^1 (0.0043)^2 \approx 0.01278 \quad (14)$$

Quindi 1.278%. Se invece mi viene chiesta la probabilità di restituirne almeno 1 su 3 allora semplicemente

$$P(Y \geq 1) = 1 - P(Y = 0) = 1 - (1 - 0.0043)^3 \approx 0.01284 \quad (15)$$

Ossia 1.284%.

Da qui si evince che in caso di due variabili aleatorie $X_1 \sim B(n_1, p)$ e $X_2 \sim B(n_2, p)$, si ha

$$Y = X_1 + X_2 = \sum_{i=1}^n X_{1,i} + \sum_{j=1}^m X_{2,j} = \sum_{i=1}^{n+m} Y_i \rightarrow Y \sim B(n_1 + n_2, p) \quad (16)$$

Modello Uniforme Discreto

5.3

Ipotizzo $n \in \mathbb{N}$ esiti (da 1 a n) equiprobabili, dunque

$$X \sim U(n) \quad n \in \mathbb{N} \quad (17)$$

- **Supporto:** $D_X = \{1, \dots, n\}$
- **Funzione di massa di probabilità:**

$$p_X(i) = P(X = i) = \frac{1}{n} \cdot \mathbb{I}_{D_X}(i) \quad (18)$$

- **Funzione di ripartizione:**

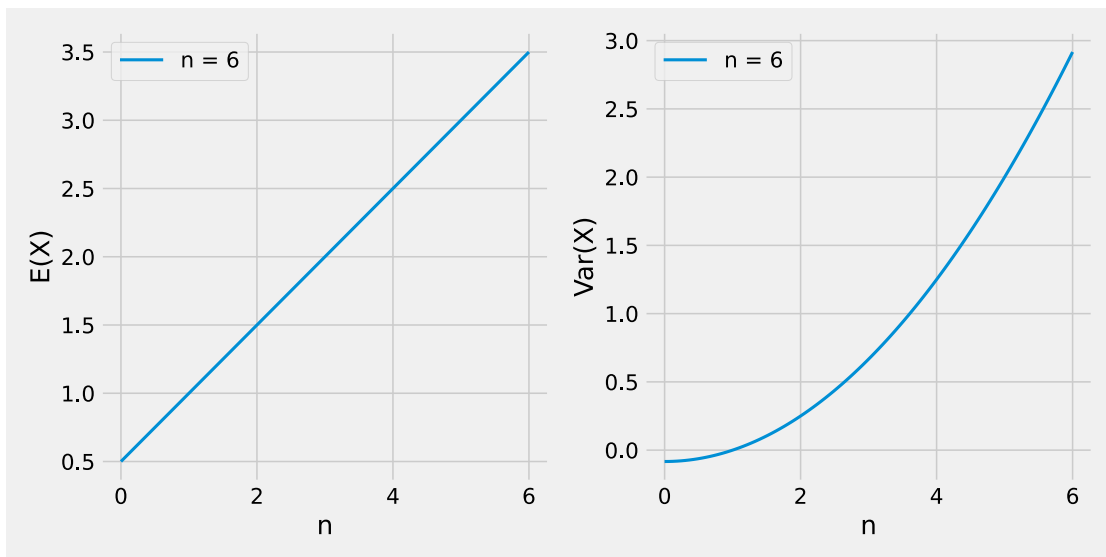
$$\begin{aligned} F_X(x) &= P(X \leq x) \\ &= \sum_{i=1}^{\lfloor x \rfloor} P(X = i) \\ &= \sum_{i=1}^{\lfloor x \rfloor} \frac{1}{n} \\ &= \frac{\lfloor x \rfloor}{n} \cdot \mathbb{I}_{[1, n]}(x) + \mathbb{I}_{[n, \infty)}(x) \end{aligned} \quad (19)$$

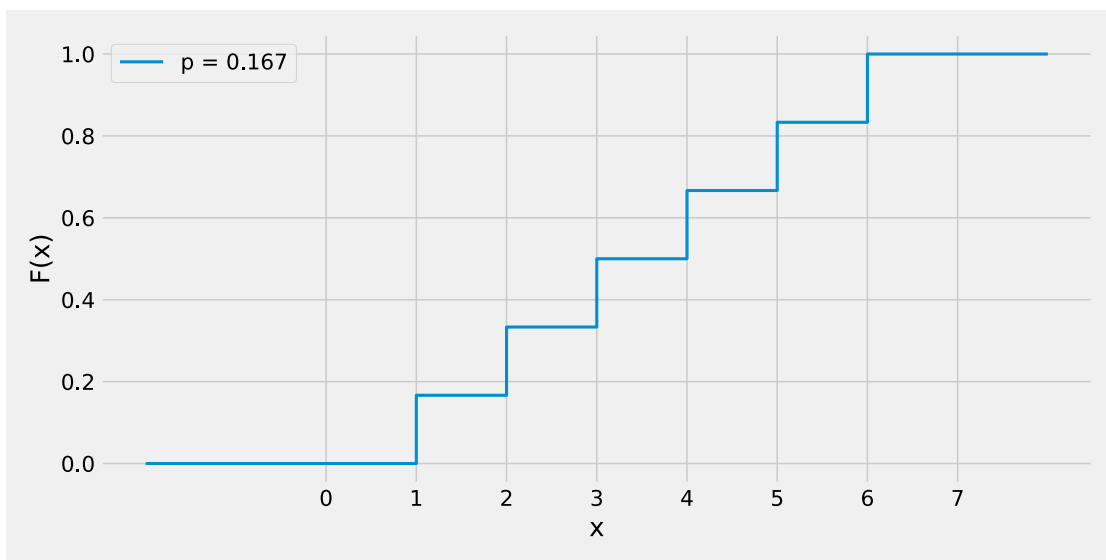
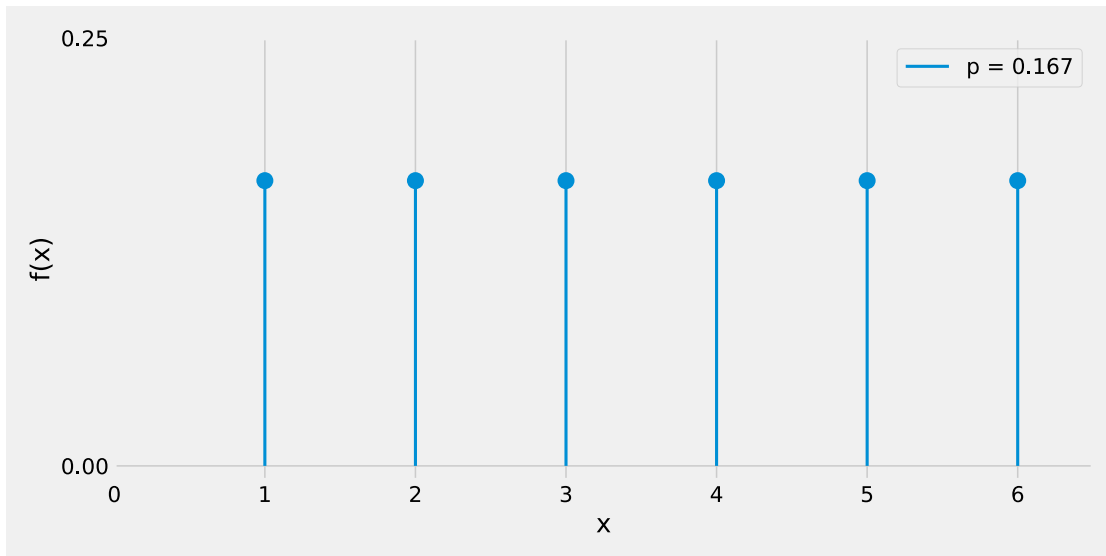
- **Valore atteso:**

$$\begin{aligned} \mathbb{E}(X) &= \sum_{i=1}^n i \cdot P(X = i) \\ &= \frac{1}{n} \sum_{i=1}^n i \\ &= \frac{1}{n} \cdot \frac{n(n+1)}{2} \\ &= \frac{n+1}{2} \end{aligned} \quad (20)$$

• Varianza:

$$\begin{aligned}
 \text{Var}(X) &= \mathbb{E}(X^2) - \mathbb{E}(X)^2 \\
 &= \mathbb{E}(X^2) - \left(\frac{n+1}{2}\right)^2 \\
 &= \sum_{i=1}^n i^2 \cdot P(X=i) - \left(\frac{n+1}{2}\right)^2 \\
 &= \frac{1}{n} \sum_{i=1}^n i^2 - \left(\frac{n+1}{2}\right)^2 \\
 &= \frac{1}{n} \cdot \frac{n(n+1)(2n+1)}{6} - \left(\frac{n+1}{2}\right)^2 \\
 &= \frac{(n+1)(2n+1)}{6} - \left(\frac{n+1}{2}\right)^2 \\
 &= (n+1) \cdot \left(\frac{2n+1}{6 - \frac{n+1}{4}}\right) \\
 &= (n+1) \cdot \frac{4n+2-3n-3}{12} \\
 &= (n+1) \cdot \frac{n-1}{12} \\
 &= \frac{n^2-1}{12}
 \end{aligned} \tag{21}$$





Modello Geometrico

5.4

Basato sul modello di Bernoulli, ripeto l'esperimento finché non ottengo successo: posso contare il numero totale di esperimenti, oppure quello di insuccessi. Nel nostro caso la nostra X sarà il numero di insuccessi prima del primo successo in una sequenza di esperimenti bernoulliani con lo stesso parametro e tra loro indipendenti.

$$X \sim G(p) \quad (22)$$

Dire $X = i$ significa dire che ho effettuato $i + 1$ esperimenti bernoulliani, di cui i primi i hanno avuto insuccesso, e l' $i + 1$ -esimo è stato un successo.

- **Supporto:** $D_X = \mathbb{N} \cup \{0\}$
- **Funzione di massa di probabilità:**

$$\begin{aligned} f_X(i) &= P(X = i) \\ &= p(1 - p)^i \mathbb{I}_{\mathbb{N} \cup \{0\}}(i) \end{aligned} \quad (23)$$

- **Funzione di ripartizione:** Inizio osservando che

$$\begin{aligned}
 P(X > n) &= \sum_{i=n+1}^{\infty} P(X = i) \\
 &= \sum_{i=n+1}^{\infty} p(1-p)^i \\
 &\rightarrow \text{moltiplico e divido per lo stesso valore} \\
 &= \sum_{i=n+1}^{\infty} p(1-p)^i \cdot \frac{(1-p)^{n+1}}{(1-p)^{n+1}} \\
 &= p(1-p)^{n+1} \cdot \sum_{i=n+1}^{\infty} (1-p)^{i-(n+1)} \\
 &\rightarrow j = i - (n+1) \text{ dunque se } i = n+1 \implies j = 0 \\
 &= p(1-p)^{n+1} \cdot \sum_{j=0}^{\infty} (1-p)^j \\
 &\rightarrow \text{per la serie geometrica } \sum_{n=0}^{\infty} x^n = \frac{1}{1-x} \\
 &= p(1-p)^{n+1} \cdot \frac{1}{1-(1-p)} \\
 &= p(1-p)^{n+1} \cdot \frac{1}{p} \\
 &= (1-p)^{n+1}
 \end{aligned} \tag{24}$$

Quindi possiamo dedurre che:

$$\begin{aligned}
 F_X(x) &= P(X \leq x) \\
 &= 1 - P(X > x) \\
 &= 1 - (1-p)^{\lfloor x \rfloor + 1} \mathbb{I}_{\mathbb{N} \cup \{0\}}(x)
 \end{aligned} \tag{25}$$

- **Valore atteso:** Prima calcolo $\sum_{i=0}^{\infty} i\alpha^i$:

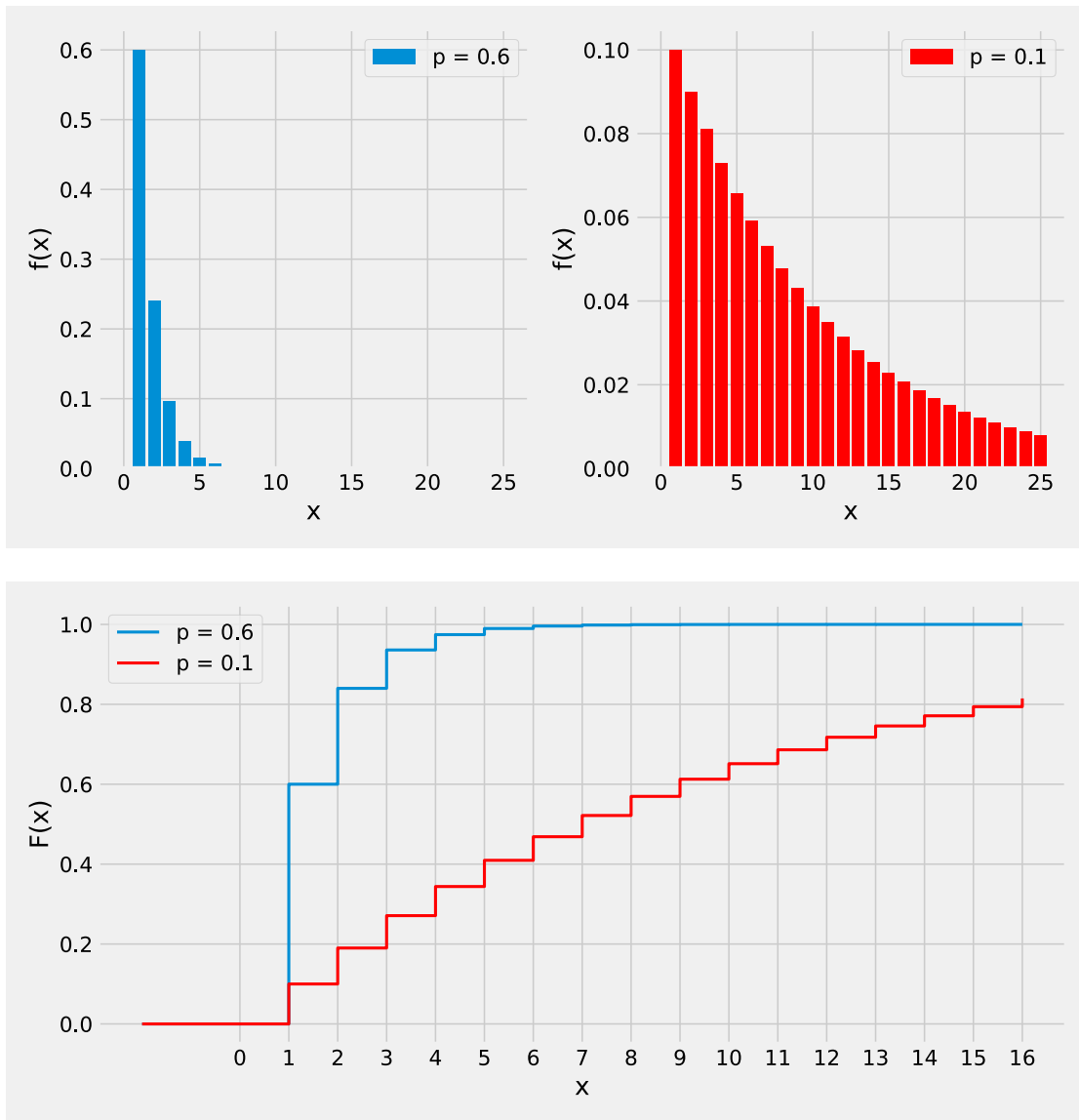
$$\begin{aligned}
 \sum_{i=0}^{\infty} i\alpha^i &= \alpha \sum_{i=0}^{\infty} i\alpha^{i-1} \\
 &\rightarrow \text{la derivata della somma è uguale alla somma delle derivate} \\
 &= \alpha \sum_{i=0}^{\infty} \frac{d}{dx} \alpha^i \\
 &= \alpha \frac{d}{dx} \sum_{i=0}^{\infty} \alpha^i \\
 &\rightarrow \text{per la serie geometrica } \sum_{n=0}^{\infty} x^n = \frac{1}{1-x} \\
 &= \alpha \frac{d}{dx} \frac{1}{1-\alpha} \\
 &= \alpha \cdot \frac{1}{(1-\alpha)^2} \\
 &= \frac{\alpha}{(1-\alpha)^2}
 \end{aligned} \tag{26}$$

Dunque:

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{i=0}^{\infty} i \cdot P(X = i) \\
&= \sum_{i=0}^{\infty} i \cdot p \cdot (1-p)^i \\
&= p \cdot \sum_{i=0}^{\infty} i \cdot (1-p)^i \\
&= p \cdot \frac{1-p}{p^2} \\
&= \frac{1-p}{p}
\end{aligned} \tag{27}$$

• **Varianza:**

$$\begin{aligned}
\text{Var}(X) &= \mathbb{E}(X^2) - \mathbb{E}(X)^2 \\
&= \mathbb{E}(X^2) - \left(\frac{1-p}{p}\right)^2 \\
&= \left(\sum_{i=0}^{\infty} i^2 \cdot p \cdot (1-p)^i\right) - \left(\frac{1-p}{p}\right)^2 \\
&= \left(p(1-p) \cdot \sum_{i=0}^{\infty} i^2 \cdot (1-p)^{i-1}\right) - \left(\frac{1-p}{p}\right)^2 \\
&= \left(p(1-p) \cdot \sum_{i=0}^{\infty} i \cdot i \cdot (1-p)^{i-1}\right) - \left(\frac{1-p}{p}\right)^2 \\
&= \left(p(1-p) \cdot \sum_{i=0}^{\infty} i \cdot \left(-\frac{d}{dp}(1-p)^i\right)\right) - \left(\frac{1-p}{p}\right)^2 \\
&= \left(-p(1-p) \cdot \sum_{i=0}^{\infty} i \cdot \frac{d}{dp}(1-p)^i\right) - \left(\frac{1-p}{p}\right)^2 \\
&= \left(-p(1-p) \cdot \frac{d}{dp} \sum_{i=0}^{\infty} i \cdot (1-p)^i\right) - \left(\frac{1-p}{p}\right)^2 \\
&= \left(-p(1-p) \cdot \frac{d}{dp} \frac{1-p}{p^2}\right) - \left(\frac{1-p}{p}\right)^2 \\
&= \left(-p(1-p) \cdot \frac{-p^2 - 2p(1-p)}{p^4}\right) - \left(\frac{1-p}{p}\right)^2 \\
&= \left(-p(1-p) \cdot \frac{p + 2(1-p)}{p^2}\right) - \left(\frac{1-p}{p}\right)^2 \\
&= \left(\frac{1-p}{p^2} \cdot (p + 2 - 2p)\right) - \left(\frac{1-p}{p}\right)^2 \\
&= \frac{(1-p)(2-p)}{p^2} - \left(\frac{1-p}{p}\right)^2 \\
&= \frac{1-p}{p^2} \cdot (2-p-1+p) \\
&= \frac{1-p}{p^2}
\end{aligned} \tag{28}$$



5.4.1 Considerazione

Un'importante proprietà del modello geometrico è l'**assenza di memoria**, pertanto vale la seguente relazione:

$$\begin{aligned}
 P(X > i + j \mid X > i) &= \frac{P(X > i + j \cap X > i)}{P(X > i)} \\
 &= \frac{P(X > i + j)}{P(X > i)} \\
 &= \frac{(1 - p)^{i+j}}{(1 - p)^i} \\
 &= (1 - p)^j \\
 &= P(X > j) \\
 \\
 &\implies P(X \geq i + j \mid X \geq i) = P(X \geq j)
 \end{aligned}
 \tag{29}$$

Quindi per esempio, se gioco d'azzardo da più tempo di qualcuno, e ho avuto molti insuccessi, NON significa che avrò più successi di qualcuno che ha appena iniziato a giocare.

5.4.2 Esempio

Se X è il numero di volte in cui mi ammalo mensilmente ed esso è $X \sim B(12, 0.01)$, la probabilità che io mi ammali 3 volte entro la fine dell'anno sarà:

$$P(X = 3) = \binom{12}{3} p^3 (1-p)^{12-3} \approx 0.085 \quad (30)$$

Se mi sono ammalato almeno 3 volte allora devo andare dal dottore:

$$\begin{aligned} P(X \geq 3) &= 1 - P(X = 0) - P(X = 1) - P(X = 2) \\ &= 1 - \binom{12}{0} p^0 (1-p)^{12-0} - \binom{12}{1} p^1 (1-p)^{12-1} - \binom{12}{2} p^2 (1-p)^{12-2} \\ &\approx 0.11 \end{aligned} \quad (31)$$

Tuttavia a me interessa andare dal medico per la prima volta tra 5 anni: mentre prima contavo iterativamente le occorrenze di ammalarmi e ho usato un modello binomiale, ora invece quando vado per la prima volta dal dottore, smetto di contare, e dunque uso un modello geometrico $Y \sim G(p')$. Sapendo che la probabilità di ammalarmi almeno 3 volte all'anno è 0.11, qual è la probabilità che la prima volta che vada dal dottore sia tra 5 anni? (Considerando che ho 4 insuccessi e il quinto è un successo)

$$\begin{aligned} P(M) &= P(Y = 4) \\ &= p' (1-p')^4 \\ &= 0.11 \cdot (1 - 0.11)^4 \\ &\approx 0.069 \end{aligned} \quad (32)$$

Ora penso: dato che sono andato dal medico, per l'ultima volta, 2 anni fa, qual è la probabilità che io debba andarci l'anno prossimo? (Almeno due insuccessi li ho visti, ne vedrò un terzo?)

$$\begin{aligned} P(Y \geq 3 \mid Y \geq 2) &= P(Y \geq 1) \\ &= 1 - p' \end{aligned} \quad (33)$$

Modello di Poisson

5.5

$$X \sim P(\lambda) \quad (34)$$

- **Supporto:** $D_X = \mathbb{N} \cup \{0\}$ ($\lambda > 0$)
- **Funzione di massa di probabilità:**

$$\begin{aligned} p_X(i) &= P(X = i) \\ &= e^{-\lambda} \frac{\lambda^i}{i!} \mathbb{I}_{\mathbb{N} \cup \{0\}}(i) \\ &= \sum_{i=0}^{\infty} e^{-\lambda} \frac{\lambda^i}{i!} \\ &= e^{-\lambda} \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} \rightarrow \text{sviluppo in serie di McLaurin di } e^\lambda \end{aligned} \quad (35)$$

• Valore atteso:

$$\begin{aligned}\mathbb{E}(X) &= \sum_{i=0}^{\infty} i \cdot e^{-\lambda} \cdot \frac{\lambda^i}{i!} \\ &\rightarrow \text{se } i \text{ vale } 0 \text{ non succede nulla} \\ &= \sum_{i=1}^{\infty} i \cdot e^{-\lambda} \cdot \frac{\lambda^i}{i!} \\ &\rightarrow \text{porto fuori un } \lambda \\ &= \sum_{i=1}^{\infty} i \cdot e^{-\lambda} \cdot \lambda \frac{\lambda^{(i-1)}}{i!} \\ &\rightarrow \text{divido per } i \\ &= \sum_{i=1}^{\infty} e^{-\lambda} \lambda \frac{\lambda^{(i-1)}}{(i-1)!} \\ &= e^{-\lambda} \lambda \sum_{i=1}^{\infty} \frac{\lambda^{(i-1)}}{(i-1)!} \\ &\rightarrow \text{considero ora } j = i - 1 \\ &= \lambda e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} \\ &= \lambda e^{-\lambda} e^{\lambda} \\ &= \lambda\end{aligned}\tag{36}$$

• **Varianza:**

$$\begin{aligned}
 \text{Var}(X) &= \mathbb{E}(X^2) - \mathbb{E}(X)^2 \\
 &= \mathbb{E}(X^2) - \lambda^2 \\
 &= \left(\sum_{i=0}^{\infty} i^2 \cdot e^{-\lambda} \cdot \frac{\lambda^i}{i!} \right) - \lambda^2 \\
 &= \left(\sum_{i=0}^{\infty} i \cdot e^{-\lambda} \cdot \frac{\lambda^i}{(i-1)!} \right) - \lambda^2 \\
 &= \left(\sum_{i=1}^{\infty} i \cdot e^{-\lambda} \cdot \frac{\lambda^i}{(i-1)!} \right) - \lambda^2 \\
 &= \left(\sum_{i=1}^{\infty} i \cdot e^{-\lambda} \cdot \lambda \frac{\lambda^{(i-1)}}{(i-1)!} \right) - \lambda^2 \\
 &= \left(\lambda \sum_{i=1}^{\infty} i \cdot e^{-\lambda} \cdot \frac{\lambda^{(i-1)}}{(i-1)!} \right) - \lambda^2 \\
 &= \left(\lambda \sum_{i=1}^{\infty} ((i-1) + 1) \cdot e^{-\lambda} \cdot \frac{\lambda^{(i-1)}}{(i-1)!} \right) - \lambda^2 \\
 &= \left(\lambda \sum_{i=1}^{\infty} \left((i-1) \cdot e^{-\lambda} \cdot \frac{\lambda^{(i-1)}}{(i-1)!} + e^{-\lambda} \cdot \frac{\lambda^{(i-1)}}{(i-1)!} \right) \right) - \lambda^2 \\
 &= \left(\lambda \sum_{i=1}^{\infty} (i-1) \cdot e^{-\lambda} \cdot \frac{\lambda^{(i-1)}}{(i-1)!} + \lambda \sum_{i=1}^{\infty} e^{-\lambda} \cdot \frac{\lambda^{(i-1)}}{(i-1)!} \right) - \lambda^2 \\
 &\rightarrow \text{considero ora } j = i - 1 \\
 &= \left(\lambda \sum_{j=0}^{\infty} j \cdot e^{-\lambda} \cdot \frac{\lambda^j}{j!} + \lambda \sum_{j=0}^{\infty} e^{-\lambda} \cdot \frac{\lambda^j}{j!} \right) - \lambda^2 \\
 &= (\lambda \cdot \lambda + \lambda \cdot 1) - \lambda^2 \\
 &= (\lambda^2 + \lambda) - \lambda^2 \\
 &= \lambda^2 + \lambda - \lambda^2 \\
 &= \lambda \\
 \\
 &\implies \mathbb{E}(X) = \text{Var}(X)
 \end{aligned} \tag{37}$$

Il parametro λ rappresenta il numero medio di eventi che accadono solitamente, con X si modella la probabilità che accada un certo numero di eventi nel relativo intervallo temporale, ma a cosa serve? Il modello di Poisson è legato al modello binomiale: immagino un modello binomiale in cui mi chiedo se, al tendere di n ad infinito, non si ottenga un comportamento particolare. La risposta è sì, infatti questo comportamento particolare lo si ha per n grande e p

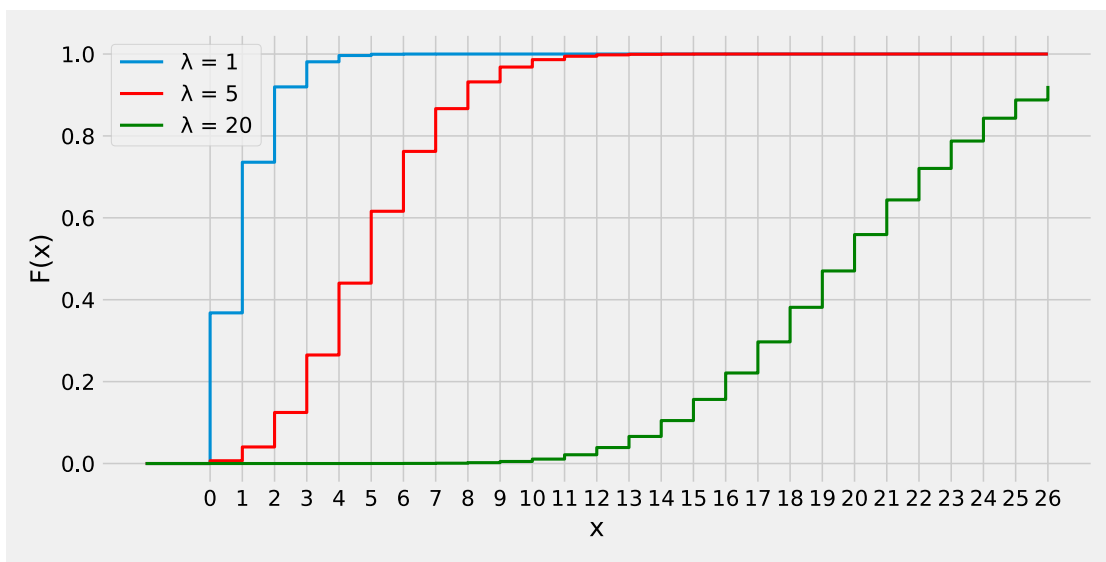
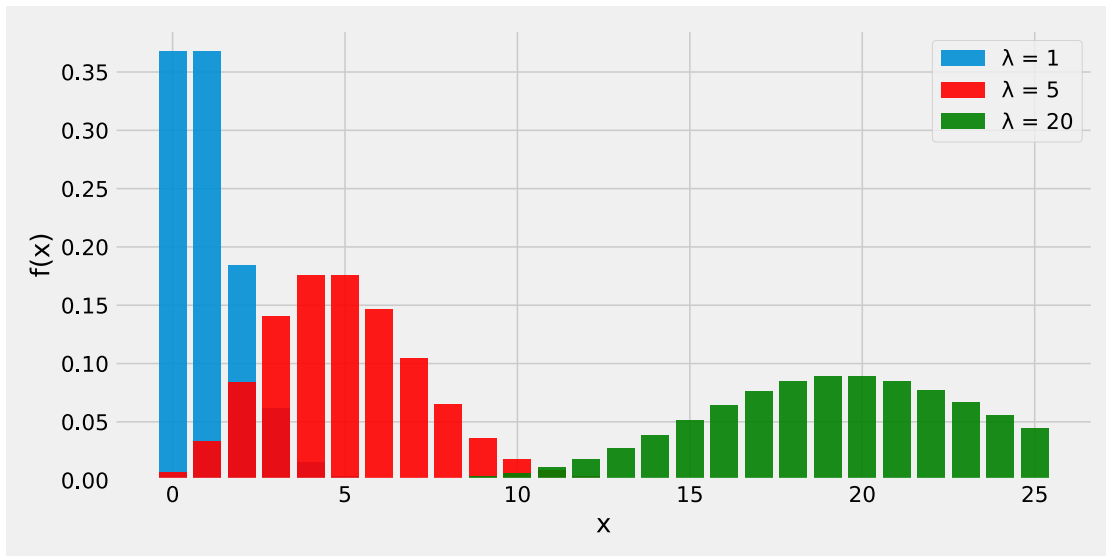
piccolo con $X \sim B(n, p)$. Da qui si può trarre che $np = \lambda$ e $p = \frac{\lambda}{n}$. Nel dettaglio:

$$\begin{aligned}
P(X = i) &= \binom{n}{i} p^i (1-p)^{n-i} \\
&= \binom{n}{i} \left(\frac{\lambda}{n}\right)^i \left(1 - \frac{\lambda}{n}\right)^{n-i} \\
&= \frac{n(n-1) \cdot \dots \cdot (n-i+1)}{i!} \left(\frac{\lambda}{n}\right)^i \left(1 - \frac{\lambda}{n}\right)^{n-i} \\
&= \frac{n(n-1) \cdot \dots \cdot (n-i+1)}{i!} \frac{\lambda^i}{n^i} \left(1 - \frac{\lambda}{n}\right)^{n-i} \\
&= \frac{n(n-1) \cdot \dots \cdot (n-i+1)}{n^i} \frac{\lambda^i}{i!} \left(1 - \frac{\lambda}{n}\right)^{n-i} \\
&= \frac{n}{n} \cdot \frac{n-1}{n} \cdot \dots \cdot \frac{n-i+1}{n} \frac{\lambda^i}{i!} \frac{\left(1 - \frac{\lambda}{n}\right)^n}{\left(1 - \frac{\lambda}{n}\right)^i} \\
&\rightarrow \lim_{n \rightarrow \infty} \frac{n}{n} \cdot \frac{n-1}{n} \cdot \dots \cdot \frac{n-i+1}{n} \frac{\lambda^i}{i!} \frac{\left(1 - \frac{\lambda}{n}\right)^n}{\left(1 - \frac{\lambda}{n}\right)^i} \\
&= 1 \cdot 1 \cdot \dots \cdot 1 \cdot \frac{\lambda^i}{i!} \frac{\left(1 - \frac{\lambda}{n}\right)^n}{\left(1 - \frac{\lambda}{n}\right)^i} \\
&= \frac{\lambda^i}{i!} \frac{\left(1 - \frac{\lambda}{n}\right)^n}{\left(1 - \frac{\lambda}{n}\right)^i} \\
&\rightarrow \text{sappiamo che } \lim_{n \rightarrow \infty} \left(1 + \frac{\alpha}{n}\right)^n = e^\alpha \\
&= \frac{\lambda^i}{i!} \frac{e^{-\lambda}}{(1-0)^i} \\
&= \frac{\lambda^i}{i!} \frac{e^{-\lambda}}{1^i} \\
&= \frac{\lambda^i}{i!} \frac{e^{-\lambda}}{1} \\
&= \frac{\lambda^i}{i!} e^{-\lambda}
\end{aligned} \tag{38}$$

Dunque una poissoniana può essere utilizzata per approssimare una binomiale quando n cresce molto e p decresce.

Indicativamente per $n > 100$ e $p < 0.1$ posso dire che $B(n, p) \sim P(np)$.

N.B. la f_X di una poissoniana può essere unimodale oppure bimodale.



5.5.1 Esempio 1

Una compagnia di assicurazioni riceve in media 5 richieste di rimborso al giorno. In che frazione delle giornate vedrò arrivare meno di 3 richieste?

Considero $X \sim P(5)$ come il numero di richieste al giorno, dunque

$$\begin{aligned}
 P(X < 3) &= P(X = 0) + P(X = 1) + P(X = 2) \\
 &= e^{-5} \left(\frac{5^0}{0!} + \frac{5^1}{1!} + \frac{5^2}{2!} \right) \\
 &\approx 0.1247
 \end{aligned}
 \tag{39}$$

Dunque il 12.5% delle volte ricevo meno di 3 richieste di rimborso al giorno.

Con che probabilità in una settimana lavorativa (5 giorni) arrivano 4 richieste in 3 giorni (4 richieste lunedì, martedì)

e venerdì per esempio)?

$$\begin{aligned}
 P(\text{in 3gg su 5 arrivano 4 richieste}) &\rightarrow P(\text{in 1gg arrivano 4 richieste}) \\
 &= P(X = 4) \\
 &= e^{-5} \frac{5^4}{4!} \\
 &\approx 0.1755 \\
 &\rightarrow Y \sim B(5, p) \\
 &\rightarrow P(Y = 3) \\
 &= \binom{5}{3} p^3 (1-p)^2 \\
 &\approx 0.0367
 \end{aligned} \tag{40}$$

5.5.2 Esempio 2

La probabilità che un pezzo sia difettoso è 0.1, ci sono 10 pezzi, qual è la probabilità che al più uno sia difettoso? Innanzitutto $X \sim B(10, 0.1)$, dunque

$$\begin{aligned}
 P(X \leq 1) &= P(X = 0) + P(X = 1) \\
 &= \binom{10}{0} 0.1^0 \cdot 0.9^{10} + \binom{10}{1} 0.1^1 \cdot 0.9^9 \\
 &\approx 0.7361
 \end{aligned} \tag{41}$$

Però sapendo che $\lambda = n \cdot p = 10 \cdot 0.1 = 1$, possiamo dire che $X' \sim P(\lambda)$ e quindi si poteva svolgere:

$$\begin{aligned}
 P(X' \leq 1) &= P(X' = 0) + P(X' = 1) \\
 &= e^{-1} \left(\frac{1^0}{0!} + \frac{1^1}{1!} \right) \\
 &\approx 0.7358
 \end{aligned} \tag{42}$$

Da qui deduciamo che iniziamo ad avere un errore a partire dalla terza cifra significativa.

Modello Ipergeometrico

5.6

Dati

- **n**: numero di estrazioni senza reimmissione da eseguire
- **N**: numero totale di oggetti "funzionanti"
- **M**: numero totale di oggetti "difettosi"

Definiamo con

$$X \sim H(n, N, M) \tag{43}$$

La distribuzione ipergeometrica, ossia quella che modella un'estrazione di oggetti binari da un'urna senza effettuare reimmissione.

Siccome l'insieme da cui si estrae è formato da variabili aleatorie bernoulliane (1 se l'i-esimo oggetto estratto funziona, 0 altrimenti), posso calcolare

$$\begin{aligned}
 \mathbb{E}(X_i) &= P(X_i = 1) \\
 &= \frac{\text{funzionante}}{\text{guasto} + \text{funzionante}} \\
 &= p
 \end{aligned} \tag{44}$$

Ossia la frequenza relativa delle vittorie. Questo ci torna utile nella definizione del modello.

Come ulteriore considerazione, al crescere del fattore $N + M$ si sta considerando un'urna di dimensioni elevate e dunque l'ipergeometrica è approssimabile con una binomiale.

$$\lim_{(N+M) \rightarrow \infty} H(n, N, M) \sim B\left(n, \frac{N}{N+M}\right) \quad (45)$$

Di seguito le caratteristiche del modello ipergeometrico (N.B. N e M sono intercambiabili, basta che ci sia logica nell'usarli e che le formule siano corrette):

- **Supporto:** $D_X = \mathbb{N} \cup \{0\}$ ($[\max\{0, n - M\}, \min\{n, N\}]$)
- **Funzione di massa di probabilità:**

$$\begin{aligned} f_X(i) &= P(X = i) \\ &= \frac{\binom{N}{i} \binom{M}{n-i}}{\binom{N+M}{n}} \mathbb{I}_{\{0, \dots, n\}}(i) \end{aligned} \quad (46)$$

- **Valore atteso:**

$$\begin{aligned} \mathbb{E}(X) &= \mathbb{E}\left(\sum_{i=1}^n X_i\right) \\ &= \sum_{i=1}^n \mathbb{E}(X_i) \\ &= \sum_{i=1}^n \frac{\text{funzionante}}{\text{guasto} + \text{funzionante}} \\ &= np \end{aligned} \quad (47)$$

\implies come nel modello binomiale

- **Varianza:** Inizio col calcolare

$$\begin{aligned} \text{Var}(X_i) &= \mathbb{E}(X_i)(1 - \mathbb{E}(X_i)) \\ &= \frac{\text{funzionante}}{\text{guasto} + \text{funzionante}} \cdot \frac{\text{guasto}}{\text{guasto} + \text{funzionante}} \\ &= \frac{\text{funzionante} \cdot \text{guasto}}{(\text{funzionante} + \text{guasto})^2} \\ &= \frac{NM}{(N + M)^2} \end{aligned} \quad (48)$$

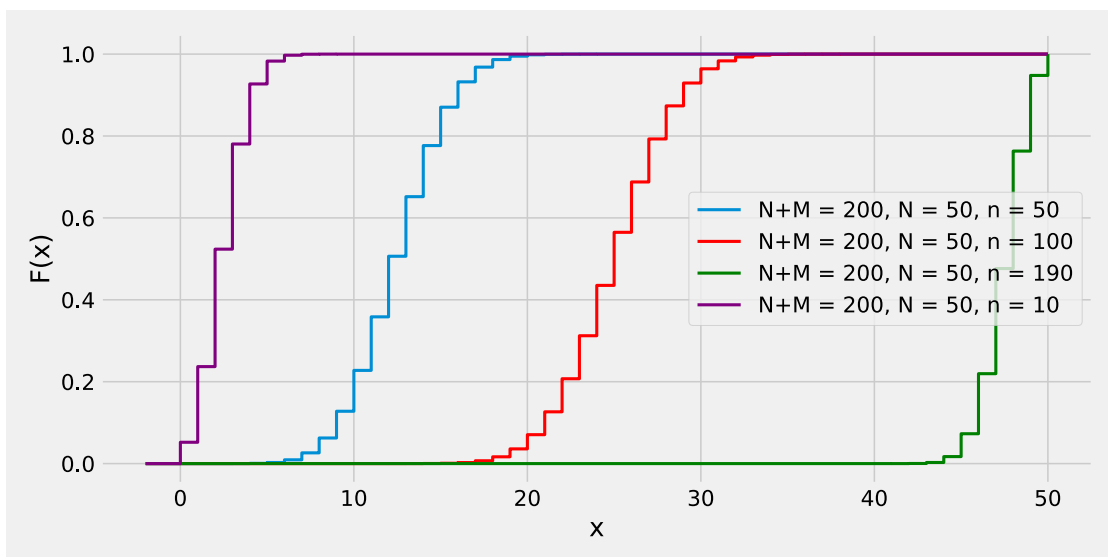
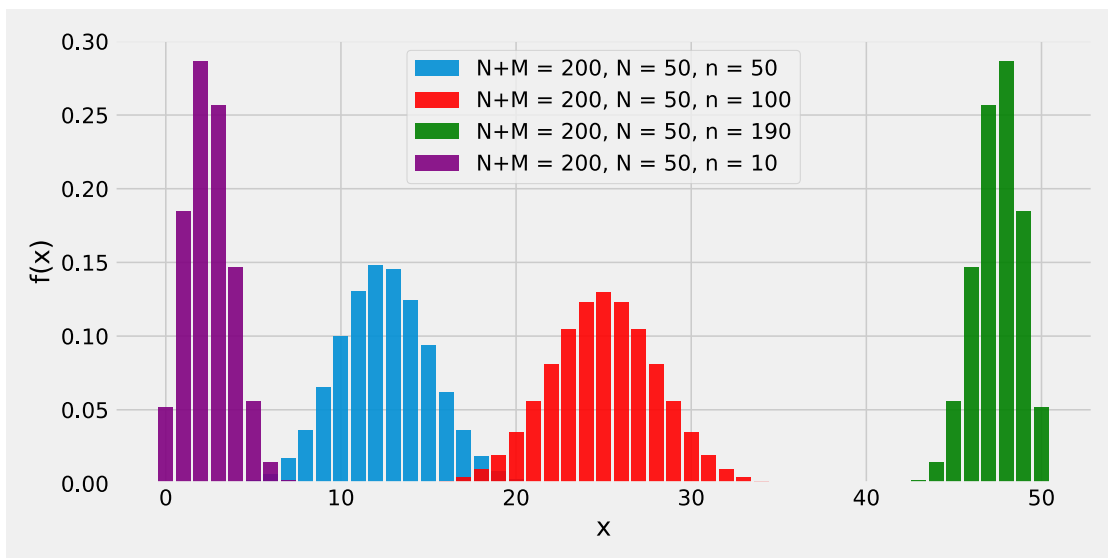
Noto che le X non sono indipendenti perché l'estrazione senza reinserimento influisce sugli esiti successivi.

Dunque proseguo:

$$\begin{aligned}
Var(X) &= Var\left(\sum_{i=1}^n X_i\right) \\
&\rightarrow \text{non sono indipendenti} \\
&= \sum_{i=1}^n Var(X_i) + \sum_{i \neq j} Cov(X_i, X_j) \\
&= \sum_{i=1}^n Var(X_i) + \sum_{i \neq j} (\mathbb{E}(X_i \cdot X_j) - \mathbb{E}(X_i)\mathbb{E}(X_j)) \\
&= \sum_{i=1}^n Var(X_i) + \sum_{i \neq j} \left(P(X_i \cdot X_j = 1) - \left(\frac{N}{N+M}\right)^2 \right) \\
&= \sum_{i=1}^n Var(X_i) + \sum_{i \neq j} \left(P(X_i \cap X_j = 1) - \left(\frac{N}{N+M}\right)^2 \right) \\
&\rightarrow P(E | F) = \frac{P(E \cap F)}{P(F)} \\
&= \sum_{i=1}^n Var(X_i) + \sum_{i \neq j} \left(P(X_i | X_j = 1)P(X_i = 1) - \left(\frac{N}{N+M}\right)^2 \right) \\
&= \sum_{i=1}^n Var(X_i) + \sum_{i \neq j} \left(\frac{N-1}{N-1+M} \cdot \frac{N}{N+M} - \left(\frac{N}{N+M}\right)^2 \right) \\
&\rightarrow \text{inverto il significato di N e M per essere coerente} \\
&\quad \text{con i miei appunti, altrimenti rischio di sbagliare} \\
&= \sum_{i=1}^n Var(X_i) + \sum_{i \neq j} \left(\frac{M-1}{N+M-1} \cdot \frac{M}{N+M} - \left(\frac{M}{N+M}\right)^2 \right) \\
&= \sum_{i=1}^n Var(X_i) + \sum_{i \neq j} \left(\frac{M}{N+M} \cdot \left(\frac{M-1}{N+M-1} - \frac{M}{N+M} \right) \right) \\
&= \sum_{i=1}^n Var(X_i) + \sum_{i \neq j} \left(\frac{M}{N+M} \cdot \frac{(N+M)(M-1) - (N+M-1)M}{(N+M-1)(N+M)} \right) \\
&= \sum_{i=1}^n Var(X_i) + \sum_{i \neq j} \left(\frac{M}{N+M} \cdot \frac{NM - N + M^2 - M - NM - M^2 + M}{(N+M-1)(N+M)} \right) \\
&= \sum_{i=1}^n Var(X_i) + \sum_{i \neq j} \left(\frac{M}{N+M} \cdot \frac{-N}{(N+M-1)(N+M)} \right) \\
&= \sum_{i=1}^n Var(X_i) + \sum_{i \neq j} \left(\frac{-NM}{(N+M-1)(N+M)^2} \right) \\
&= \sum_{i=1}^n \frac{NM}{(N+M)^2} + \sum_{i \neq j} \left(\frac{-NM}{(N+M-1)(N+M)^2} \right) \\
&= n \cdot \frac{NM}{(N+M)^2} - n(n-1) \cdot \frac{NM}{(N+M-1)(N+M)^2}
\end{aligned} \tag{49}$$

Continuando...

$$\begin{aligned}
 &= \frac{nNM}{(N+M)^2} - \frac{n(n-1)NM}{(N+M-1)(N+M)^2} \\
 &= \frac{nNM}{(N+M)^2} \cdot \left(1 - \frac{(n-1)}{N+M-1}\right) \\
 &= \frac{nNM}{(N+M)^2} \cdot \left(1 - \frac{n-1}{N+M-1}\right) \\
 &= n \frac{N}{N+M} \frac{M}{N+M} \cdot \left(1 - \frac{n-1}{N+M-1}\right) \\
 &= n \left(1 - \frac{M}{N+M}\right) \frac{M}{N+M} \cdot \left(1 - \frac{n-1}{N+M-1}\right) \\
 &= n \frac{M}{N+M} \left(1 - \frac{M}{N+M}\right) \left(1 - \frac{n-1}{N+M-1}\right) \\
 &= np(1-p) \left(1 - \frac{n-1}{N+M-1}\right) \\
 &\rightarrow \text{per } N+M \rightarrow \infty \\
 &= np(1-p)
 \end{aligned} \tag{50}$$



5.6.1 Esempio 1

Ipotizzo $N = 5$ oggetti funzionanti e $M = 1$ oggetti difettosi. Posso effettuare $n = 2$ estrazioni senza reimmissione: qual è la probabilità che gli X oggetti funzionanti che pesco siano i , e dunque $P(X = i)$?

Innanzitutto, dati i valori precedenti, può essere $P(X = 0)$? Solo se $P(X = 0) = 0$, perché so che $M = 1$ e dunque se sono sfortunato lo pesco, ma poi non essendoci reimmissione non ci sono altri oggetti non funzionanti da pescare, quindi per forza di cose non posso avere due estrazioni, senza reimmissione, di oggetti non funzionanti quando di oggetti non funzionanti ce n'è solo uno. Nel dettaglio

$$P(X = i) = \frac{\binom{N}{i} \binom{M}{n-i}}{\binom{N+M}{n}} \mathbb{I}_{\{0, \dots, n\}}(i) \quad (51)$$

E da qui sappiamo che $\binom{M}{n-i} = \binom{1}{2}$, ossia stiamo estraendo 2 elementi da un universo che ne contiene 1, è strano, tant'è che $\frac{1!}{2!(-1)!} = 0$ e quindi

$$\frac{\binom{N}{i} \cdot 0}{\binom{N+M}{n}} \mathbb{I}_{\{0, \dots, n\}}(i) = 0 \quad (52)$$

Quindi il dominio in realtà è

$$[\max\{0, n - M\}, \min\{n, N\}] \quad (53)$$

Ossia l'intervallo tra il numero minimo di oggetti funzionanti che io devo necessariamente trovare in quelli che ho estratto (scelto come il massimo valore tra $n - M$ e 0 in caso $n - M$ fosse negativo) e il numero di estrazioni funzionanti (il minimo tra le n estrazioni e le N possibilità funzionanti).

5.6.2 Esempio 2

Ipotizzo $M = 15$ oggetti funzionanti e $N = 5$ oggetti guasti. Immagino ora di estrarre casualmente $n = 6$ oggetti e so che il sistema funziona quando al più 2 componenti sono guasti, quindi quando su $n = 6$ estrazioni almeno 4 funzionano, allora:

$$\begin{aligned} P(X \geq 4) &= P(X = 4 \cup X = 5 \cup X = 6) \\ &= P(X = 4) + P(X = 5) + P(X = 6) \\ &= \frac{\binom{15}{4} \binom{5}{2} + \binom{15}{5} \binom{5}{1} + \binom{15}{6} \binom{5}{0}}{\binom{20}{6}} \\ &\approx 0.8687 \end{aligned} \quad (54)$$

Capitolo 6

Modelli Continui

Siccome ora si lavora su intervalli continui, non ha più senso parlare della funzione di massa, in quanto ogni valore preciso avrebbe probabilità pari a 0. Fissato un punto a e un ε piccolo a piacere, definisco dunque la funzioni di densità nel seguente modo

$$\begin{aligned} f_X(a) &= \int_{a-\frac{\varepsilon}{2}}^{a+\frac{\varepsilon}{2}} f_x(x) dx \\ &\approx \varepsilon f(a) \end{aligned} \quad (1)$$

E posso dire che

$$\int_{-\infty}^{+\infty} f_x(x) dx = 1 \quad (2)$$

Inoltre definisco la funzione di ripartizione come

$$F_X(x) = \int_{-\infty}^x f_x(u) du \quad (3)$$

Da cui si evince che la funzione di ripartizione non è altro che la primitiva della funzione di densità.

Modello Uniforme Continuo

6.1

$$X \sim U([a, b]) \quad (4)$$

- **Supporto:** $D_X = [a, b] \subseteq \mathbb{R}$
- **Funzione di densità:**

$$f_X(x) = \frac{1}{b-a} \mathbb{I}_{[a,b]}(x) \quad (5)$$

- **Funzione di ripartizione:**

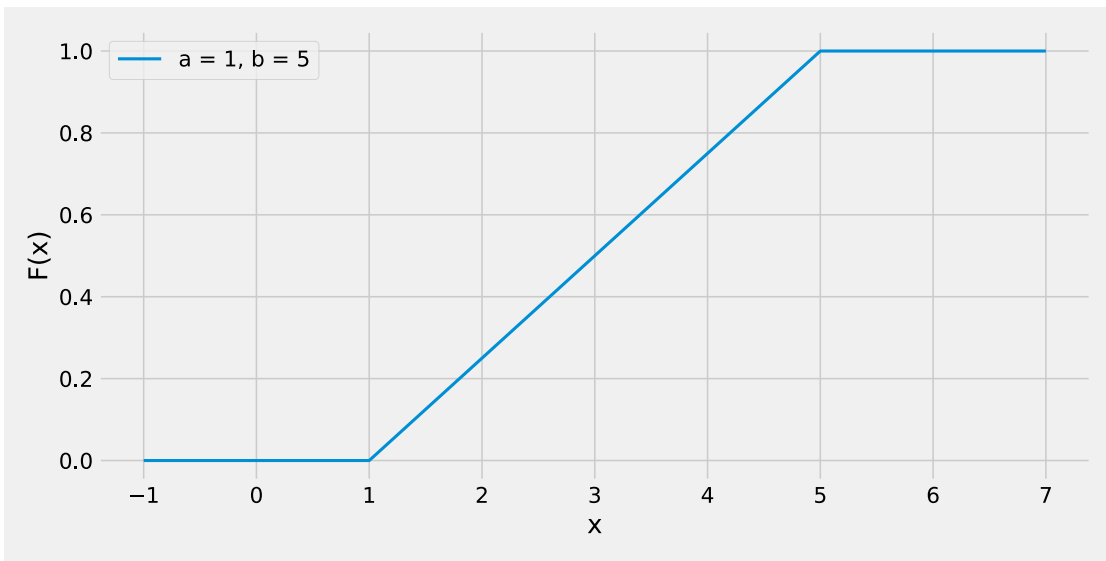
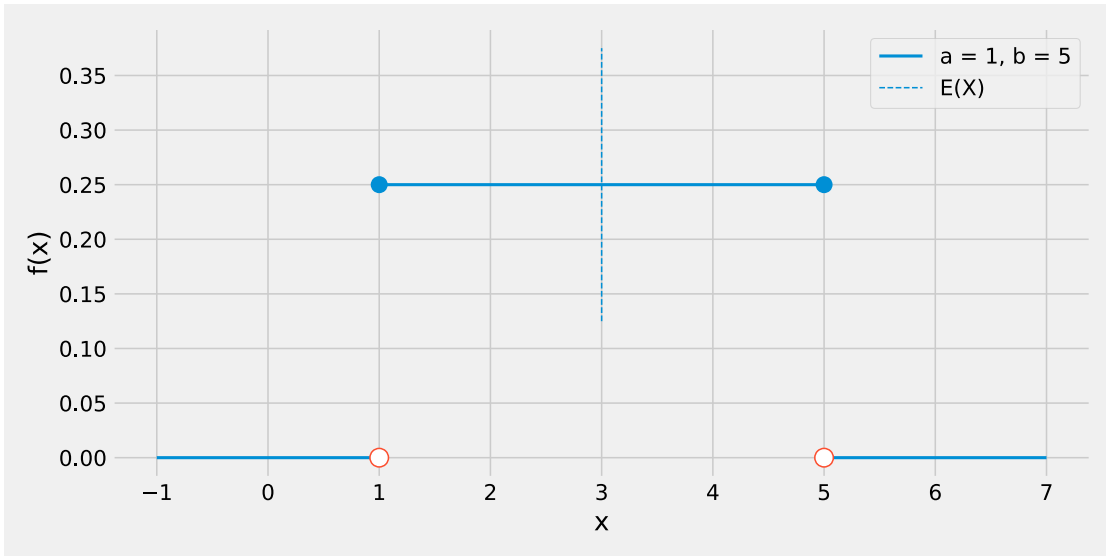
$$\begin{aligned} F_X(x) &= P(X \leq x) \\ &= \int_{-\infty}^x f_X(u) du \\ &= \int_a^x \frac{1}{b-a} du \\ &= \frac{1}{b-a} u \Big|_a^x \\ &= \frac{x-a}{b-a} \mathbb{I}_{[a,b]}(x) + \mathbb{I}_{(b,\infty)}(x) \end{aligned} \quad (6)$$

• Valore atteso:

$$\begin{aligned}
 \mathbb{E}(X) &= \int_{-\infty}^{\infty} x f_X(x) dx \\
 &= \int_a^b \frac{x}{b-a} dx \\
 &= \frac{1}{b-a} \int_a^b x dx \\
 &= \frac{1}{b-a} \frac{x^2}{2} \Big|_a^b \\
 &= \frac{1}{b-a} \frac{b^2 - a^2}{2} \\
 &= \frac{1}{b-a} \frac{(b+a)(b-a)}{2} \\
 &= \frac{b+a}{2}
 \end{aligned} \tag{7}$$

• Varianza:

$$\begin{aligned}
 Var(X) &= \mathbb{E}(X^2) - \mathbb{E}(X)^2 \\
 &= \mathbb{E}(X^2) - \left(\frac{b+a}{2}\right)^2 \\
 &= \mathbb{E}(X^2) - \left(\frac{b+a}{2}\right)^2 \\
 &= \left(\int_a^b x^2 f_X(x) dx\right) - \left(\frac{b+a}{2}\right)^2 \\
 &= \left(\frac{1}{b-a} \int_a^b x^2 dx\right) - \left(\frac{b+a}{2}\right)^2 \\
 &= \left(\frac{1}{b-a} \frac{x^3}{3} \Big|_a^b\right) - \left(\frac{b+a}{2}\right)^2 \\
 &= \left(\frac{b^3 - a^3}{3(b-a)}\right) - \left(\frac{b+a}{2}\right)^2 \\
 &= \left(\frac{(b-a)(a^2 + ab + b^2)}{3(b-a)}\right) - \left(\frac{b+a}{2}\right)^2 \\
 &= \left(\frac{a^2 + ab + b^2}{3}\right) - \left(\frac{b+a}{2}\right)^2 \\
 &= \frac{a^2 + ab + b^2}{3} - \frac{(a+b)^2}{4} \\
 &= \frac{a^2 + ab + b^2}{3} - \frac{a^2 + 2ab + b^2}{4} \\
 &= \frac{4a^2 + 4ab + 4b^2 - 3a^2 - 6ab - 3b^2}{12} \\
 &= \frac{a^2 - 2ab + b^2}{12} \\
 &= \frac{(b-a)^2}{12}
 \end{aligned} \tag{8}$$



6.1.1 Considerazione 1

$$\begin{aligned}
 P(X \in \mathbb{I}) &= \int f_X(x) dx \\
 &= \int_{\mathbb{I}} \frac{1}{b-a} dx \\
 &= \frac{1}{b-a} \int_{\mathbb{I}} dx \\
 &= \frac{1}{b-a} x \Big|_{\mathbb{I}} \\
 &= \frac{|\mathbb{I}|}{b-a}
 \end{aligned}
 \tag{9}$$

6.1.2 Considerazione 2

$$\begin{aligned}\int_{-\infty}^{\infty} f_X(x) dx &= \int_a^b \frac{1}{b-a} dx \\ &= \frac{1}{b-a} x \Big|_a^b \\ &= \frac{b-a}{b-a} \\ &= 1\end{aligned}\tag{10}$$

6.1.3 Esempio

Sappiamo che $X \sim U([0, 30])$ e X è il minuto, dalle 07:00 in poi, nel quale arrivo alla fermata dell'autobus: qual è la probabilità che io attenda meno di 5 minuti per l'autobus, sapendo che l'autobus passa una volta ogni 15 minuti (alle 07:00, 07:15 e 07:30)?

$$\begin{aligned}P(\text{attendo meno di 5 minuti}) &= P(X = 0 \cup 10 < X \leq 15 \cup 25 < X \leq 30) \\ &= P(X = 0) + P(10 < X \leq 15) + P(25 < X \leq 30) \\ &\rightarrow P(X = 0) = 0 \\ &= P(10 < X \leq 15) + P(25 < X \leq 30) \\ &\rightarrow P(\alpha < X \leq \beta) = F_X(\beta) - F_X(\alpha) \\ &= F_X(15) - F_X(10) + F_X(30) - F_X(25) \\ &= \frac{15-0}{30-0} - \frac{10-0}{30-0} + \frac{30-0}{30-0} - \frac{25-0}{30-0} \\ &= \frac{15-10}{30} + \frac{30-25}{30} \\ &= \frac{5}{30} + \frac{5}{30} \\ &= \frac{10}{30} \\ &= \frac{1}{3}\end{aligned}\tag{11}$$

Modello Esponenziale

6.2

La distribuzione esponenziale descrive la "durata di vita" di un fenomeno senza memoria. Per esempio può essere usata per calcolare l'occorrenza di eventi come terremoti o predizioni varie. Essa è rappresentata da:

$$X \sim E(\lambda) \quad \lambda \in \mathbb{R}^+ \setminus \{0\}\tag{12}$$

- **Supporto:** $D_x = \mathbb{R}^+$
- **Funzione di densità:**

$$f_X(x) = \lambda e^{-\lambda x} \mathbb{I}_{[0, +\infty)}(x)\tag{13}$$

• Funzione di ripartizione:

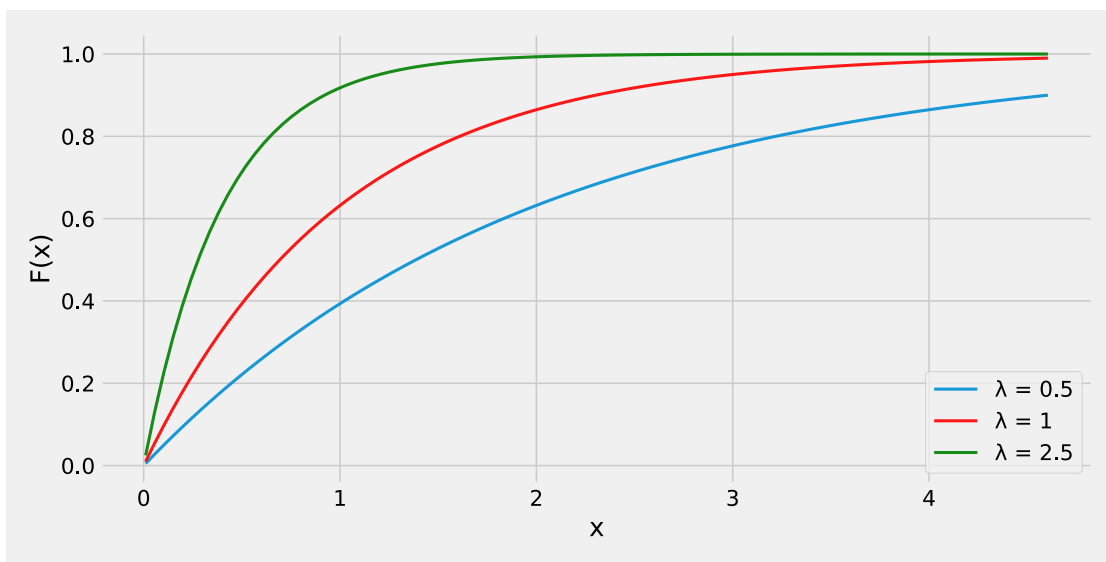
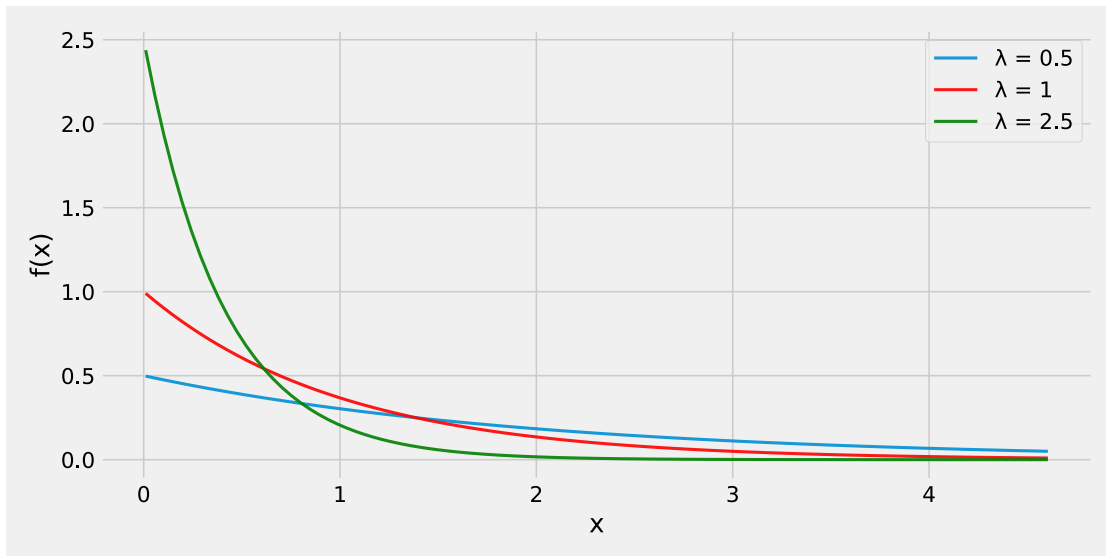
$$\begin{aligned} F_X(x) &= \int_0^x f_X(y) dy \\ &= \int_0^x \lambda e^{-\lambda y} dy \\ &\rightarrow z = \lambda y, \quad dz = \lambda dy \\ &= \int_0^{\lambda x} e^{-z} dz \\ &= -e^{-z} \Big|_0^{\lambda x} \\ &= -e^{-\lambda x} + e^0 \\ &= (1 - e^{-\lambda x}) \mathbb{I}_{R^+}(x) \end{aligned} \tag{14}$$

• Valore atteso:

$$\begin{aligned} \mathbb{E}(X) &= \int_0^\infty x f_X(x) dx \\ &= \int_0^\infty x \lambda e^{-\lambda x} dx \\ &\rightarrow f(x) = x, \quad g'(x) = \lambda e^{-\lambda x}, \quad g(x) = -e^{-\lambda x} \\ &= -x e^{-\lambda x} \Big|_0^\infty + \int_0^\infty e^{-\lambda x} dx \\ &= 0 + \int_0^\infty e^{-\lambda x} dx \\ &= \int_0^\infty e^{-\lambda x} dx \\ &\rightarrow \text{introduco } \lambda \\ &= \frac{1}{\lambda} \int_0^\infty \lambda e^{-\lambda x} dx \\ &\rightarrow \text{guardo la Considerazione 1 qui sotto} \\ &= \frac{1}{\lambda} \cdot 1 \\ &= \frac{1}{\lambda} \end{aligned} \tag{15}$$

• **Varianza:**

$$\begin{aligned}
 \text{Var}(X) &= \mathbb{E}(X^2) - \mathbb{E}(X)^2 \\
 &= \mathbb{E}(X^2) - \left(\frac{1}{\lambda}\right)^2 \\
 &= \mathbb{E}(X^2) - \frac{1}{\lambda^2} \\
 &= \left(\int_0^\infty x^2 \lambda e^{-\lambda x} dx\right) - \frac{1}{\lambda^2} \\
 &\rightarrow f(x) = x^2, \quad g'(x) = \lambda e^{-\lambda x}, \quad g(x) = -e^{-\lambda x} \\
 &= \left(-x^2 e^{-\lambda x} \int_0^\infty + \int_0^\infty 2x e^{-\lambda x} dx\right) - \frac{1}{\lambda^2} \\
 &= \left(0 + 2 \int_0^\infty x e^{-\lambda x} dx\right) - \frac{1}{\lambda^2} \\
 &= \left(2 \int_0^\infty x e^{-\lambda x} dx\right) - \frac{1}{\lambda^2} \\
 &\rightarrow \text{introduco } \lambda \\
 &= \left(\frac{2}{\lambda} \int_0^\infty \lambda x e^{-\lambda x} dx\right) - \frac{1}{\lambda^2} \\
 &\rightarrow f(x) = x, \quad g'(x) = \lambda e^{-\lambda x}, \quad g(x) = -e^{-\lambda x} \\
 &= \left(\frac{2}{\lambda} \cdot \left(-x e^{-\lambda x} \int_0^\infty + \int_0^\infty e^{-\lambda x} dx\right)\right) - \frac{1}{\lambda^2} \\
 &= \left(\frac{2}{\lambda} \cdot \left(0 + \int_0^\infty e^{-\lambda x} dx\right)\right) - \frac{1}{\lambda^2} \\
 &= \left(\frac{2}{\lambda} \cdot \left(\int_0^\infty e^{-\lambda x} dx\right)\right) - \frac{1}{\lambda^2} \\
 &\rightarrow \text{introduco } \lambda \\
 &= \left(\frac{2}{\lambda} \cdot \left(\frac{1}{\lambda} \int_0^\infty \lambda e^{-\lambda x} dx\right)\right) - \frac{1}{\lambda^2} \\
 &\rightarrow \text{guardo la Considerazione 1 qui sotto} \\
 &= \left(\frac{2}{\lambda} \cdot \left(\frac{1}{\lambda} \cdot 1\right)\right) - \frac{1}{\lambda^2} \\
 &= \left(\frac{2}{\lambda} \cdot \frac{1}{\lambda}\right) - \frac{1}{\lambda^2} \\
 &= \frac{2}{\lambda^2} - \frac{1}{\lambda^2} \\
 &= \frac{1}{\lambda^2} \\
 \\
 &\implies \sigma_X = \mathbb{E}(X)
 \end{aligned} \tag{16}$$



6.2.1 Considerazione 1

$$\begin{aligned}
 \int_0^{\infty} f_X(x) dx &= \int_0^{\infty} \lambda e^{-\lambda x} dx \\
 &\rightarrow y = \lambda x, \quad dy = \lambda dx \\
 &= \int_0^{\infty} e^{-y} dy \\
 &= -e^{-y} \Big|_0^{\infty} \\
 &= 0 + e^{-0} \\
 &= 1
 \end{aligned}
 \tag{17}$$

6.2.2 Considerazione 2

Come il modello geometrico, così il modello esponenziale gode dell'assenza di memoria. Ipotizziamo infatti di avere giustamente

$$F_X(x) = 1 - e^{-\lambda x} \tag{18}$$

E dunque

$$\begin{aligned} P(X > x) &= 1 - F_X(x) \\ &= e^{-\lambda x} \end{aligned} \tag{19}$$

Quindi

$$\begin{aligned} P(X > s + t) &= e^{-\lambda(s+t)} \\ &= e^{-\lambda s} \cdot e^{-\lambda t} \end{aligned} \tag{20}$$

Infatti

$$\begin{aligned} P(X > s + t) &= P(X > s) \cdot P(X > t) \\ \rightarrow P(X > s) &= \frac{P(X > s + t)}{P(X > t)} \\ \implies P(X > s + t \mid X > t) &= P(X > s) \end{aligned} \tag{21}$$

6.2.3 Considerazione 3

Ora immagino x_1, \dots, x_n indipendenti e $Y_i = \max_i X_i$, allora:

$$\begin{aligned} F_Y(x) &= P(Y \leq x) \\ &= P(\max_i X_i \leq x) \\ &\rightarrow \max_i X_i \leq x \iff \forall i X_i \leq x \\ &= P(\forall i X_i \leq x) \\ &\rightarrow \text{per indipendenza } P(X_1 > x, X_2 > x, \dots) \\ &= \prod_{i=1}^n P(X_i \leq x) \\ &= \prod_{i=1}^n F_{X_i}(x) \end{aligned} \tag{22}$$

Quindi se X_1, \dots, X_n sono i.i.d. (*indipendenti e identicamente distribuite*) secondo F allora

$$\begin{aligned} F_Y(x) &= \prod_{i=1}^n F(x) \\ &= F(x)^n \end{aligned} \tag{23}$$

Per cui, date X_1, \dots, X_n indipendenti (e non i.i.d), allora

$$\begin{aligned}
Z &= \min_{i=1, \dots, n} X_i \\
&= P(Z > x) \\
&\rightarrow \min_i X_i > x \iff \forall i X_i > x \\
&= P(\forall i X_i > x) \\
&= \prod_{i=1}^n P(X_i > x) \\
&= \prod_{i=1}^n (1 - P(X_i \leq x)) \\
&= \prod_{i=1}^n (1 - F_{X_i}(x)) \\
&\rightarrow \text{possiamo vedere } P(Z > x) \text{ come } 1 - P(Z \leq x) \\
&= 1 - \prod_{i=1}^n (1 - F_{X_i}(x)) \\
&\rightarrow X_i \text{ i.i.d secondo } F_Z(x) = 1 - (1 - F_Z(x))^n \\
&\rightarrow X_i \sim E(\lambda_i) \text{ e indipendente} \\
&\rightarrow F_{X_i}(x) = 1 - e^{-\lambda_i x} \\
&= 1 - \prod_{i=1}^n (e^{-\lambda_i x}) \\
&= 1 - e^{\sum_{i=1}^n -\lambda_i x} \\
&= 1 - e^{-x \sum_{i=1}^n \lambda_i} \\
&= 1 - e^{-x\lambda} \\
&= 1 - e^{-\lambda x} \\
&\implies F_Z = 1 - e^{-\lambda x} \\
&\implies Z \sim E(\lambda)
\end{aligned} \tag{24}$$

Ho dimostrato che se considero X_1, \dots, X_n e $\forall i X_i \sim E(\lambda_i)$ allora $\min X_i \sim E(\sum_{i=1}^n \lambda_i)$

6.2.4 Considerazione 4

Dato $X \sim E(\lambda)$ con $c \in R^+$ e $Y = c \cdot X$, allora:

$$\begin{aligned}
F_Y(x) &= P(Y \leq x) \\
&= P(cX \leq x) \\
&= P\left(X \leq \frac{x}{c}\right) \\
&= F_X\left(\frac{x}{c}\right) \\
&= 1 - e^{-\lambda \frac{x}{c}} \\
&= 1 - e^{-\frac{\lambda}{c} x}
\end{aligned} \tag{25}$$

6.2.5 Esempio 1

La mia $X \sim E(1)$ rappresenta le ore necessarie per riparare un macchinario, qual è la probabilità che io ci impieghi più di 2 ore per ripararlo?:

$$\begin{aligned} P(X > 2) &= 1 - F_X(2) \\ &= 1 - (1 - e^{-2}) \\ &= e^{-2} \\ &\approx 0.1353 \end{aligned} \tag{26}$$

Qual è invece la probabilità che io ci impieghi 3 ore dato che prima ce ne ho messe 2 di ore?

$$\begin{aligned} P(X > 3 \mid X > 2) &= P(X > 1 + 2 \mid X > 2) \\ &= P(X > 1) \\ &= e^{-1} \\ &\approx 0.3678 \end{aligned} \tag{27}$$

6.2.6 Esempio 2

La mia $X \sim E\left(\frac{1}{20}\right)$ rappresenta le migliaia di km che l'auto percorre prima di rompersi. Ho usato l'auto per 10k km, la vendo e il compratore si domanda la probabilità che duri altri 20km, quindi:

$$\begin{aligned} P(X > 30 \mid X > 10) &= P(X > 20 + 10 \mid X > 10) \\ &= P(X > 20) \\ &= e^{-\frac{20}{20}} \\ &= e^{-1} \\ &\approx 0.3678 \end{aligned} \tag{28}$$

Come cambierebbe la risposta se $X \sim U((0, 40))$?

$$\begin{aligned} P(X > 30 \mid X > 10) &= \frac{P(X > 30 \cap X > 10)}{P(X > 10)} \\ &\rightarrow \text{se si verifica } X > 30 \text{ allora si è verificato anche } X > 10 \\ &= \frac{P(X > 30)}{P(X > 10)} \\ &= \frac{P(30 < X < 40)}{P(10 < X < 40)} \\ &= \frac{F_X(40) - F_X(30)}{F_X(40) - F_X(10)} \\ &= \frac{\frac{40-0}{40-0} - \frac{30-0}{40-0}}{\frac{40-0}{40-0} - \frac{10-0}{40-0}} \\ &= \frac{\frac{10}{40}}{\frac{30}{40}} \\ &= \frac{1}{3} \end{aligned} \tag{29}$$

Questo modello è noto anche come *Modello Normale*, tuttavia a volte specifica un caso particolare del Modello Gaussiano.

- **Supporto:** $D_x = \mathbb{R}$

- **Funzione di densità:**

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (31)$$

- **Funzione di ripartizione:**

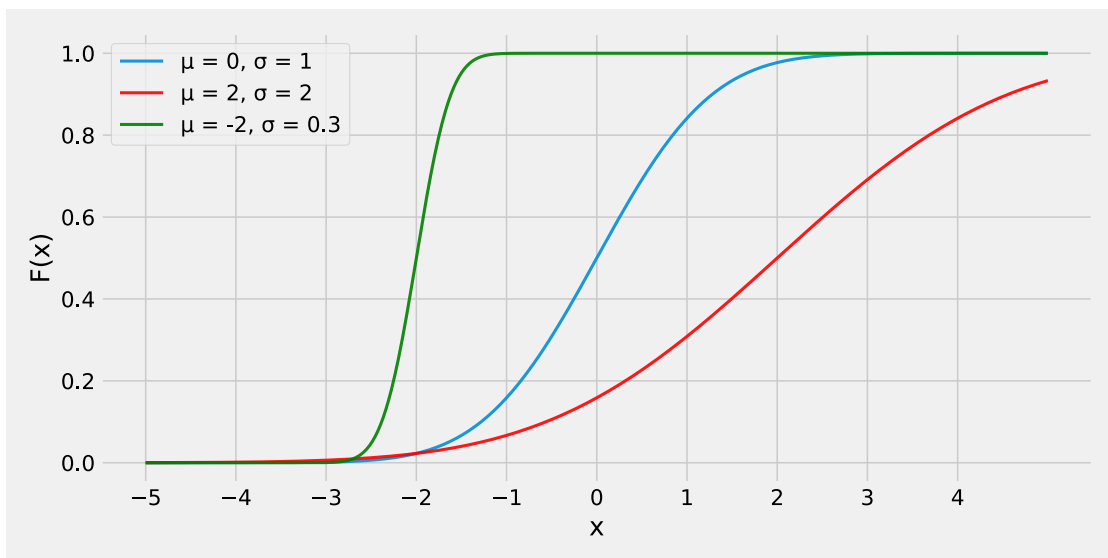
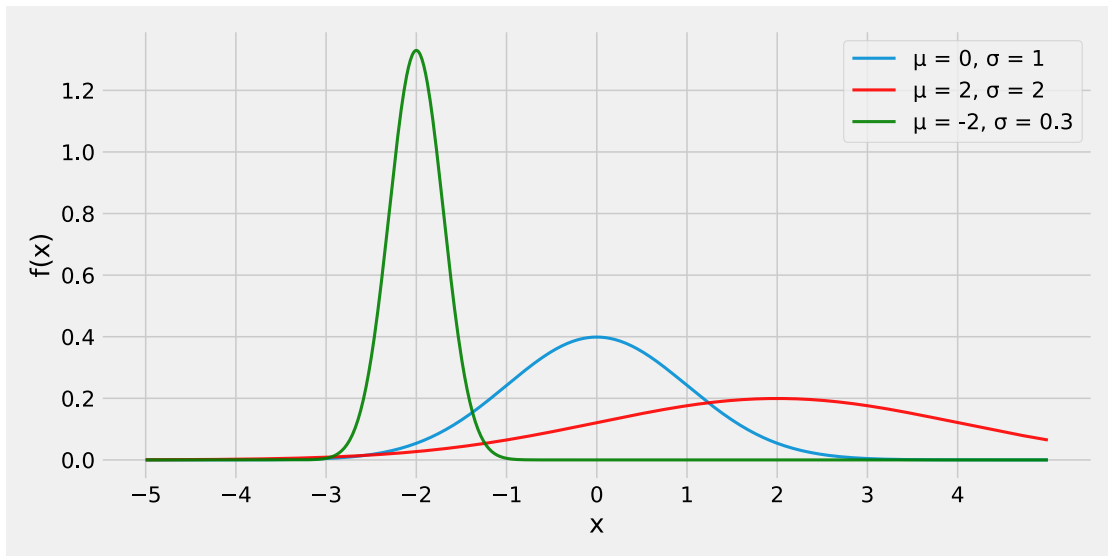
$$F_X(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} du \quad (32)$$

- **Valore atteso:**

$$E(X) = \mu \quad (33)$$

- **Varianza:**

$$Var(X) = \sigma^2 \quad (34)$$



6.3.1 Considerazione 1

Consideriamo

$$\lim_{x \rightarrow \infty} f_X(x) = \lim_{x \rightarrow \infty} \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (35)$$

Per prima cosa calcoliamo la derivata prima $f'_X(x)$:

$$\begin{aligned} f'_X(x) &= -\frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} \cdot \frac{1}{2\sigma^2} 2(x-\mu) \\ &= -\frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} \cdot \frac{1}{\sigma^2} (x-\mu) \\ &= -\frac{1}{\sqrt{2\pi}\sigma^3} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} \cdot (x-\mu) \\ &= \frac{1}{\sqrt{2\pi}\sigma^3} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} \cdot (\mu-x) \\ &\rightarrow f'_X(x) \geq 0 \text{ quando } \mu-x \geq 0 \rightarrow x \leq \mu \end{aligned} \quad (36)$$

\Rightarrow cresce verso μ per poi decrescere

Ora calcoliamo la derivata seconda $f''_X(x)$:

$$\begin{aligned} f''_X(x) &= \frac{1}{\sqrt{2\pi}\sigma^3} \cdot \left(e^{-\frac{(x-\mu)^2}{2\sigma^2}} \cdot \left(-\frac{2(x-\mu)}{2\sigma^2} \right) \cdot (\mu-x) - e^{-\frac{(x-\mu)^2}{2\sigma^2}} \right) \\ &\quad \frac{1}{\sqrt{2\pi}\sigma^3} \cdot \left(e^{-\frac{(x-\mu)^2}{2\sigma^2}} \cdot \left(-\frac{(x-\mu)}{\sigma^2} \right) \cdot (\mu-x) - e^{-\frac{(x-\mu)^2}{2\sigma^2}} \right) \\ &\quad \frac{1}{\sqrt{2\pi}\sigma^3} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} \cdot \left(\frac{(x-\mu)^2}{\sigma^2} - 1 \right) \\ &\rightarrow f''_X(x) \geq b \text{ quando } \frac{(x-\mu)^2}{\sigma^2} - 1 \geq 0 \rightarrow \left(\frac{x-\mu}{\sigma} \right)^2 \geq 1 \end{aligned} \quad (37)$$

Questi calcoli dimostrano che non solo $G(\mu, \sigma)$ è convessa fino a $\mu - \sigma$ e da $\mu + \sigma$, con μ al centro, ma anche che quindi essa assumerà la classica forma a campana. Inoltre se abbiamo $\mu_2 < \mu_1$, allora μ_2 farà traslare la funzione a sinistra rispetto a μ_1 , siccome stiamo spostando a sinistra il valore atteso; mentre se $\sigma_2 < \sigma_1$ allora avremo un picco, sempre in μ ma più evidenziato per σ_2 , in quando stiamo dicendo che i dati sono meno devianti rispetto a σ_1 e quindi più concentrati.

6.3.2 Considerazione 2

$$\begin{aligned} f_X(\mu) &= \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{(\mu-\mu)^2}{2\sigma^2}} \\ &= \frac{1}{\sqrt{2\pi}\sigma} \end{aligned} \quad (38)$$

Quindi la densità in μ sarà più grande tanto più σ è piccola.

6.3.3 Considerazione 3

Ipotizzo di essere abbastanza confidente del fatto che

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\left(\frac{x-\mu}{\sigma}\right)^2} dx = 1 \quad (39)$$

Ma come lo provo?

Innanzitutto considero $z = \frac{x-\mu}{\sigma}$ e quindi $dz = \frac{dx}{\sigma}$, dunque

$$\begin{aligned}
 I &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2} dz \\
 &\rightarrow I \cdot I \\
 &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-x^2} dx \int_{-\infty}^{\infty} e^{-y^2} dy \\
 &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-x^2} e^{-y^2} dx dy \\
 &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-(x^2+y^2)} dx dy \\
 &\rightarrow p \sin \theta = y, \quad p \cos \theta = x \text{ e matrice jacobiana} \\
 &= 1
 \end{aligned} \tag{40}$$

6.3.4 Considerazione 4

Partendo da $F_X(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} du$ io so che $\mu = \mathbb{E}(X)$ e che $\sigma^2 = Var(X)$, dunque se io prendo $y = ax + b$, se $X \sim G(\mu, \sigma)$ allora $Y \sim G(a\mu + b, a\sigma)$.

Inoltre so che se ho $X_1 \sim G(\mu_1, \sigma_1)$ e $X \sim G(\mu_2, \sigma_2)$ più indipendenza, allora avrò

$$X_1 + X_2 \sim G\left(\mu_1 + \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2}\right) \tag{41}$$

N.B. Sto sommando variabili aleatorie, non funzioni. Nel caso della somma di funzioni non posso fare come ho scritto e il risultato anzi sarà una funzione multimodale.

6.3.5 Considerazione 5

Se usassi $Y = \frac{X-\mu}{\sigma}$ otterrei

$$\begin{aligned}
 \mathbb{E}(Y) &= \mathbb{E}\left(\frac{X-\mu}{\sigma}\right) \\
 &= \frac{1}{\sigma} \mathbb{E}(X-\mu) \\
 &= \frac{1}{\sigma} (\mathbb{E}(X) - \mu) \\
 &= \frac{1}{\sigma} (\mathbb{E}(X) - \mathbb{E}(X)) \\
 &= 0
 \end{aligned} \tag{42}$$

Inoltre:

$$\begin{aligned}
 Var(Y) &= Var\left(\frac{X-\mu}{\sigma}\right) \\
 &= \frac{1}{\sigma^2} Var(X-\mu) \\
 &= \frac{1}{\sigma^2} Var(X) \\
 &= 1
 \end{aligned} \tag{43}$$

Dunque otteniamo

$$X \sim G(\mu, \sigma) \rightarrow Y \sim G(0, 1) \tag{44}$$

Dunque $Y = \frac{X-\mu}{\sigma}$ non è altro che la gaussiana normale standard. Essa mi è utile perché tramite tabulati posso convertire i miei dati per usarli nella gaussiana standard.

6.3.6 Legge empirica

Se un campione numerico è approssimativamente normale, ha media campionaria \bar{x} e deviazione standard s , allora:

- (1) Circa il 68% dei suoi dati cade nell'intervallo $\bar{x} \pm s$
- (2) Circa il 95% dei suoi dati cade nell'intervallo $\bar{x} \pm 2s$
- (3) Circa il 99.7% dei suoi dati cade nell'intervallo $\bar{x} \pm 3s$

6.3.7 Considerazione 6

Partendo dalla Considerazione 4, se invece di applicare una trasformazione lineare io avessi usato variabili aleatorie gaussiane, il risultato sarebbe stata una variabile aleatoria gaussiana, pertanto, avendo X_1, \dots, X_n indipendenti, $\forall i X_i \sim G(\mu_i, \sigma_i)$ allora $Y = \sum_{i=1}^n X_i$. In particolare $\mathbb{E}(Y) = \sum_{i=1}^n \mu_i$, $Var(Y) = \sum_{i=1}^n \sigma_i^2$ e nel complesso

$$Y \sim G\left(\sum_{i=1}^n \mu_i, \sqrt{\sum_{i=1}^n \sigma_i^2}\right) \quad (45)$$

Quindi sommando variabili aleatorie che hanno un modello comune, ottengo una variabile aleatoria che fa ancora parte di quel modello comune, questo concetto è noto come **riproducibilità**. Quindi se X è una variabile casuale con media μ e deviazione standard σ , allora la standardizzazione di X tramite la trasformazione $Z = \frac{X-\mu}{\sigma} \sim G(0, 1)$ segue una distribuzione normale standard $N(0, 1)$. In altre parole, la variabile standardizzata Z ha media 0 e deviazione standard 1, indipendentemente dalla distribuzione originale di X . Questo processo rende le misurazioni comparabili, poiché esprime ciascun valore in termini di deviazioni standard dal valore atteso, facilitando così la valutazione delle relazioni e delle differenze tra dati che potrebbero avere scale diverse.

E' importante notare che la riproducibilità vale, oltre che per le variabili aleatorie gaussiane, anche per le variabili aleatorie poissoniane e per le binomiali, in condizioni specifiche, tant'è che

$$X_1 \sim B(n, p) \quad X_2 \sim B(m, p) \quad X_1 + X_2 \sim B(n + m, p) \quad (46)$$

$$X_1 \sim P(\lambda_1) \quad X_2 \sim P(\lambda_2) \quad X_1 + X_2 \sim P(\lambda_1 + \lambda_2) \quad (47)$$

A proposito della legge di riproducibilità, è giusto tenere in considerazione che la differenza di variabili aleatorie distribuite secondo un modello normale è anch'essa una variabile aleatoria normale. Infine, la distribuzione gaussiana e la distribuzione esponenziale, godono entrambe del concetto di scalabilità.

6.3.8 Considerazione 7

Cosa succede se voglio calcolare $F_X(x)$? Viene un integrale non risolvibile in forma analitica, quindi uso la Normale Standard:

$$\begin{aligned} F_X(x) &= P(X \leq x) \\ &= P\left(\frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma}\right) \\ &= P\left(Z \leq \frac{x - \mu}{\sigma}\right) \\ &= F_Z\left(\frac{x - \mu}{\sigma}\right) \\ &= \Phi \end{aligned} \quad (48)$$

E nel dettaglio valgono le seguenti proprietà:

- $\Phi(0) = \frac{1}{2}$
- $\Phi(-x) = P(Z > x) = 1 - P(Z \leq x) = 1 - \Phi(x)$

- $P(|X - \mu| \leq n\sigma) \implies P(-n \leq Z \leq n) = P(|Z| \leq n)$

In generale:

$$\begin{aligned}
 P(-n \leq Z \leq n) &= P(|Z| \leq n) \\
 &= \Phi(n) - \Phi(-n) \\
 &= \Phi(n) - (1 - \Phi(n)) \\
 &= 2\Phi(n) - 1
 \end{aligned}
 \tag{49}$$

6.3.9 Quantili

La mediana di X è $m \in \mathbb{R}$ tale che $P(X \leq m) = P(X \geq m) = \frac{1}{2}$ (ossia 0, il valore atteso).

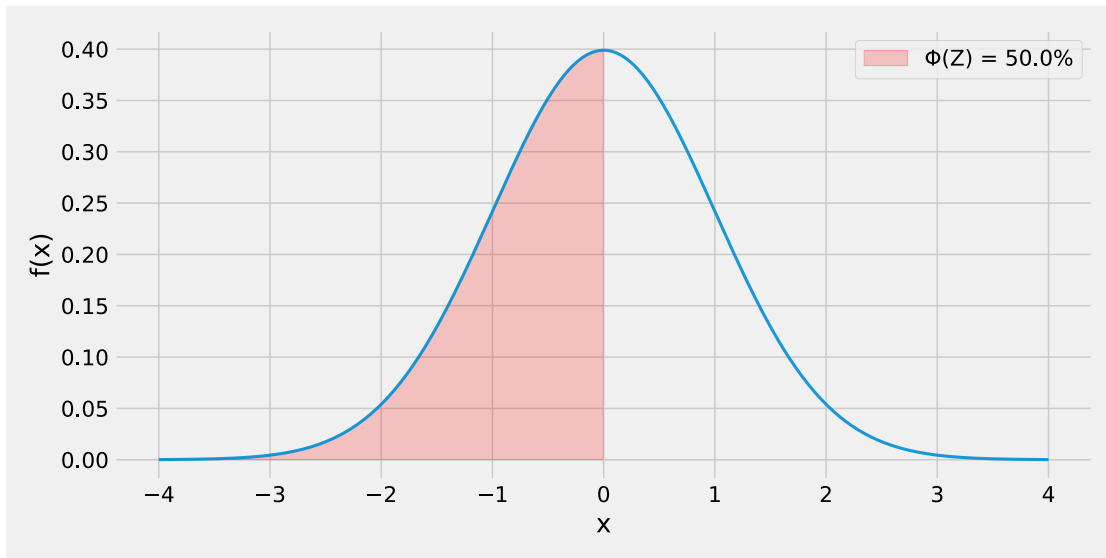
La moda di X è la specificazione che massimizza la densità o la massa di probabilità (ossia 0, il valore atteso).

Il quantile di livello $q \in [0, 1]$ di X è la specificazione $x_q \in \mathbb{R}$ tale che $P(X \leq x) = q$. Dunque pongo $X \sim G(\mu, \sigma)$, da qui, se volessi calcolare $q = P(X \leq 2)$ procedo come

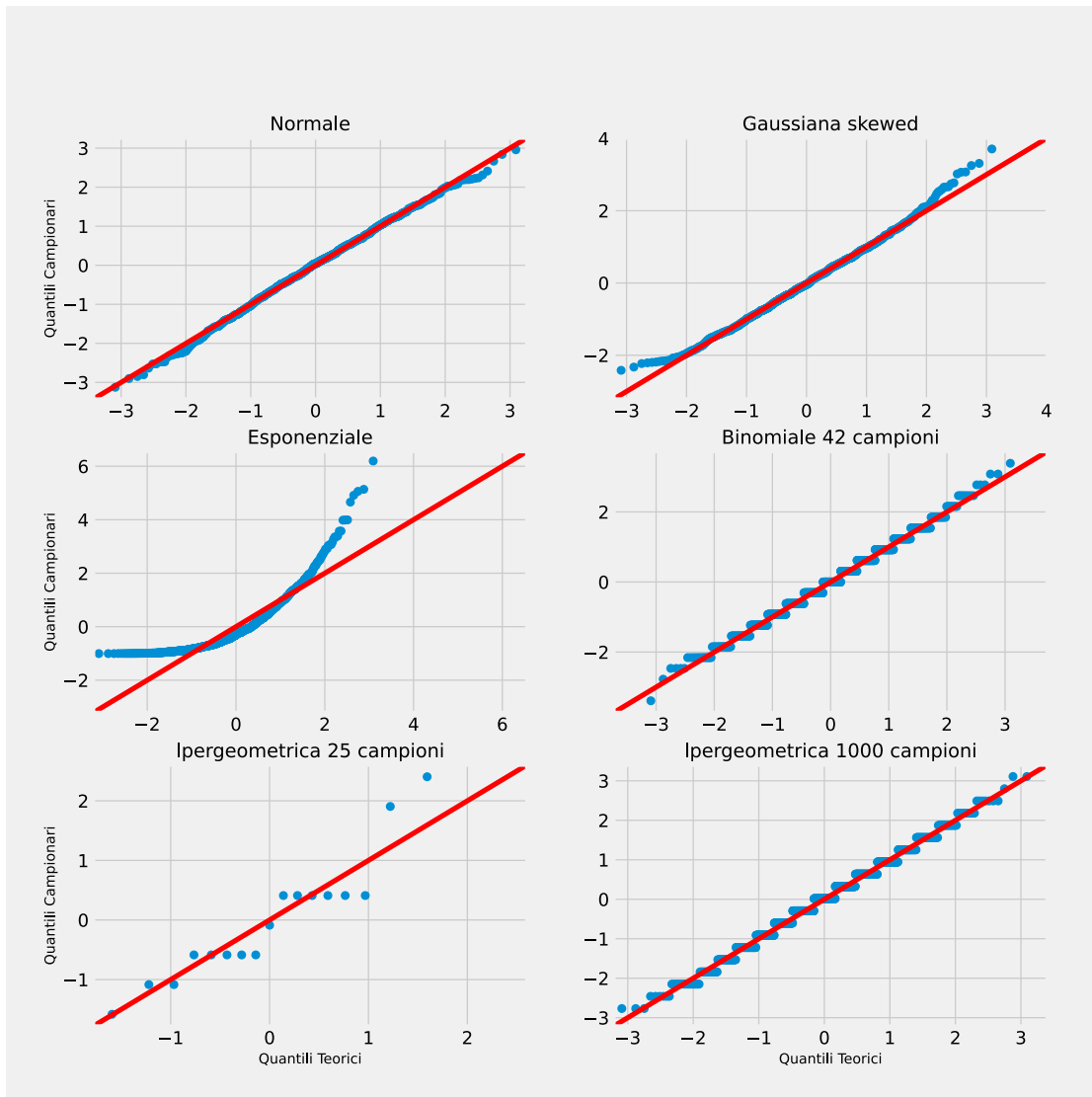
$$\begin{aligned}
 q &= P(X \leq x_q) \\
 &= F_X(x_q) \\
 &= F_X^{-1}(q)
 \end{aligned}
 \tag{50}$$

Da qui

- $F_X(F_X^{-1}(q))$
- $F_X^{-1}(q) = x_q$ (Funzione ppf di Python)
- $F_X(x_q) = q$



Risulta importante notare che il quantile Q1 e il quantile Q3 si trovano rispettivamente a $\mu - z \cdot \sigma$ e $\mu + z \cdot \sigma$, con $z = 0.675$. Questo perché osservando le tabelle della distribuzione cumulativa normale standardizzata, possiamo notare che $\Phi(-0.675) = 0.25 = 25\%$ e $\Phi(0.675) = 0.75 = 75\%$.



Per disegnare un Q-Q plot, ossia i grafici sovrastanti, basta seguire i seguenti passi:

- (1) Raccolgo i dati e mi assicuro di avere un numero abbastanza elevato di valori così da ottenere un'analisi significativa
- (2) Ordino i dati in ordine crescente
- (3) Calcolo i quantili teorici, solitamente di una distribuzione normale standard, e i quantili campionari
- (4) Creo una coppia di coordinate (x_i, y_i) dove le x_i saranno i quantili teorici mentre le y_i quelli campionari
- (5) Traccio la linea bisettrice di riferimento per osservare se i punti giacciono su essa

Esempio

Si determini il più piccolo valore di X , chiamiamolo $x_{0.8}$, tale che $P(X \leq x_{0.8}) \geq 0.8$. Questo non è altro che $F_X(x) \geq 0.8$ ossia il quantile 0.8 di X , che può essere facilmente calcolato con `X.ppf(0.8)`

6.3.10 Teorema Centrale del Limite

Il Teorema Centrale del Limite o Teorema del Limite Centrale asserisce che, dati X_1, \dots, X_n i.i.d. allora $\forall i \mathbb{E}(X_i) = \mu$, $Var(X_i) = \sigma^2$ si ha che

$$\sum_{i=1}^n X_i \sim N(n\mu, \sqrt{n}\sigma) \quad (51)$$

E quindi, posto

$$Y_n = \frac{\frac{1}{n} \sum_{j=1}^n X_j - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1) \quad (52)$$

Allora

$$\frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n}\sigma} \sim N(0, 1) \quad (53)$$

E dunque

$$\lim_{n \rightarrow \infty} P\left(\frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n}\sigma} \leq x\right) = \Phi(x) \quad (54)$$

Dove tanto più è grande n , tanto migliore sarà l'approssimazione.

Empiricamente si evidenzia come per $n \geq 30$ si iniziano ad avere risultati accettabili. In ogni caso se voglio verificare qualitativamente la bontà dell'approssimazione posso usare un QQ-plot.

Il TCL è valido per ognuna delle distribuzioni viste in questo corso e ricordiamoci infine che quindi, nel caso della media campionaria, avremo

$$\frac{1}{n} \sum_{i=1}^n X_i \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \quad (55)$$

6.3.11 Teorema di De Moivre-Laplace

Una distribuzione normale può essere usata per approssimare una distribuzione binomiale, a patto di avere un parametro n sufficientemente grande. Dato $X \sim B(n, p)$ con n grande, allora

$$X = \sum_{i=1}^n X_i \sim N\left(np, \sqrt{np(1-p)}\right) \rightarrow \frac{X - np}{\sqrt{np(1-p)}} \sim N(0, 1) \quad (56)$$

In generale

$$\frac{X - \mu}{\sqrt{Var(X)}} \sim N(0, 1) \quad (57)$$

N.B. Su Python andranno usati i comandi `Z = st.norm()` per definire una distribuzione normale e `Z.rvs()` per ricevere una specificazione plausibile per una variabile aleatoria della distribuzione normale create precedentemente. Entrambe le funzioni appartengono al modulo `scipy.stats`.

6.3.12 Esempio 1

Per $n > 1$ si controlli che $P(-\varepsilon < T_n - p < \varepsilon) \approx 2\Phi\left(\frac{\varepsilon\sqrt{n}}{\sigma}\right) - 1$:

$$\begin{aligned} P(-\varepsilon < T_n - p < \varepsilon) &\approx P\left(-\frac{\varepsilon\sqrt{n}}{\sigma} < Z < \frac{\varepsilon\sqrt{n}}{\sigma}\right) \\ &\approx 2\Phi\left(\frac{\varepsilon\sqrt{n}}{\sigma}\right) - 1 \end{aligned} \quad (58)$$

6.3.13 Esempio 2

Verifico che $P(\mu - k \leq X \leq \mu + k) \leq 2\Phi\left(\frac{k}{\sigma}\right) - 1$:

$$\begin{aligned}
 P(\mu - k \leq X \leq \mu + k) &= P(\mu \leq X - k \leq \mu) \\
 &= P(\mu \leq X^* \leq \mu) \\
 &= F_X\left(\frac{k}{\sigma}\right) - F_X\left(-\frac{k}{\sigma}\right) \\
 &\approx \Phi\left(\frac{k}{\sigma}\right) - \Phi\left(-\frac{k}{\sigma}\right) \\
 &= \Phi\left(\frac{k}{\sigma}\right) - \left(1 - \Phi\left(\frac{k}{\sigma}\right)\right) \\
 &= 2\Phi\left(\frac{k}{\sigma}\right) - 1
 \end{aligned} \tag{59}$$

6.3.14 Esempio 3

Ipotizzo di avere 25000 polizze auto e X è il risarcimento annuo di un cliente. Mi aspetto di risarcire intorno ai 320 euro ogni cliente, quindi $\mathbb{E}(X) = 320$ euro, con una deviazione di $\sigma_x = 540$ euro: qual è la probabilità di dover risarcire più di $8.3 \cdot 10^6$ euro?

$$\begin{aligned}
 P(\text{più di } 8.3 \cdot 10^6 \text{ di risarcimento/anno}) &= P(Y > 8.3 \cdot 10^6) \\
 &\rightarrow \text{per il TCL } \sum_{i=1}^n X_i \sim N(n\mu, \sqrt{n}\sigma) \\
 &\rightarrow \sum_{i=1}^{25000} X_i = Y \sim N(25000 \cdot 320, \sqrt{25000} \cdot 540) \\
 &= Y \sim N(8 \cdot 10^6, 8.54 \cdot 10^6) \\
 &\rightarrow \text{standardizzo} \\
 &= P\left(\frac{Y - 8 \cdot 10^6}{8.54 \cdot 10^6} > \frac{8.3 \cdot 10^6 - 8 \cdot 10^6}{8.54 \cdot 10^6}\right) \\
 &= P((\sim N(0, 1)) > (\approx 3.51)) \\
 &\approx P(Z > 3.51) \\
 &= 1 - \Phi(3.51) \\
 &\approx 2.2 \cdot 10^{-4}
 \end{aligned} \tag{60}$$

Tuttavia sto approssimando due volte, qual è la soglia entro cui questa approssimazione continua a valere? Lo domando alla *Funzione Cumulativa Empirica (ECDF)*

$$\hat{F}(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{(-\infty, x]}(X_i) \tag{61}$$

E quindi

$$\hat{F}(8.3 \cdot 10^6) = \frac{1}{25000} \sum_{i=1}^n \mathbb{I}_{(-\infty, 8.3 \cdot 10^6]}(X_i) \tag{62}$$

Ipotizzo (con dati immaginari) che $\hat{F}(8.3 \cdot 10^6) \approx 0.85$, questo significa che l'85% delle osservazioni è inferiore o uguale a $8.3 \cdot 10^6$ e che quindi l'approssimazione normale sembra funzionare abbastanza bene fino a quel punto di divergenza. Quindi se per esempio avessi avuto bisogno di una precisione del 95% quel valore non sarebbe andato bene.

6.3.15 Esempio 4

Il 30% degli iscritti è frequentante, quindi $P(\text{uno studente sia frequentante}) = 0.3$. Modelliamo X come il numero di studenti frequentanti su 450, tale che $X \sim B(450, 0.3)$. La nostra aula ha 150 posti: qual è la probabilità di superare la capienza massima dell'aula?

Sappiamo che

$$X \dot{\sim} N\left(450 \cdot 0.3, \sqrt{450 \cdot 0.3 \cdot 0.7}\right) \rightarrow X \dot{\sim} N(135, 9.72) \quad (63)$$

Quindi servendoci della distribuzione normale standard $Z \sim N(0, 1)$

$$\begin{aligned} P(X > 150) &= P\left(\frac{X - 135}{9.72} > \frac{150 - 135}{9.72}\right) \\ &\approx P(Z > 1.59) \\ &\approx 0.06 \end{aligned} \quad (64)$$

Dunque ho un 6% di giornate di lezione in cui gli studenti iscritti non trovano posto in aula.

6.3.16 Esempio 5

La media delle precipitazioni del 2022 è 12.08cm mentre la deviazione standard è 9.1cm. Le precipitazioni del 2022 sommate a quelle del 2023, supereranno i 25cm?

Sappiamo che $X_{2022}, X_{2023} \sim N(\mu, \sigma)$, dunque

$$Y = X_{2022} + X_{2023} \sim N\left(2\mu, \sqrt{2}\sigma\right) \quad (65)$$

Allora

$$\begin{aligned} P(X_{2022} + X_{2023} > 25) &= P(Y > 25) \\ &= P\left(\frac{Y - 2 \cdot 12.08}{\sqrt{2} \cdot 3.1} > \frac{25 - 2 \cdot 12.08}{\sqrt{2} \cdot 3.1}\right) \\ &= P(Z > 0.1916) \\ &\approx 0.4240 \end{aligned} \quad (66)$$

Cosa succede se volessi calcolare la probabilità che nel 2022 ci sia stata più pioggia (di almeno 3cm)?

$$P(X_{2022} > X_{2023} + 3) = P(X_{2022} - X_{2023} > 3)$$

Quindi se io definisco $T = X_{2022} - X_{2023}$, il cui risultato sarà una distribuzione normale standard, grazie alla proprietà della riproducibilità, allora

$$\begin{aligned} \mathbb{E}(T) &= \mathbb{E}(X_{2022}) - \mathbb{E}(X_{2023}) \\ &= 0 \end{aligned} \quad (67)$$

E infine

$$\begin{aligned} \text{Var}(T) &= \text{Var}(X_{2022} - X_{2023}) \\ &= \text{Var}(X_{2022}) + \text{Var}(X_{2023}) \\ &= 2\sigma^2 \end{aligned} \quad (68)$$

6.3.17 Esempio 6

So che $P(|\bar{X}_n - \mu| \leq \varepsilon)$, determino n dato α tale che $P(|\bar{X}_n - \mu| \leq \varepsilon) \geq \alpha$. Risolvo il seguente

$$\begin{aligned}
 P(|\bar{X}_n - \mu| \leq \varepsilon) &= P(-\varepsilon \leq \bar{X}_n - \mu \leq \varepsilon) \\
 &\rightarrow TCL \\
 &= P\left(-\frac{\varepsilon}{\frac{\sigma}{\sqrt{n}}} \leq \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \leq \frac{\varepsilon}{\frac{\sigma}{\sqrt{n}}}\right) \\
 &\approx P\left(-\frac{\varepsilon}{\frac{\sigma}{\sqrt{n}}} \leq Z \leq \frac{\varepsilon}{\frac{\sigma}{\sqrt{n}}}\right) \\
 &= \Phi\left(\frac{\varepsilon}{\frac{\sigma}{\sqrt{n}}}\right) - \Phi\left(-\frac{\varepsilon}{\frac{\sigma}{\sqrt{n}}}\right) \\
 &= 2\Phi\left(\frac{\varepsilon}{\frac{\sigma}{\sqrt{n}}}\right) - 1 \\
 &= 2\Phi\left(\frac{\varepsilon\sqrt{n}}{\sigma}\right) - 1 \tag{69} \\
 &\rightarrow 2\Phi\left(\frac{\varepsilon\sqrt{n}}{\sigma}\right) - 1 \geq \alpha \\
 &\rightarrow \Phi\left(\frac{\varepsilon\sqrt{n}}{\sigma}\right) \geq \frac{\alpha + 1}{2} \\
 &\rightarrow \frac{\varepsilon\sqrt{n}}{\sigma} \geq \Phi^{-1}\left(\frac{\alpha + 1}{2}\right) \\
 &\rightarrow \sqrt{n} \geq \Phi^{-1}\left(\frac{\alpha + 1}{2}\right) \cdot \frac{\sigma}{\varepsilon} \\
 &\rightarrow n \geq \left(\Phi^{-1}\left(\frac{\alpha + 1}{2}\right) \cdot \frac{\sigma}{\varepsilon}\right)^2
 \end{aligned}$$

Esercizi

6.4

Uno stabilimento ha 6 macchinari che usano in media energia elettrica per 20 minuti ogni ora.

- Se i macchinari vengono usati indipendentemente, mostrare che la probabilità che 4 o più macchinari usino energia elettrica contemporaneamente è 0.1.

Consideriamo la variabile casuale X = "numero di macchine che consumano energia", possiamo quindi assumere che $X \approx B(n, p)$, dove $n = 6$, numero totale di macchinari disponibili e $p = \frac{20}{60} = \frac{1}{3} = 0.33$, "successo" ossia minuti in cui i macchinari usano energia ogni ora. Dunque $X \approx B(6, 0.33)$. A questo punto basterà considerare il seguente

$$\begin{aligned}
 P(X \geq 4) &= P(X = 4) + P(X = 5) + P(X = 6) \\
 &= 0.0823 + 0.0165 + 0.0014 \\
 &= 0.1002 \tag{70}
 \end{aligned}$$

Ovviamente potevo anche considerare

$$P(X \geq 4) = 1 - P(X = 3) - P(X = 2) - P(X = 1) - P(X = 0) = \dots$$

- Se lo stabilimento avesse 60 macchinari, quale sarebbe la probabilità di avere al massimo 30 macchinari in funzione contemporaneamente?

Come prima $X \approx B(60, 0.33)$, ma siccome abbiamo un valore di n elevato, possiamo ricorrere all'approssimazione normale della distribuzione binomiale considerando

$$N\left(np, \sqrt{np(1-p)}\right)$$

Il nostro μ non sarà altro che $np = 60 \cdot \frac{1}{3} = 20$ mentre la varianza $np(1-p) = 20 \cdot \frac{2}{3} = 13.33$ e quindi la deviazione standard sarà uguale a $\sqrt{13.33} = 3.65$. A questo punto la probabilità di avere al massimo 30 macchinari in funzione risulta essere

$$\begin{aligned} P(X \leq 30) &\approx \Phi\left(\frac{30 + 0.5 - 20}{3.65}\right) \\ &= \Phi(2.877) \\ &= 0.998 \end{aligned} \tag{71}$$

- **Sempre considerando 60 macchinari trovare un numero approssimato r , tale che la probabilità che più di r macchinari usino energia elettrica allo stesso tempo sia 0.1.**

Vogliamo cercare r tale che

$$P(X > r) = 0.1$$

Dunque procediamo semplicemente nel seguente modo

$$\begin{aligned} P(X > r) &= 1 - P(X \leq r) \\ &= 1 - \Phi\left(\frac{r + 0.5 - 20}{3.65}\right) \\ &\rightarrow 1 - \Phi\left(\frac{r + 0.5 - 20}{3.65}\right) = 0.1 \\ &\rightarrow -\Phi\left(\frac{r - 19.5}{3.65}\right) = 0.1 - 1 \\ &\rightarrow \Phi\left(\frac{r - 19.5}{3.65}\right) = -0.1 + 1 \\ &\rightarrow \Phi\left(\frac{r - 19.5}{3.65}\right) = 0.9 \\ &\rightarrow \frac{r - 19.5}{3.65} = \Phi^{-1}(0.9) \\ &\rightarrow r - 19.5 = \Phi^{-1}(0.9) \cdot 3.65 \\ &\rightarrow r = 19.5 + \Phi^{-1}(0.9) \cdot 3.65 \\ &\rightarrow r = 19.5 + 1.29 \cdot 3.65 \\ &\rightarrow r = 19.5 + 4.70 \\ &\rightarrow r = 24.2 \\ &\rightarrow r \approx 24 \end{aligned} \tag{72}$$

Parte III

Statistica Inferenziale

Capitolo 7

Statistiche e Stimatori

Statistica

7.1

Definiamo una **popolazione** come una variabile aleatoria $X \sim F(\theta)$ di supporto D_x e un **campione** X_1, \dots, X_n i.i.d. come una conformazione aleatoria. Nota bene che il campione, non ancora osservato, differisce da x_1, \dots, x_n che invece rappresentano i valori osservati.

Una **statistica** o **stimatore**, non è altro che una funzione $t : D_x^n \rightarrow \mathbb{R}$ dove presi n valori del campione produce una stima $\hat{\tau} = t(X_1, \dots, X_n)$. Ovvero un algoritmo che applico, il cui input è il mio campione e l'output è la stima di qualcosa che non conosco. Così la mia statistica mi permette di conoscere F :

- **Statistica non parametrica** $\rightarrow F$ è completamente sconosciuta e non si fanno ipotesi sulla forma specifica della distribuzione sottostante
- **Statistica parametrica** $\rightarrow F$ è parzialmente sconosciuta e si assume che i valori siano generati da una particolare famiglia di distribuzioni

Quindi, data una variabile aleatoria $T = t(X_1, \dots, X_n)$ e dato che voglio stimare $\tau(\theta)$, cerco una stima $\hat{\tau} = t(X_1, \dots, X_n)$ tale che $\hat{\tau} \approx \tau(\theta)$.

Dunque, una statistica è una variabile aleatoria che è semplicemente una funzione dei dati di un campione, e una qualunque statistica il cui scopo sia quello di dare una stima di un parametro θ si dice stimatore di θ .

7.1.1 Esempio

Mi ritrovo con una popolazione $X \sim B(p)$ e voglio stimare $\Theta = p$. Ipotizzo che $t(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}$ sia una stima accettabile e lo verifico risolvendo il seguente

$$\begin{aligned}\mathbb{E}(\bar{X}) &= \mathbb{E}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) \\ &= \frac{1}{n} \sum_{i=1}^n \mathbb{E}(X_i) \\ &\rightarrow \text{le } X_i \text{ appartengono allo stesso campione, quindi hanno la stessa distribuzione di } X \\ &= \frac{1}{n} \sum_{i=1}^n \mathbb{E}(X) \\ &\rightarrow \mathbb{E}(X) = p \text{ perché lavoriamo con una variabile aleatoria bernoulliana} \\ &= \frac{1}{n} \sum_{i=1}^n p \\ &= \frac{1}{n} np \\ &= p\end{aligned}\tag{1}$$

Però cosa significa? Significa che per esempio farò un esperimento bernoulliano e otterrò certi risultati i quali, se faccio la media campionaria di quei valori, oscilleranno attorno a p .

In realtà capita spesso di non voler stimare direttamente θ , bensì un valore dipendente da esso come la varianza e il valore atteso. Quando approssimo voglio valutare la bontà della mia approssimazione secondo due metriche: il **bias** (o **distorsione**) e l'**errore quadratico medio**, noto come **MSE**.

Distorsione

7.2

Se ho θ e voglio stimarlo, allora deve essere che $\tau(\theta) = \theta$ e ricordiamo che $T = t(X_1, \dots, X_n)$ è una variabile aleatoria: dunque per farmi prediligere una variabile aleatoria anziché un'altra, devo avere $\mathbb{E}(T) = \tau(\theta)$. In questo caso quindi diciamo che t , la mia statistica, è uno stimatore **non distorto**, **non deviato** o **unbiased** di $\tau(\theta)$.

7.2.1 Esempio 1

Definisco con $X \sim E(\lambda)$ il numero di minuti attesi in coda. Sapendo che il parametro della mia distribuzione esponenziale è λ , suppongo di voler stimare il tempo medio di attesa $\mathbb{E}(X) = \frac{1}{\lambda} = \tau(\lambda)$ e per farlo provo ad usare la media campionaria \bar{X}

$$\begin{aligned}\mathbb{E}(\bar{X}) &= \mathbb{E}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) \\ &= \frac{1}{n} \sum_{i=1}^n \mathbb{E}(X_i) \\ &= \frac{1}{n} \sum_{i=1}^n \mathbb{E}(X) \\ &= \mathbb{E}(X) \\ &= \frac{1}{\lambda}\end{aligned}\tag{2}$$

Dunque la media campionaria \bar{X} è non deviata in questo caso, ma non solo: essa è sempre non deviata rispetto a $\mathbb{E}(X)$, il valore atteso della popolazione.

7.2.2 Esempio 2

Cambiando i dati all'esempio precedente, ipotizziamo che $X \sim B(n, p)$, dove conosco n ma ignoro p . Dunque $\theta = p$, $\tau(\theta) = \tau(p) = p$ e $\mathbb{E}(\bar{X}) = \mathbb{E}(X) = np$.

Se io adesso dovessi ipotizzare che non vale più np ma mp tale che $Y_{n,m} = X_{1,1}, \dots, X_{1,m}, \dots, X_{n,m}$, allora

$$\begin{aligned}\bar{X} &= \bar{Y} \\ &= \frac{1}{nm} \sum_{i=1}^{nm} Y_i \\ &= \frac{1}{nm} \sum_{i=1}^n X_i \\ &= \frac{1}{nm} \sum_{i=1}^n X_i \\ &= \frac{1}{m} \bar{X}\end{aligned}\tag{3}$$

Dunque sapendo che ora vale $\mathbb{E}(\bar{X}) = mp$ dobbiamo vedere se $T = \frac{1}{m} \bar{X}$ è deviato rispetto a p

$$\begin{aligned}\mathbb{E}(\bar{X}) &= mp \\ \rightarrow \frac{1}{m} \mathbb{E}(\bar{X}) &= p \\ \rightarrow \mathbb{E}\left(\frac{1}{m} \bar{X}\right) &= p\end{aligned}\tag{4}$$

Quindi $T = \frac{1}{m} \bar{X}$ non è deviato rispetto a p .

7.2.3 Esempio 3

Solito esempio, ma ora voglio stimare $\tau(\theta) = \lambda$. Inizio col dire che \bar{X} è uno stimatore non deviato di $\mathbb{E}(X) = \frac{1}{\lambda}$ (infatti $\mathbb{E}(X) = \mathbb{E}(\bar{X})$).

Dunque inizio dicendo che $\mathbb{E}(\bar{X}) = \frac{1}{\lambda}$ e che posso semplificarlo con $\lambda = \frac{1}{\mathbb{E}(\bar{X})}$. Tuttavia dopo questo mi devo fermare perché non posso più compiere trasformazioni lineari.

7.2.4 Considerazione 1

Se noi abbiamo uno stimatore non deviato per una certa quantità, calcolo la varianza di quello stimatore, per esempio quello della media campionaria, e osservo cosa succede

$$\begin{aligned} \text{Var}(\bar{X}) &= \text{Var}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) \\ &= \frac{1}{n^2} \text{Var}\left(\sum_{i=1}^n X_i\right) \\ &\rightarrow \text{sono v.a. indipendenti quindi } \text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}(X_i) \\ &= \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) \\ &= \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X) \\ &= \frac{1}{n^2} \sum_{i=1}^n \sigma^2 \\ &= \frac{\sigma^2}{n} \end{aligned} \tag{5}$$

7.2.5 Considerazione 2

Ci sono casi particolari in cui si può risolvere il problema della stima in modo esatto, per esempio con $\theta = \mathbb{N}$ e $X \sim U((\theta - \frac{1}{2}, \theta + \frac{1}{2}))$. A questo punto mi basta conoscere un valore della popolazione per capire il valore di θ ; infatti se $x_1 = 3.1$ allora $\theta = 3$ in quanto numero naturale più vicino a 3.1 (ricordo che avevamo definito $\theta = \mathbb{N}$). Dunque, in questo contesto, il mio stimatore, distorto in quanto operazione non lineare, risulta essere $\lfloor X_i \rfloor$.

Mean Square Error

7.3

Il *Mean Square Error (MSE)* o *scarto quadratico medio* è definito partendo da $T = t(X_1, \dots, X_n)$ come

$$MSE_{\tau(\theta)}(T) = \mathbb{E}((T - \tau(\theta))^2) \tag{6}$$

E più precisamente

$$\begin{aligned}
MSE_{\tau(\theta)}(T) &= \mathbb{E}((T - \tau(\theta))^2) \\
&= \mathbb{E}((T - \mathbb{E}(T) + \mathbb{E}(T) - \tau(\theta))^2) \\
&= \mathbb{E}\left(\underbrace{(T - \mathbb{E}(T))}_{\text{errore}} + \underbrace{(\mathbb{E}(T) - \tau(\theta))}_{\text{bias}}\right)^2 \\
&= \mathbb{E}\left((T - \mathbb{E}(T))^2 + 2(T - \mathbb{E}(T))(\mathbb{E}(T) - \tau(\theta)) + (\mathbb{E}(T) - \tau(\theta))^2\right) \\
&= \mathbb{E}\left((T - \mathbb{E}(T))^2\right) + \mathbb{E}(2(T - \mathbb{E}(T))(\mathbb{E}(T) - \tau(\theta))) + \mathbb{E}\left((\mathbb{E}(T) - \tau(\theta))^2\right) \\
&= Var(T) + \mathbb{E}(2(T - \mathbb{E}(T))(\mathbb{E}(T) - \tau(\theta))) + \mathbb{E}\left((\mathbb{E}(T) - \tau(\theta))^2\right) \\
&= Var(T) + \mathbb{E}(T - \mathbb{E}(T))\mathbb{E}(2(\mathbb{E}(T) - \tau(\theta))) + \mathbb{E}\left((\mathbb{E}(T) - \tau(\theta))^2\right) \\
&= Var(T) + (\mathbb{E}(T) - \mathbb{E}(\mathbb{E}(T)))\mathbb{E}(2(\mathbb{E}(T) - \tau(\theta))) + \mathbb{E}\left((\mathbb{E}(T) - \tau(\theta))^2\right) \\
&= Var(T) + 0 \cdot \mathbb{E}(2(\mathbb{E}(T) - \tau(\theta))) + \mathbb{E}\left((\mathbb{E}(T) - \tau(\theta))^2\right) \\
&= Var(T) + \mathbb{E}\left((\mathbb{E}(T) - \tau(\theta))^2\right) \\
&= Var(T) + (\mathbb{E}(T) - \tau(\theta))^2
\end{aligned} \tag{7}$$

Bias

7.4

Definisco il bias come

$$b_{\tau(\theta)}(T) = \mathbb{E}(T) - \tau(\theta) \tag{8}$$

E dunque posso riscrivere l'errore quadratico medio come

$$MSE_{\tau(\theta)}(T) = Var(T) + b_{\tau(\theta)}(T)^2 \tag{9}$$

Infine posso dire che se lo stimatore T è non deviato per $\tau(\theta)$, allora

$$b_{\tau(\theta)}(T) = 0 \tag{10}$$

e dunque

$$\mathbb{E}(T) = \tau(\theta) \tag{11}$$

e

$$MSE_{\tau(\theta)}(T) = Var(T) \tag{12}$$

Consistenza in media quadratica

7.5

Uno stimatore T si dice **consistente in media quadratica** per $\tau(\theta)$ quando

$$\lim_{n \rightarrow \infty} MSE_{\tau(\theta)}(T) = 0 \quad \forall \rightarrow T_n \quad MSE_{\tau(\theta)}(T) \xrightarrow{n \rightarrow \infty} 0 \tag{13}$$

Uno stimatore T si dice invece **debolmente consistente** rispetto a $\tau(\theta)$ se

$$\lim_{n \rightarrow \infty} P(\tau(\theta) - \varepsilon \leq T_n \leq \tau(\theta) + \varepsilon) = 1 \quad \forall \varepsilon > 0 \tag{14}$$

7.5.1 Considerazione

Posso facilmente dimostrare che se T è consistente in media quadratica per $\tau(\theta)$, allora T è debolmente consistente per $\tau(\theta)$

$$\begin{aligned}
 P(\tau(\theta) - \varepsilon \leq T_n \leq \tau(\theta) + \varepsilon) &= P(-\varepsilon \leq T_n - \tau(\theta) \leq \varepsilon) \\
 &= P(|T_n - \tau(\theta)| \leq \varepsilon) \\
 &= P((T_n - \tau(\theta))^2 \leq \varepsilon^2) \\
 &= 1 - P((T_n - \tau(\theta))^2 > \varepsilon^2) \\
 &\rightarrow \text{applico Markov } P(X > a) < \frac{\mathbb{E}(X)}{a} \\
 &\rightarrow 1 - P((T_n - \tau(\theta))^2 > \varepsilon^2) > 1 - \frac{\mathbb{E}((T_n - \tau(\theta))^2)}{\varepsilon^2} \\
 &= 1 - \frac{MSE_{\tau(\theta)}(T)}{\varepsilon^2} \xrightarrow{n \rightarrow \infty} 1
 \end{aligned} \tag{15}$$

Dunque ho dimostrato che avendo $a > b$ e $b \rightarrow 1$, allora per forza anche $a \rightarrow 1$. Se non ho capito il senso, ricordo che la probabilità massima è 1, quindi se a supera b e b tende ad 1, a comunque oltre 1 non va e tenderà anch'esso a 1.

7.5.2 Esempio 1

Ipotizzo una popolazione $X \sim B(p)$ e voglio conoscere il parametro $\theta = p$, dunque $\tau(\theta) = \tau(p) = p$. Siamo in possesso di un campione X_1, X_2, X_3 e lo vogliamo usare nello stimatore $t(X_1, X_2, X_3) = \frac{X_1 + X_2 + X_3}{5}$: il nostro stimatore è corretto per p ?

$$\begin{aligned}
 \mathbb{E}(T) &= \mathbb{E}\left(\frac{X_1 + X_2 + X_3}{5}\right) \\
 &= \frac{1}{5}\mathbb{E}(X_1 + X_2 + X_3) \\
 &= \frac{3}{5}p
 \end{aligned} \tag{16}$$

Ovviamente è distorto (è presente bias) perché dovrei avere una media ma faccio diviso 5. Qual è il suo bias?

$$\begin{aligned}
 b_p(T) &= \mathbb{E} - p \\
 &= \frac{3}{5}p - p \\
 &= -\frac{2}{5}p
 \end{aligned} \tag{17}$$

Qual è invece il suo MSE?

$$\begin{aligned}
 MSE_p(T) &= Var(T) + b_p(T)^2 \\
 &= Var\left(\frac{1}{5}(X_1 + X_2 + X_3)\right) + \frac{4}{25}p^2 \\
 &= \frac{3}{25}Var(X) + \frac{4}{25}p^2 \\
 &\rightarrow \text{ci troviamo in un contesto bernoulliano} \\
 &= \frac{3}{25}p(1-p) + \frac{4}{25}p^2
 \end{aligned} \tag{18}$$

7.5.3 Esempio 2

Quanto può saltare un atleta sapendo in media quanto salta e che minimo salta 0 e massimo θ ? Riassumendo, considero che $X \sim U((0, \theta))$ e $\tau(\theta) = \theta$ dunque provo a calcolare

$$\begin{aligned}\mathbb{E}(\bar{X}) &= \mathbb{E}(X) \\ &= \frac{\theta}{2}\end{aligned}\tag{19}$$

In questo caso la media campionaria non è un'idea eccelsa dato che mi trova $\frac{\theta}{2}$, dunque \bar{X} è deviato per θ , tuttavia

$$\begin{aligned}2\mathbb{E}(\bar{X}) &= 2\mathbb{E}(X) \\ &= 2\frac{\theta}{2} \\ &= \theta\end{aligned}\tag{20}$$

Dunque $T = 2\bar{X}$ risulta non deviato per θ . E' consistente in media quadratica per θ ?

$$\begin{aligned}MSE_p(T) &= Var(T) \\ &= Var(2\bar{X}) \\ &= 4Var(\bar{X}) \\ &= \frac{4}{n}Var(X) \\ &\rightarrow \text{in un contesto uniforme } Var(X) = \frac{(b-a)^2}{12} \\ &= \frac{4}{n} \frac{\theta^2}{12} \\ &= \frac{\theta^2}{3n} \xrightarrow{n \rightarrow \infty} 0\end{aligned}\tag{21}$$

Dunque T è consistente in media quadratica per θ . Ma se volessi considerare $T' = \max : 1 \leq i \leq n X_i$? Inizio con

$$\begin{aligned}F_{T'}(x) &= P(T' \leq x) \\ &= P(\max X - i \leq x) \\ &= P(\forall i X_i \leq x) \\ &= \prod_{i=1}^n P(X_i \leq x) \\ &= \prod_{i=1}^n F_X(x) \\ &= F_X(x)^n \\ &= \left(\sum_{i=1}^{\lfloor x \rfloor} P(X = i) \right)^n \\ &\rightarrow \text{in un contesto uniforme } P(X = i) = \frac{1}{n} \\ &= \left(\sum_{i=1}^{\lfloor x \rfloor} \frac{1}{n} \right)^n \\ &= \left(\frac{\lfloor x \rfloor}{n} \right)^n \\ &= \left(\frac{\lfloor x \rfloor}{\theta} \right)^n\end{aligned}\tag{22}$$

A questo punto calcolo

$$\begin{aligned}
\mathbb{E}(T') &= \int_0^\theta x f_{T'}(x) dx \\
&= \int_0^\theta x \frac{d}{dx} F_{T'}(x) dx \\
&= \int_0^\theta x \frac{d}{dx} \left(\frac{\lfloor x \rfloor}{\theta} \right)^n dx \\
&= \int_0^\theta xn \left(\frac{\lfloor x \rfloor}{\theta} \right)^{n-1} \frac{1}{\theta} dx \\
&= \int_0^\theta xn \frac{x^{n-1}}{\theta^n} dx \\
&= \frac{n}{\theta^n} \int_0^\theta x \cdot x^{n-1} dx \\
&= \frac{n}{\theta^n} \int_0^\theta x^n dx \\
&= \frac{n}{\theta^n} \frac{x^{n+1}}{n+1} \Big|_0^\theta \\
&= \frac{n}{\theta^n} \frac{\theta^{n+1}}{n+1} \\
&= \frac{n}{n+1} \theta
\end{aligned} \tag{23}$$

Vediamo quindi che $\mathbb{E}(T') = \frac{n}{n+1}\theta$ che è deviato per θ . Per risolvere si potrebbe considerare $T'' = \frac{n+1}{n}T'$ che risulta

$$\begin{aligned}
\mathbb{E}(T'') &= \frac{n+1}{n} \mathbb{E}(T') \\
&= \frac{n+1}{n} \frac{n}{n+1} \theta \\
&= \theta
\end{aligned} \tag{24}$$

E dunque non deviato per θ . Il suo MSE?

$$\begin{aligned}
MSE_p(T'') &= Var(T'') \\
&= Var\left(\frac{n+1}{n} \max_i X_i\right) \\
&= \left(\frac{n+1}{n}\right)^2 \cdot Var\left(\max_i X_i\right) \\
&= \left(\frac{n+1}{n}\right)^2 \cdot Var(T') \\
&= \left(\frac{n+1}{n}\right)^2 \cdot \left(\mathbb{E}(T'^2) - \mathbb{E}(T')^2\right) \\
&= \left(\frac{n+1}{n}\right)^2 \cdot \left(\left(\int_0^\theta x^2 f_{T'}(x) dx\right) - \mathbb{E}(T')^2\right)
\end{aligned} \tag{25}$$

Oppure potevo anche fare

$$\begin{aligned}
MSE_p(T'') &= Var(T'') \\
&= \mathbb{E}(T''^2) - \mathbb{E}(T'')^2 \\
&= \mathbb{E}(T''^2) - \theta^2 \\
&= \mathbb{E}\left(\left(\frac{n+1}{n}T'\right)^2\right) - \theta^2 \\
&= \left(\frac{n+1}{n}\right)^2 \mathbb{E}(T^2) - \theta^2 \\
&= \left(\frac{n+1}{n}\right)^2 \left(\int_0^\theta x^2 f_{T'}(x) dx\right) - \theta^2 \\
&= \left(\frac{n+1}{n}\right)^2 \left(\int_0^\theta x^2 n \frac{x^{n-1}}{\theta^n} dx\right) - \theta^2 \\
&= \left(\frac{n+1}{n}\right)^2 \left(\frac{n}{\theta^2} \int_0^\theta x^{n+1} dx\right) - \theta^2 \\
&= \left(\frac{n+1}{n}\right)^2 \left(\frac{n}{\theta^2} \frac{x^{n+2}}{n+2} \Big|_0^\theta\right) - \theta^2 \\
&= \left(\frac{n+1}{n}\right)^2 \left(\frac{n}{\theta^n} \frac{\theta^{n+2}}{n+2}\right) - \theta^2 \\
&= \left(\frac{n+1}{n}\right)^2 \left(\frac{n}{n+2} \theta^2\right) - \theta^2 \\
&= \frac{(n+1)^2}{n^2} \cdot \frac{n}{n+2} \theta^2 - \theta^2 \\
&= \frac{(n+1)^2}{n(n+2)} \theta^2 - \theta^2 \\
&= \theta^2 \left(\frac{(n+1)^2}{n(n+2)} - 1\right) \\
&= \theta^2 \left(\frac{n^2 + 2n + 1 - n^2 - 2n}{n(n+2)}\right) \\
&= \theta^2 \left(\frac{1}{n(n+2)} - 1\right) \\
&= \frac{\theta^2}{n(n+2)}
\end{aligned} \tag{26}$$

Dunque per riassumere:

- $MSE_\theta(T'') = MSE_\theta\left(\frac{n+1}{n} \max X_i\right) = \frac{\theta^2}{n(n+2)}$
- $MSE_\theta(T) = MSE_\theta(\bar{X}) = \frac{\theta^2}{3n}$
- $\mathbb{E}(T) = \mathbb{E}(T'') = \theta$

Per quali valori di n , $MSE_\theta(T'') \leq MSE_\theta(T)$?

$$\begin{aligned}
 \frac{\theta^2}{n(n+2)} &\leq \frac{\theta^2}{2n} \\
 \rightarrow \frac{1}{n(n+2)} &\leq \frac{1}{2n} \\
 \rightarrow n(n+2) &\geq 3n \\
 \rightarrow n &\geq 1
 \end{aligned} \tag{27}$$

Ma n è sempre ≥ 1 , dunque ho sempre almeno un campione e questo implica che $MSE_\theta(T'')$ è sempre più basso di $MSE_\theta(T)$.

7.5.4 Esempio 3

Ipotizzo $X \sim N(\mu, \sigma)$ e immagino di avere un campione X_1, X_2 . Mi piacerebbe conoscere σ , dunque $\theta = \sigma$, $\tau(\theta) = \tau(\sigma) = \sigma^2$ (stimare in maniera non distorta la deviazione standard non mi è facile dunque provo a stimare la varianza). Sono in possesso dello stimatore $T = X_1^2 - X_1X_2$, è buono?

$$\begin{aligned}
 \mathbb{E}(T) &= \mathbb{E}(X_1^2 - X_1X_2) \\
 &= \mathbb{E}(X_1^2) - \mathbb{E}(X_1X_2) \\
 &= \mathbb{E}(X^2) - \mathbb{E}(X)\mathbb{E}(X) \\
 &= \mathbb{E}(X^2) - \mathbb{E}(X)^2 \\
 &= \sigma^2
 \end{aligned} \tag{28}$$

7.5.5 Esempio 4

Sono in possesso di un campione X_1, \dots, X_n estratto da una popolazione $X \sim F$. So che $\mathbb{E}(X) = \mu$ e che $Var(F) = \sigma^2$, dunque $T_n = \frac{1}{3} \sum_{i=1}^n X_i$ è non deviato per μ ?

$$\begin{aligned}
 \mathbb{E}(T) &= \mathbb{E}\left(\frac{1}{3} \sum_{i=1}^n X_i\right) \\
 &= \frac{1}{3} \sum_{i=1}^n \mathbb{E}(X_i) \\
 &= \frac{n}{3} \mu \\
 &= \mu \iff n = 3
 \end{aligned} \tag{29}$$

Dunque osservo

$$\begin{aligned}
 MSE_\mu(T_n) &= Var(T_n) + b_\mu(T_n)^2 \\
 &= Var\left(\frac{1}{3} \sum_{i=1}^n X_i\right) + (\mu - \mathbb{E}(T_n))^2 \\
 &= \left(\frac{1}{9} \sum_{i=1}^n Var(X_i)\right) + \left(\mu - \frac{n}{3}\mu\right)^2 \\
 &= \frac{n}{9}\sigma^2 + \left(1 - \frac{n}{3}\right)^2 \mu^2
 \end{aligned} \tag{30}$$

7.5.6 Esempio 5

Considero un campione X_1, \dots, X_n di $X \sim N(\mu, \sigma)$. Inoltre $\theta = \mu$ e $\tau(\theta) = \mathbb{E}(X) = \mu$. Se prendo $T_1 = \bar{X}$ questa stima è non deviata. Se invece prendo $T_2 = X_3$?

$$\begin{aligned}\mathbb{E}(T_2) &= \mathbb{E}(X_3) \\ &= \mathbb{E}(X) \\ &= \mu\end{aligned}\tag{31}$$

Anch'esso non deviato, ma noto che i due MSE sono diversi. Questo l'MSE per T_1

$$\begin{aligned}MSE_\mu(T_1) &= Var(T_1) \\ &= \frac{\sigma^2}{n}\end{aligned}\tag{32}$$

Questo invece l'MSE per T_2

$$\begin{aligned}MSE_\mu(T_2) &= Var(T_2) \\ &= \sigma^2\end{aligned}\tag{33}$$

Proviamo ora invece con $T_3 = \sum_{i=1}^n \lambda_i X_i$ sapendo che $\forall i \lambda_i > 0$ e che $\sum_i \lambda_i = 1$. Esso è distorto?

$$\begin{aligned}\mathbb{E}(T_3) &= \sum_{i=1}^n \lambda_i \mathbb{E}(X_i) \\ &= \sum_{i=1}^n \lambda_i \mu \\ &= 1 \cdot \mu \\ &= \mu\end{aligned}\tag{34}$$

Non distorto, e il suo MSE?

$$\begin{aligned}MSE_\mu(T_3) &= Var(T_3) \\ &= Var\left(\sum_{i=1}^n \lambda_i X_i\right) \\ &= \sum_{i=1}^n Var(\lambda_i X_i) \\ &= \sum_{i=1}^n \lambda_i^2 Var(X_i) \\ &= \sigma^2 \sum_{i=1}^n \lambda_i^2\end{aligned}\tag{35}$$

7.5.7 Esempio 6

Prendo X_1, X_2 oltre che a $T_1 = \frac{1}{2}(X_1 + X_2)$, ossia la media, e $T_2 = \lambda_1 X_1 + (1 - \lambda_1) X_2$. Dato $\lambda_1^2 + (1 - \lambda_1)^2$, qual è il λ_1 che mi rende l'espressione il più piccolo possibile?

$$\begin{aligned}\min_{\text{val}} \text{ expression} &= \min_{\lambda_1} \lambda_1^2 + (1 - \lambda_1)^2 \\ &\rightarrow \text{derivo} \\ &= 2\lambda_1 - 2(1 - \lambda_1) \\ &= 2\lambda_1 - 2 + 2\lambda_1 \\ &= 4\lambda_1 - 2 \\ &\rightarrow \lambda_1 = \frac{1}{2}, \quad \lambda_2 = 1 - \lambda_1 = \frac{1}{2}\end{aligned}\tag{36}$$

7.5.8 Esempio 7

Il TCL ci dice che prese X_1, \dots, X_n i.i.d., allora $\sum_{i=1}^n X_i \sim N(n\mu, \sqrt{n}\sigma)$. Dunque

$$\begin{aligned}\bar{X} &= \frac{1}{n} \sum_{i=1}^n X_i \sim N(\mu_{\bar{X}}, \sigma_{\bar{X}}) \\ &= N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)\end{aligned}\tag{37}$$

E guarda caso $\mathbb{E}(\bar{X}) = \mu$ e $Var(\bar{X}) = \frac{\sigma^2}{n}$.

Legge dei Grandi numeri

7.6

7.6.1 Forma forte

All'aumentare di n , la media campionaria è equivalente al valore atteso della popolazione

$$P\left(\lim_{n \rightarrow \infty} \bar{X}_n = \mathbb{E}(X)\right) = 1\tag{38}$$

7.6.2 Forma debole

La probabilità che la media si discosti dal valore atteso, per al più di ε , tende a 0 all'aumentare di n

$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - \mathbb{E}(X)| > \varepsilon) = 0, \quad \forall \varepsilon > 0\tag{39}$$

7.6.3 Esempio 1

Ipotizzo un campione X_1, \dots, X_n estratto da una popolazione X e ipotizzo di voler stimare $\tau(\theta) = Var(X)$. Lo stimatore $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$ è non distorto per $Var(X)$, lo dimostro ricordandomi che

$$\begin{aligned}\sum_{i=1}^n (x_i - \bar{x})^2 &= \sum_{i=1}^n (x_i^2 - 2\bar{x}x_i + \bar{x}^2) \\ &= \sum_{i=1}^n x_i^2 - 2\bar{x} \sum_{i=1}^n x_i + n\bar{x}^2 \\ &= \sum_{i=1}^n x_i^2 - 2n\bar{x}^2 + n\bar{x}^2 \\ &= \sum_{i=1}^n x_i^2 - n\bar{x}^2\end{aligned}\tag{40}$$

Noto che

$$\begin{aligned}(n-1)s^2 &= \sum_{i=1}^n (x_i - \bar{x})^2 \\ &= \sum_{i=1}^n x_i^2 - n\bar{x}^2\end{aligned}\tag{41}$$

Quindi tramite metodo plug-in

$$\begin{aligned}
 \mathbb{E}((n-1)s^2) &= \mathbb{E}\left(\sum_{i=1}^n x_i^2 - n\bar{x}^2\right) \\
 &= \sum_{i=1}^n \mathbb{E}(X_i^2) - n\mathbb{E}(\bar{X}^2) \\
 &= n\left(\mathbb{E}(X^2) - \mathbb{E}(\bar{X}^2)\right) \\
 &= n\left(\text{Var}(X) + \mathbb{E}(X)^2 - \text{Var}(\bar{X}) - \mathbb{E}(\bar{X})^2\right) \\
 &= n\left(\sigma^2 + \mu^2 - \frac{\sigma^2}{n} - \mu^2\right) \\
 &= n\sigma^2 - \sigma^2 \\
 &= (n-1)\sigma^2
 \end{aligned} \tag{42}$$

Dunque

$$\mathbb{E}((n-1)s^2) \implies (n-1)\mathbb{E}(s^2) = (n-1)\sigma^2 \implies \mathbb{E}(s^2) = \sigma^2 \tag{43}$$

Risulta importante sottolineare che il metodo plug-in non mi dice se il risultato è deviato o non deviato, ma solo che è ragionevole.

7.6.4 Esempio 2

Ipotizzo di avere un campione X_1, \dots, X_n tale che $X \sim N(d, 2)$. So che $\theta = d$ e la mia statistica è \bar{X}_n . Inoltre sono a conoscenza del fatto che $\mathbb{E}(\bar{X}_n) = d$, $\text{Var}(\bar{X}_n) = \frac{4}{n}$ e $Z = N(0, 1)$. Inoltre $P(|\bar{X}_n - d| \leq 0.5) \geq 0.95$, ma per quale

n ? Oltre quale soglia?

$$\begin{aligned}P(|\bar{X}_n - d| \leq 0.5) &= P\left(\left|\frac{\bar{X}_n - d}{\frac{2}{\sqrt{n}}}\right| \leq \frac{0.5}{\frac{2}{\sqrt{n}}}\right) \\&= P\left(\left|\frac{\bar{X}_n - d}{\frac{2}{\sqrt{n}}}\right| \leq \frac{\sqrt{n}}{4}\right) \\&\rightarrow TCL \\&\approx P\left(|Z| \leq \frac{\sqrt{n}}{4}\right) \\&= \Phi\left(\frac{\sqrt{n}}{4}\right) - \Phi\left(-\frac{\sqrt{n}}{4}\right) \\&= \Phi\left(\frac{\sqrt{n}}{4}\right) - 1 + \Phi\left(\frac{\sqrt{n}}{4}\right) \\&= 2\Phi\left(\frac{\sqrt{n}}{4}\right) - 1 \\&\rightarrow 2\Phi\left(\frac{\sqrt{n}}{4}\right) - 1 \geq 0.95 \\&\rightarrow \Phi\left(\frac{\sqrt{n}}{4}\right) \geq \frac{1.95}{2} \\&\rightarrow \Phi\left(\frac{\sqrt{n}}{4}\right) \geq 0.975 \\&\rightarrow \Phi^{-1}\left(\Phi\left(\frac{\sqrt{n}}{4}\right)\right) \geq \Phi^{-1}(0.975) \\&\rightarrow \frac{\sqrt{n}}{4} \geq 1.96 \\&\rightarrow n \geq (4 \cdot 1.96)^2 \\&\rightarrow n \geq \approx 61.45 \\&\rightarrow n \geq 62 : P(|\bar{X}_n - d| \leq 0.5) \geq 0.95\end{aligned}\tag{44}$$

Più in generale, la stima dell'errore avviene tramite l'impiego del TCL nel seguente modo

$$\begin{aligned}
 P(|\bar{X}_n - \mu| \leq r) &\geq 1 - \delta \rightarrow P\left(\left|\frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}}\right| \leq \frac{r}{\frac{\sigma}{\sqrt{n}}}\right) \\
 &= P\left(\left|\frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}}\right| \leq \frac{r}{\sigma}\sqrt{n}\right) \\
 &\approx P\left(|Z| \leq \frac{r}{\sigma}\sqrt{n}\right) \\
 &= P\left(-\frac{r}{\sigma}\sqrt{n} \leq Z \leq \frac{r}{\sigma}\sqrt{n}\right) \\
 &= \Phi\left(\frac{r}{\sigma}\sqrt{n}\right) - \Phi\left(-\frac{r}{\sigma}\sqrt{n}\right) \\
 &= \Phi\left(\frac{r}{\sigma}\sqrt{n}\right) - 1 + \Phi\left(\frac{r}{\sigma}\sqrt{n}\right) \\
 &= 2\Phi\left(\frac{r}{\sigma}\sqrt{n}\right) - 1 \\
 &\rightarrow 2\Phi\left(\frac{r}{\sigma}\sqrt{n}\right) - 1 \geq 1 - \delta \\
 &\rightarrow \Phi\left(\frac{r}{\sigma}\sqrt{n}\right) \geq 1 - \frac{\delta}{2}
 \end{aligned} \tag{45}$$

Da qui posso trovare quello che succede con una probabilità di almeno $(1 - \delta)$

$$\delta \geq 2\left(1 - \Phi\left(\frac{r}{\sigma}\sqrt{n}\right)\right) \tag{46}$$

Oppure posso trovare

$$n \geq \left(\frac{\sigma}{r}\Phi^{-1}(1 - \delta)\right)^2 \tag{47}$$

In alternativa

$$r \geq \frac{\sigma}{\sqrt{n}}\Phi^{-1}\left(1 - \frac{\delta}{2}\right) \tag{48}$$

In realtà tutti questi conti possono essere svolti diversamente applicando la disuguaglianza di Chebyshev

$$P(|X - \mu| \geq r) \leq \frac{\sigma^2}{r^2} \tag{49}$$

e proseguendo così con

$$\begin{aligned}
 P(|X - \mu| \geq r) &\leq \frac{\sigma^2}{r^2} \rightarrow P(|X - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2} \\
 &\rightarrow P(|X - \mu| < \varepsilon) \implies 1 - P(|X - \mu| \leq \varepsilon) \geq 1 - \frac{\sigma^2}{\varepsilon^2} \\
 &\rightarrow \text{Var}(\bar{X}) = \frac{\text{Var}(X)}{n} \\
 &\rightarrow P(|\bar{X} - \mu| \leq \varepsilon) \geq 1 - \frac{\sigma^2}{n\varepsilon^2} \geq 1 - \delta \\
 &\rightarrow 1 - \frac{\sigma^2}{n\varepsilon^2} \geq 1 - \delta
 \end{aligned} \tag{50}$$

Dunque avrò:

- $\delta \geq \frac{\sigma^2}{n\varepsilon^2}$

- $n \geq \frac{\sigma^2}{\delta \varepsilon^2}$
- $\varepsilon \geq \sqrt{\frac{\sigma^2}{n \delta}}$

Questo è possibile siccome sono quantità piccole e posso non preoccuparmi della differenza tra $<$ e \leq , dunque

$$P(|\bar{X} - \mu| \leq \varepsilon) \geq 1 - \frac{\sigma^2}{n\varepsilon^2} \longleftrightarrow P(|\bar{X} - \mu| < \varepsilon) \geq 1 - \delta \quad (51)$$

Infine, per diletto, confronto TLC con Chebyshev e mi domando $\frac{\sigma^2}{\varepsilon^2} \Phi^{-1}(1 - \frac{\delta}{2})^2$ quando è $\leq \frac{\sigma^2}{\delta \varepsilon^2}$?

$$\frac{\sigma^2}{\varepsilon^2} \Phi^{-1}\left(1 - \frac{\delta}{2}\right)^2 \leq \frac{\sigma^2}{\delta \varepsilon^2} \rightarrow \Phi^{-1}\left(1 - \frac{\delta}{2}\right)^2 \leq \frac{1}{\delta} \quad (52)$$

Preferire il metodo del TCL è spesso una scelta più vantaggiosa quando si ha ragione di credere che la media campionaria sia approssimativamente distribuita in modo normale. Questo perché il TCL offre non solo un limite superiore sulla probabilità di discostamento della variabile aleatoria dalla media, ma fornisce anche una approssimazione più precisa della distribuzione della media campionaria. D'altra parte, il metodo con la disuguaglianza di Chebyshev, offre solo limiti superiori basati sulla deviazione standard, senza specificare la forma precisa della distribuzione. Di conseguenza, quando la normalità della media campionaria è plausibile, il TCL è preferibile per ottenere una descrizione più dettagliata e un'approssimazione più accurata della distribuzione della media campionaria.

Riepilogo stimatori

7.8

7.8.1 Distorsione

Un'importante caratteristica di uno stimatore è la sua non distorsione, che si verifica quando il suo valore atteso è uguale al parametro che si intende stimare. Per garantire l'assenza di distorsione, è fondamentale limitarsi a trasformazioni lineari. Ciò significa che sono consentite solo operazioni di addizione, sottrazione, moltiplicazione per uno scalare, derivazione e integrazione.

7.8.2 Stimatore media campionaria

Se mi ritrovo con n variabili aleatorie $X_1 \dots X_n$ i.i.d., allora posso usare la media campionaria come stimatore del valore atteso della popolazione, e valgono le seguenti proprietà:

- $\mathbb{E}(\bar{X}) = \frac{1}{n} \mathbb{E}(X_1 + \dots + X_n) = \frac{1}{n} (\mu + \dots + \mu) = \mu$
- $Var(\bar{X}) = \frac{1}{n^2} Var(X_1 + \dots + X_n) = \frac{1}{n^2} (\sigma^2 + \dots + \sigma^2) = \frac{\sigma^2}{n}$

7.8.3 Stimatore varianza e deviazione standard

Lo stimatore della varianza è ottenibile tramite il metodo plug-in visto precedentemente, mentre per la deviazione standard

$$s = \sqrt{s^2} \not\Rightarrow \mathbb{E}(s) = \sigma \quad (53)$$

Questo perché ho eseguito un'operazione non lineare (ho estratto una radice quadrata), quindi non posso dire nulla. Per ottenere uno stimatore non deviato della deviazione standard, esso dipende dal modello seguito dalla popolazione.

Capitolo 8

Processo di Poisson

Il processo di Poisson

8.1

Un processo di Poisson è un modello stocastico usato per descrivere l'occorrenza di eventi rari in un tempo continuo, dove gli eventi sono casuali e indipendenti, ma avvengono ad una frequenza costante. Esso è spesso usato per modellare fenomeni come per esempio il numero di chiamate che un call center riceverà in un certo lasso di tempo, oppure per stimare il traffico ad un server o la frequenza di attacchi informatici.

Per modellarlo inizio sapendo che, dato tempo $t = 0$ e una variabile aleatoria $N(t) =$ numero di eventi in $[0, t)$, al variare di t ottengo una famiglia di variabili aleatorie, ossia i miei **processi stocastici**. Sapendo che λ rappresenta il numero medio di eventi che capitano per unità di tempo, definisco le seguenti regole:

(1) $N(0) = 0 \rightarrow$ il processo inizia a tempo 0

(2) Indipendenza per intervalli disgiunti \rightarrow gli eventi in t sono indipendenti dagli eventi in $t + s$

(3)

$$\lim_{h \rightarrow 0} \frac{P(N(h) = 1)}{h} = \lambda \rightarrow$$

la probabilità di un evento in un intervallo di tempo è proporzionale alla lunghezza dell'intervallo, infatti, dato h piccolo avrò $P(N(h) = 1) \approx \lambda h$

(4)

$$\lim_{h \rightarrow 0} \frac{P(N(h) \geq 2)}{h} = 0 \rightarrow$$

la probabilità che accadano 2 o più eventi in un piccolo intervallo di tempo è trascurabile, infatti, dato h piccolo avrò $P(N(h) \geq 2) \approx 0$

Definisco infine $N(t) \sim P(\lambda t)$, dove $N(t) = k \in \mathbb{N}$, $n \in \mathbb{N}$ e $n > k$. Inoltre assicuro che ci sia un incremento stazionario degli eventi da 0 a t . Dunque posso dichiarare i seguenti eventi mutualmente esclusivi:

- **Evento A:** l'evento che si verifica quando ogni avvenimento cade in un intervallo diverso. Dunque $N(t) = k$ e massimo un elemento ogni sotto-intervallo.
- **Evento B:** l'evento che si verifica tutte le altre volte (tutti gli altri modi per cui $N(t) = k$). Dunque $N(t) = k$ e almeno un intervallo ha 2 o più eventi.

Dunque $\{N(t) = k\} = A \cup B$, $A \cap B = \emptyset$ e quindi

$$P(N(t) = k) = P(A) + P(B) \tag{1}$$

Dove $P(A)$ sono k intervalli a partire dai miei n ($\binom{n}{k}$) e immagino di aver preso i primi k sotto-intervalli: sempre per il fatto che il numero di occorrenze all'interno di due intervalli disgiunti è indipendente, io posso calcolare la probabilità che abbia una e una sola occorrenza nell'ennesimo intervallo, k volte. Questo si traduce in

$P(\text{occorrenza nel primo intervallo}) = \lambda \cdot \text{grandezza intervallo} = \lambda \frac{t}{n}$, che per k intervalli diventa $(\lambda \frac{t}{n})^k$. Per quanto riguarda $P(B)$ invece, più n è grande e più i singoli intervalli sono piccoli; per cui da un certo n in avanti la probabilità di avere il numero di avvenimenti entro un particolare intervallo è zero. Ma siccome sto moltiplicando (per 2) è intersezione indipendente (quindi l'evento è risultato dalla produttoria degli eventi), allora $P(B)$ è zero. Allora la probabilità di avere k avvenimenti del mio evento, che non avvengono ognuno in un intervallo diverso, è zero. Infine devo dire che so che posso avere 0 occorrenze, 1 occorrenza o 2 o più occorrenze, tuttavia l'ultimo caso ho visto che ha probabilità zero ($P(B) = 0$). Dunque gli unici casi possibili sono 1 occorrenza o nessuna. "Nessuna occorrenza" si traduce semplicemente in $1 - (1 \text{ occorrenza})$, ossia $(1 - \lambda \frac{t}{n})^{n-k}$.

Quanto detto può sembrare confusionario ma è possibile riassumerlo in punti per chiarire:

- **Calcolo di $P(A)$**

- Si considerano k intervalli da un totale di n . Questo è rappresentato dalla combinazione binomiale $\binom{n}{k}$, che rappresenta il numero di modi in cui è possibile selezionare k intervalli da un totale di n .
- Per ogni intervallo, si calcola la probabilità che si verifichi esattamente un evento. Dato che gli eventi sono indipendenti tra intervalli disgiunti, si può calcolare la probabilità di una singola occorrenza nel primo intervallo come λ (il tasso medio di arrivo degli eventi) moltiplicato per la dimensione dell'intervallo, che è $\frac{t}{n}$.
- Questo si traduce in $P(\text{occorrenza nel primo intervallo}) = \lambda \cdot \text{grandezza intervallo} = \lambda \frac{t}{n}$, che per k intervalli diventa $(\lambda \frac{t}{n})^k$

- **Calcolo di $P(B)$**

- Quando si tratta di $P(B)$, si osserva che all'aumentare di n , gli intervalli diventano più piccoli. Infatti considerando un numero crescente di intervalli (n crescente), si distribuisce la probabilità di eventi in modo più uniforme lungo l'asse temporale, riducendo la probabilità di avere due o più eventi simultaneamente in uno stesso sotto-intervallo.
- Quando gli intervalli diventano sufficientemente piccoli, la probabilità di avere due o più eventi all'interno di un singolo intervallo diventa trascurabile, e quindi $P(B)$ tende a 0.
- Poiché la probabilità di $P(B)$ è trascurabile all'aumentare del numero di intervalli n , allora $P(B) = 0$ e l'intersezione $A \cap B$ è essenzialmente nulla.
- Dunque gli unici casi possibili sono 1 occorrenza o nessuna, che si traduce in $1 - (1 \text{ occorrenza})$, ossia $(1 - \lambda \frac{t}{n})^{n-k}$.

Unendo il tutto ottengo una combinazione tra le occorrenze singole e l'assenza di occorrenze, ossia

$$P(N(t) = k) = \binom{n}{k} \cdot \left(\lambda \frac{t}{n}\right)^k \cdot \left(1 - \lambda \frac{t}{n}\right)^{n-k} \quad (2)$$

Quindi $P(N(t) = k) = \binom{n}{k} \cdot (\lambda \frac{t}{n})^k \cdot (1 - \lambda \frac{t}{n})^{n-k}$ dato $\frac{\lambda t}{n} = p \implies np = \lambda t$ ricorda $B(n, p)$ e per essere esaustivi, siccome $N(t) \sim P(\lambda t)$

$$P(N(t) = k) = e^{-\lambda t} \frac{(\lambda t)^k}{k!} I_{\mathbb{N} \cup \{0\}}(k) \quad (3)$$

Dove le specificazioni sono numeri interi non negativi.

Ora ipotizzo di sapere che l'evento X_1 occorre in $[0, t_1)$, l'evento X_2 occorre in $[t_1, t_2)$ e così via: io voglio sapere quanto tempo deve passare tra il tempo in cui occorre un evento e il tempo in cui occorre il successivo

$$\begin{aligned} P(X_1 > t) &= P(\text{non si è verificato } X_1 \text{ entro il tempo } t) \\ &= P(N(t) = 0) \\ &= e^{-\lambda t} \frac{(\lambda t)^0}{0!} \\ &= e^{-\lambda t} \end{aligned} \quad (4)$$

E dunque

$$\begin{aligned}F_{X_1}(t) &= P(X_1 \leq t) \\ &= 1 - P(X_1 > t) \\ &= 1 - e^{-\lambda t} \\ &\implies X_1 \sim E(\lambda)\end{aligned}\tag{5}$$

Dove nella prima equazione voglio sapere qual è la probabilità che il primo evento accada oltre un certo tempo t , mentre nella seconda il suo duale. Si noti che la seconda equazione è la funzione di ripartizione di una distribuzione esponenziale, infatti l'intertempo degli eventi segue una distribuzione esponenziale.

Invece per X_2 cosa posso dire?

$$P(X_2 > t \mid X_1 = s) = P(\text{nessun evento in } [s, s + t] \mid X_1 = s)\tag{6}$$

Ma a cosa serve condizionare rispetto ad un evento indipendente? A nulla, infatti

$$P(\text{nessun evento in } [s, s + t] \mid X_1 = s) = P(N(t) = 0) = e^{-\lambda t}\tag{7}$$

Questo implica che $X_2 \sim E(\lambda)$ e quindi $\forall i X_i \sim E(\lambda)$, dunque $\forall i X_i$ i.i.d. e quindi $X_2 \perp X_1$.

Dunque possiamo ricapitolare dicendo che

$$\begin{aligned}P(X_2 > t \mid X_1 = s) &= P(X_2 > t) \\ &= 1 - F_{X_2}(t) \\ &= 1 - (1 - e^{-\lambda t}) \\ &= e^{-\lambda t}\end{aligned}\tag{8}$$

8.1.1 Esempio

Supponiamo di avere una sorgente radioattiva che emette particelle alfa. Il numero medio di particelle alfa emesse da questa sorgente in 1 minuto è di $\lambda = 0.5$. Vogliamo calcolare la probabilità di non osservare alcuna particella alfa nei prossimi 5 minuti. Poiché il numero di particelle alfa emesse in un dato intervallo di tempo segue una distribuzione di Poisson, possiamo utilizzare un modello esponenziale per descrivere il tempo tra due emissioni successive.

Dato che λ rappresenta il tasso di emissione medio su 1 minuto, possiamo calcolare il parametro della distribuzione esponenziale come $\frac{1}{\lambda} = \frac{1}{0.5} = 2$. Infine, dato $X \sim E(\lambda)$, che su Python sarebbe `st.expon(scale=1/lambda)`, possiamo considerare $F_X(x)$ che, sempre su Python, equivale a `expon.cdf(5)`.

Parte IV
Appendice

Capitolo 9

Cheatsheet Matematica

Qui di seguito alcune regole matematiche utili.

Dai dati alla distribuzione

9.1

Dati	Modello
Esito binario, 1 esperimento	Bernoulli
Esito binario, n esperimenti, conteggio dei successi	Binomiale
Spazio equiprobabile discreto	Uniforme Discreta
Esito binario, n esperimenti, conteggio degli insuccessi	Geometrico
Binomiale, $\mathbb{E}(X) = \text{Var}(X)$	Poisson
Estrazione di oggetti binari senza reimmissione	Ipergeometrico
Spazio equiprobabile e continuo	Uniforme Continuo
Durata di vita di un fenomeno, $\sigma_X = \mathbb{E}(X)$	Esponenziale
Variabile casuale continua, media = mediana	Gaussiano (aka Normale)
Gaussiano, $\mu = 0, \sigma = 1$	Normale Standard
Conteggio di eventi con frequenza media variabile nel tempo	Processo di Poisson

Derivata	Risultato
$[k]'$	0
$[kf(x)]'$	$kf'(x)$
$[x^n]'$	nx^{n-1}
$[e^x]'$	e^x
$[f(x) \pm g(x)]'$	$f'(x) \pm g'(x)$
$[f(x) \cdot g(x)]'$	$f'(x) \cdot g(x) + f(x) \cdot g'(x)$
$\left[\frac{f(x)}{g(x)}\right]'$	$\frac{f'(x) \cdot g(x) - f(x) \cdot g'(x)}{g(x)^2}$

Integrale	Risultato
$\int k dx$	$kx + c$
$\int kf(x) dx$	$k \int f(x) dx$
$\int x^n dx$	$\frac{x^{n+1}}{n+1} + c$
$\int e^x dx$	$e^x + c$
$\int [f(x) \pm g(x)] dx$	$\int f(x) dx \pm \int g(x) dx$
Integrazione per sostituzione	
$\int f(g(x)) \cdot g'(x) dx$	$\int f(u) du$, dove $u = g(x)$
Integrazione per parti	
$\int f'(x) \cdot g(x) dx$	$f(x) \cdot g(x) - \int f(x) \cdot g'(x) dx$

Capitolo 10

Cheatsheet Python

Serie di appunti Python presi a lezione o durante la correzione di esami passati.

```
1 %matplotlib inline
2 import itertools
3 import graphviz
4 import matplotlib.pyplot as plt
5 import numpy as np
6 import pandas as pd
7 import scipy.stats as st
8 import statsmodels.api as sm
9
10 from sklearn import tree
11 from sklearn.preprocessing import LabelEncoder
12
13 plt.style.use('fivethirtyeight')
14 plt.rc('figure', figsize=(10.0, 7.0))
```

Alberi di decisione

10.1

```
1 # Eseguo tutto in ordine
2
3 # Preparo i dati
4 good_guys = heroes.loc[['Superman', 'Flash', 'Batman', 'Wolverine']]
5 bad_guys = heroes.loc[['Kingpin', 'Galactus', 'Joker', 'Penguin', 'Magneto']]
6 all_guys = pd.concat([good_guys, bad_guys])
7 features = ['Hair color', 'Intelligence']
8 X = all_guys[features]
9 Y = pd.concat([
10     pd.DataFrame(['good guy'] * len(good_guys), index=good_guys.index),
11     pd.DataFrame(['bad guy'] * len(bad_guys), index=bad_guys.index)
12 ])
13 # Y[X['Intelligence'] <= 40]
14
15 # Label Encoder
16 hair_col_encoder = LabelEncoder()
17 hair_col_encoder.fit(all_guys['Hair color'])
```

```

18 all_guys['Hair color'] = hair_col_encoder.transform(all_guys['Hair color'])
19 all_guys['Intelligence'] = intelligence_encoder.transform(all_guys['Intelligence'])
20 X = all_guys[features]
21
22 # Classificatore binario
23 clf = tree.DecisionTreeClassifier()
24 #clf = tree.DecisionTreeClassifier(criterion='entropy') # di norma usa 'gini'
25 clf = clf.fit(X, Y)
26 predictions = clf.predict([X.loc[name] for name in X.index])
27
28 # Per il classificatore binario posso anche fare
29 model = GaussianNB()
30 model.fit(X,Y)
31 model.predict(X)
32
33 # Mostrare l'albero graficamente
34 graphviz.Source(
35     tree.export_graphviz(
36         clf,
37         out_file=None,
38         class_names=['bad guy', 'good guy'],
39         feature_names=features
40     )
41 )

```

Calcolo combinatorio

10.2

```

1 # Disposizioni con ripetizione (3 oggetti, 2 posti)
2 list(itertools.product(range(3), repeat=2))
3
4 # Disposizioni senza ripetizione (3 oggetti, 2 posti)
5 list(itertools.permutations(range(3), repeat=2))
6
7 # Permutazioni (3 oggetti)
8 list(itertools.permutations(range(3), repeat=3))
9
10 # Combinazioni (4 oggetti, 2 posti)
11 list(itertools.combinations(range(4), repeat=2))

```

Errore nella stima

10.3

```

1 # Quanto siamo disposti a sbagliare nella stima con una probabilità del 95 per cento? Qual è
  ↪ il margine di errore che dobbiamo tollerare?
2 Z = st.norm()
3 eps = 0.95
4 n = len(vals)

```

```

5 sigma = vals.std()
6 (
7     (Z.ppf((1 + eps) / 2) * (n ** 0.5)) / sigma,
8     (Z.ppf((1 + eps) / 2) * sigma) / (n ** 0.5)
9 )

```

Frequenze

10.4

```

1 # Frequenze congiunte relative
2 pd.crosstab(
3     index=df['col1'],
4     columns=df['col2'],
5     margins=True,           # congiunte (default=False)
6     normalize=True,        # relative (default=False)
7     colnames='',
8     # dropna=False         se serve
9 )
10
11 # Frequenze assolute
12 pd.crosstab(
13     index=df['col1'],
14     columns=['Freq. Assolute'],
15     colnames='')
16 )
17
18 # Frequenze cumulate
19 (
20     pd.crosstab(
21         index=df['col1'],
22         columns=['Freq. Assolute'],
23         normalize=True,
24         colnames='')
25     )
26 ).cumsum()
27
28 # Frequenze per raggruppamento
29 pd.crosstab(
30     index=pd.cut(
31         df['col1'],
32         bins=[30,50,80,100,200,500,1000],
33         right=False),
34     columns=[df['col2']]
35 )

```

```

1 # Barplot con subplot
2 # Permette di rappresentare meglio i dati nominali o categorici, calcolare la frequenza o
  ↳ fare df.groupby(['col1', 'col2']).sum()
3 plt.subplot(1,2,1)
4 df['val1'].value_counts().plot.bar()
5 plt.subplot(1,2,2)
6 df['val2'].plot.bar()
7 plt.show()
8
9 # Barplot da frequenze
10 freqs.plot.bar(color=['pink', 'blue'], stacked=False)
11 plt.show()
12
13 # Boxplot
14 # Mostra com'è distribuita la frequenza: la linea centrale è la mediana, i lati orizzontali
  ↳ sono primo e terzo quartile mentre i pallini, se presenti, sono gli outlier.
15 plt.boxplot(
16     df['val1'],
17     df['val2'],
18     labels=['val1', 'val2'],
19     whisk=1.5
20 )
21 plt.show()
22
23 # Due grafici nello stesso grafico
24 data = [df['val1'], df['val2']]
25 fig,ax = plt.subplots()
26 ax.boxplot(data)
27 ax.set_xticklabels(['Label1', 'Label2'])
28 plt.show()
29
30 # ECDF
31 # Evidenzia valori mancanti, distribuzione unimodale e/o simmetrica, distribuzione
  ↳ asimetrica a sinistra (mediana < media) o distribuzione asimmetrica a destra (mediana >
  ↳ media)
32 ecdf = sm.distributions.ECDF(df['col'])
33 x = np.arange(1980, 1991)
34 y = ecdf(x)
35 plt.step(x,y)
36 plt.show()
37
38 # Grafico di dispersione
39 # Determina la relazione tra più attributi del dataframe
40 plt.scatter(df['val1'], df['val2'])
41 plt.show()
42
43 # Istogramma

```

```

44 # Suddivide la popolazione in classi e ne rappresenta le frequenze: mette in risalto
    ↳ simmetria, dispersione, intervalli di concentrazione, vuoti nei dati e distacco nei dati
45 df['val'].hist()
46 #df['val'].hist(bins=10)
47 #df['val'].hist(bins=[0, 30, 50, 70, 100])
48 #df['val'].hist(bins=np.arange(0,1001,100))
49 plt.show()
50
51 # QQplot
52 # Stima se la popolazione segue una normale (o qualsiasi distribuzione scelta)
53 sm.qqplot(df['val'].dropna(), fit=True, line='45')
54 #sm.qqplot(df['val'].dropna(), dist=st.expon(), line='45')
55 #sm.qqplot(df['val'].dropna(), dist=st.distributions.poisson(lambda), line='45')
56 plt.show()

```

Indici

10.6

```

1 # Descrizione generale
2 df.describe()
3
4 # Indici di centralità
5 (vals.mean(), vals.median())
6
7 # Dato più presente
8 vals.mode()
9
10 # Indici di dispersione
11 vals.var()
12 vals.quantile(0.75) - vals.quantile(0.25)
13
14 # Indice di correlazione
15 vals1.corr(vals2)
16
17 # Indice di Gini (il dato 'col' è rappresentato in maniera uniforme?)
18 def gini(series):
19     return 1 - sum(series.value_counts(normalize=True).map(lambda f: f**2))
20
21 def norm_gini(series):
22     s = len(series.unique())
23     return s * gini(series) / (s-1)
24
25 norm_gini(df['col'])
26
27 # Entropia
28 def entropy(series, normalized=False):
29     freq = series.value_counts(normalize=True)
30     e = sum((freq.map(lambda f: -f * np.log2(f))))
31     if normalized:
32         e /= np.log2(len(freq))

```

```

33     return e
34
35 entropy(df['col'], normalized=True)

```

Miscellanea

10.7

```

1  # Aggiungere una colonna
2  # Aggiungo una colonna 'new_val' che vale 1 se 'val > 50', 0 altrimenti
3  df['new_val'] = [1 if (val > 50) else 0 for i in val]
4
5  # Applicare una trasformazione su una serie di dati
6  # Ogni valore * 2 ed elevato al quadrato
7  serie.apply(lambda x: (x*2)**2)
8
9  # Classificatori
10 # Dato s come terzo quantile per esempio
11 VP = len([x for x in df[col1==1]['col2'] if x >= s])
12 FP = len([x for x in df[col1==0]['col2'] if x >= s])
13 FN = len([x for x in df[col1==1]['col2'] if x < s])
14 VN = len([x for x in df[col1==0]['col2'] if x < s])
15
16 # Genero un campione casuale di 32 elementi
17 X = st.norm(mean, std)
18 simulati = X.rvs(taglia)
19
20 # Leggere un dataset
21 df = pd.read_csv(
22     'file.csv',
23     delimiter=';',
24     decimal=',',
25     index_col='col_name',
26     parse_dates=True
27 )
28
29 # Minimo e massimo
30 (df['val'].min(), df['val'].max())
31
32 # Raggruppamento
33 # Ipotizzo un dataframe con studenti, materie e voti: come calcolo la media dei voti delle
34   ↳ materie per ciascuno studente?
35 grouped_df = df.groupby(['Studente', 'Materia'])
36 medie = grouped_df['Voto'].mean()
37 # Ipotizzo delle biblioteche, ad ognuna ogni anno si iscrive gente: quanti iscritti totali
38   ↳ ogni anno?
39 iscritti = df[['Anno', 'Iscritti']].groupby(df['Anno']).sum()
40
41 # Rimuovere valori mancanti
42 # Utile per quando vengono richieste frequenze o si esegue uno scatter plot
43 # Rimuove tutta la riga se in quella riga c'è una colonna mancante

```

```

42 df.dropna()
43
44 # Selezione
45 # Selezionare i valori della colonna 'val2' che sono nel range [40, 60]
46 df[(df['val2'] >= 40) & (df['val2'] <= 60)]
47
48 # Somma dei valori in una colonna
49 df['val'].sum()
50
51 # Sorteggio
52 # Qual è la probabilità che sorteggiando un valore dal dataset, quel valore sia di un dato
  ↪ tipo?
53 # (dopo aver eseguito prendo il valore * 100 per ottenere la percentuale)
54 df['val'].value_counts(normalize=True)
55
56 # Stima
57 # Quando voglio "stimare" uso la media
58 len(esito) / len(totale)
59
60 # Taglia del campione
61 len(vals)
62
63 # Verificare l'assenza di valori mancanti
64 pd.isnull(df['val']).any()
65 # Oppure
66 df['val'].isnull().any()
67
68 # Seleziona i valori unici nella colonna
69 df['val'].unique()

```

Ordinamento

10.8

```

1 # Ordinare una lista in base alla lunghezza degli elementi, in ordine decrescente
2 sorted(list, key=len, reverse=True)
3
4 # Ordinare una lista in base ad un attributo 'attributo'
5 sorted(tuple_list, key=lambda x: getattr(x, 'attributo'))
6
7 # Ordinare una lista in base alla lunghezza usando .sort()
8 list.sort(key=lambda n: len(n))
9
10 # Ordinare un dataframe in base ai valori di due colonne, al contrario e basandosi sulla
  ↪ stringa in minuscolo
11 df.sort_values(by=['col1', 'col2'], ascending=False, key=lambda s: s.str.lower())
12
13 # Ordinare una colonna
14 col.sort_values()
15 # Oppure
16 np.sort(col)

```