

31 Variabili aleatorie binomiali e bimodali

In questa Lezione tratteremo le variabili aleatorie binomiali e quelle bimodali, due questioni molto diverse, da mettere comunque a confronto.

31.1 Le variabili aleatorie binomiali

Si dice che una variabile aleatoria X è binomiale di parametri n e p (e dev'essere n intero ≥ 1 e $0 \leq p \leq 1$) e si scrive $X \sim B(n, p)$, se ha *densità binomiale*, cioè

$$p_k = P(X = k) := \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \dots, n. \quad (101)$$

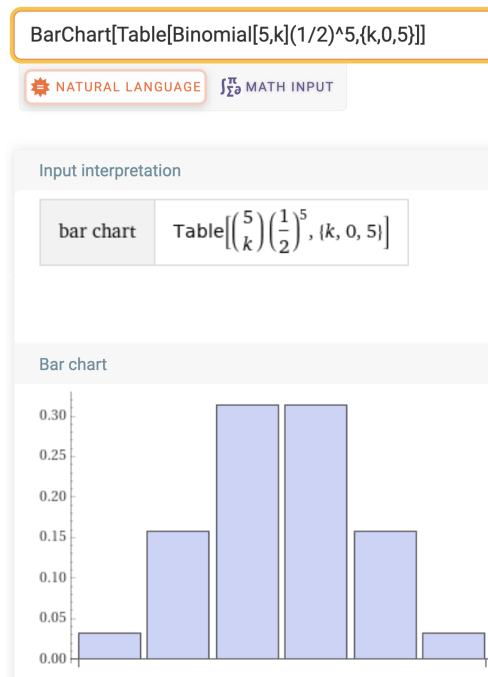
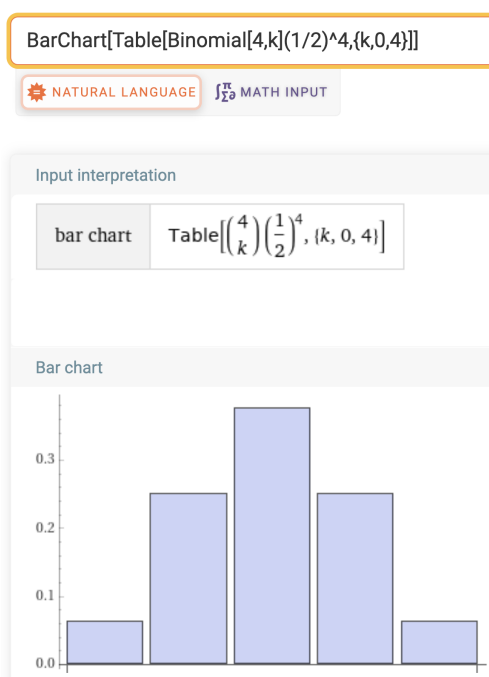
La densità a binomiale rappresenta, al variare di k intero da 0 a n , la probabilità di ottenere k teste in n lanci di una moneta che ha $P(\text{testa}) = p$. Più in generale:

la densità a binomiale dà, al variare di k intero da 0 a n , la probabilità di un certo numero k di successi in uno schema successo-insuccesso con n prove indipendenti avendo il successo probabilità p ad ogni prova.

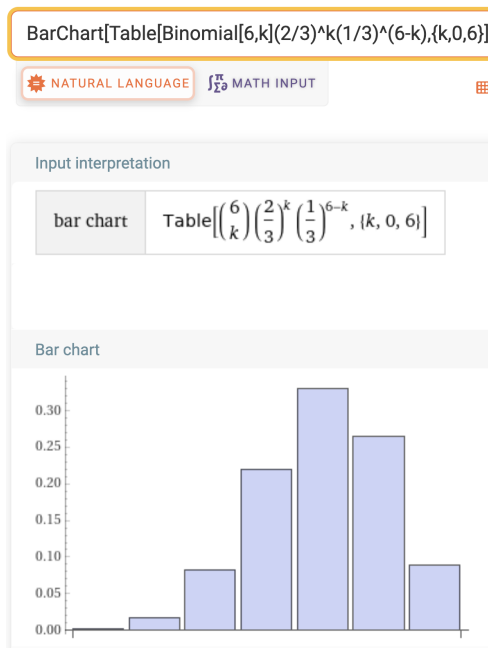
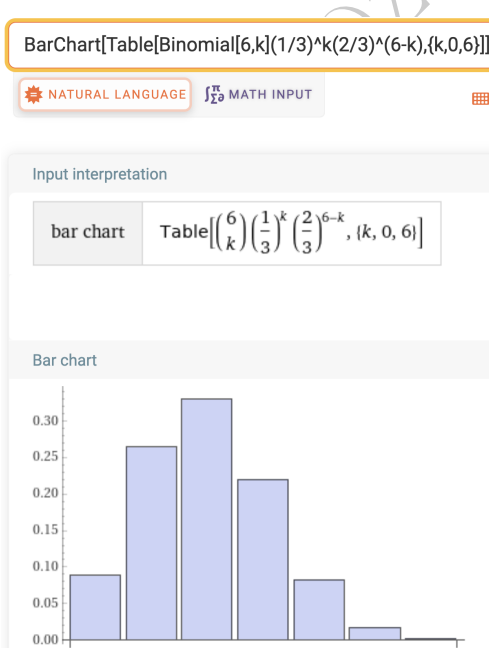
Gli istogrammi a barre rappresentativi della densità binomiale tendono ad avere forme più o meno a campana.

La forma a campana è tanto più riconoscibile quanto più n è grande e p è lontano da 0 e da 1. E la densità è tanto più "decentrata" rispetto a $\frac{n}{2}$ quanto più p è vicino a 0 o a 1; nei 2 casi il punto di massimo si avvicina a 0 e n rispettivamente.

Ecco 2 istogrammi a barre tratti da WolframAlpha, con $n = 4$ e $n = 5$ ed entrambi con $p = \frac{1}{2}$, e quest'ultimo valore causa la simmetria delle campane.



Ecco 2 istogrammi a barre tratti da WolframAlpha, con $n = 6$, e con rispettivamente $p = \frac{1}{3}$ e $p = \frac{2}{3}$, campane asimmetriche, con skewness non nulla.



L'allontanarsi di p dal valore centrale $\frac{1}{2}$ causerà inizialmente asimmetria nella forma a campana (si faccia il bar chart relativo a

$B\left(3, \frac{1}{4}\right)$) ma al crescere di n l'asimmetria tenderà ad attenuarsi sempre più, rimanendo invece la centratura spostata rispetto alla metà $\frac{n}{2}$ di n : la campana tenderà a centrarsi intorno a np .

Esempio 1. Con $n = 2$ prove si possono ottenere 0 o 1 o 2 teste (che adesso consideriamo successi) e se la moneta è equilibrata è

$$P(k = 0) = \binom{2}{0} \left(\frac{1}{2}\right)^0 \left(1 - \frac{1}{2}\right)^{2-0} = \frac{1}{4} \leftarrow P(0 \text{ teste})$$

$$P(k = 1) = \binom{2}{1} \left(\frac{1}{2}\right)^1 \left(1 - \frac{1}{2}\right)^{2-1} = \frac{1}{2} \leftarrow P(1 \text{ teste})$$

$$P(k = 2) = \binom{2}{2} \left(\frac{1}{2}\right)^2 \left(1 - \frac{1}{2}\right)^{2-2} = \frac{1}{4} \leftarrow P(2 \text{ teste})$$

Esempio 2. Qual è la probabilità di ottenere [esattamente] 3 volte il numero 1 su 7 lanci di un dado [regolare]?

Il successo ad ognuna delle $n = 7$ prove ha probabilità $\frac{1}{6}$ e allora, con la densità $B\left(7, \frac{1}{6}\right)$, la probabilità cercata è:

$$P(k = 3) = \binom{7}{3} \left(\frac{1}{6}\right)^3 \left(1 - \frac{1}{6}\right)^{7-3} = \frac{7!}{3!(7-3)!} \left(\frac{1}{6}\right)^3 \left(\frac{5}{6}\right)^4 =$$

$$= \frac{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdot 7 \cdot 5^4}{1 \cdot 2 \cdot 3 \cdot 1 \cdot 2 \cdot 3 \cdot 4 \cdot 6^7} = 5 \cdot 7 \cdot \frac{625}{279\,936} = \frac{21\,875}{279\,936} \approx 0.0781 = 7.81\%.$$

(Diciamo pure, a parole, circa 8%).

31.2 La binomiale come spiegazione delle campane

Ecco il grafico della densità dell'Esempio 1 visto, che è un bar chart, qua disposto orizzontalmente:

0 XXXX 0.25 qua c'è 1 solo caso su 4: CC cioè croce, croce
 1 XXXXXXXX 0.5 qua ci sono 2 casi su 4: TC e CT
 2 XXXX 0.25 qua c'è 1 solo caso su 4: TT cioè testa, testa

Si noti che nel soprastante bar chart si vede la distribuzione a campana molto più chiaramente che nel caso di un solo lancio

0 XXXXXX 0.5 (croce)

1 XXXXXX 0.5 (testa)

e si comincia a capire la natura della quasi onnipresente forma più o meno a campana: essa si forma in fenomeni composti da tante “cose che si sommano”, parlando in modo (necessariamente) impreciso. Similmente le altezze degli individui di una popolazione, o qualunque altro parametro biometrico, biomedico o fisiologico, tenderanno ad avere una distribuzione con forma più o meno a campana, perché ognuno di quei parametri è risultante da molti altri: per esempio in un individuo altissimo potranno aver concorso nella stessa direzione (“successi”) la genetica, l’alimentazione, le cause fortuite imponderabili, e tale concordanza sarà rara mentre molti più individui avranno un’altezza media, e di nuovo pochi saranno bassissimi. La forma a campana diventa ancora più evidente con 3 lanci di moneta, ovvero $B(3, \frac{1}{2})$, e sempre più al crescere di n . Si faccia il bar chart di $B(3, \frac{1}{2})$ e $B(4, \frac{1}{2})$.

Prova tu stesso. Leggiamo sul web⁽¹⁷⁰⁾ “Il termometro segna già -41° , ma da queste parti è normale. Siamo nella nella Jacuzzi, nel nord-est della Siberia, chiamata anche la regione del “Polo del Freddo Boreale”. Qui d’inverno la temperatura non supera mai gli zero gradi, e rimane quasi sempre al di sotto dei -30°C , raggiungendo talvolta i -60°C .” Disegna un ipotetico grafico della densità della funzione temperatura in quel luogo, segnando anche sull’asse delle ascisse il dato rilevato. Osserva sul grafico il significato delle parole usate “mai”, “quasi sempre” e “talvolta”: corrispondono a 3 aree di sottografico, se la temperatura viene considerato un parametro “continuo” – cioè, in pratica, che può assumere valori con infiniti decimali – e a somme di probabilità delle aste di un bar chart, se le temperature ambientali vengono discretizzate –

¹⁷⁰<http://www.rainews.it/dl/rainews/media/Siberia-la-raccolta-del-ghiaccio-nella-citta-piu-fredda.html> Dalle informazioni sappiamo che la funzione densità vale 0 già da 0° in poi, e certamente prima di -273° ma in pratica, usando le conoscenze usuali della realtà sensibile, da -100° .

come si usa – in valori interi. La parola “normale” riferita al valore -41° si riferisce a una quarta area, purché la normalità si riferisca al significato $\leq -41^\circ$, altrimenti nel caso continuo la probabilità dei -41° esatti è 0, e nel caso discreto sarà senz’altro *piccola*, tutt’altro che *grande* (“normale”).

In mancanza di altre informazioni – che sarebbe Medicina, che qua non vogliamo e non possiamo fare – per le temperature corporee umane possiamo aspettarci similmente un grafico a campana, seppure situato altrove, intorno ai 37° .

Similmente per glicemie, sideremie, colesterolemie, e quant’altro.

Esercizio. Si disegni le funzione di ripartizione della $B(2, \frac{1}{2})$ prima considerata, nell’Esempio 1. (Si osserverà che i 3 salti hanno ampiezza pari alle corrispondenti 3 probabilità, e questo è un fatto generale che succede per tutte le variabili aleatorie discrete, come già osservato.)

Abbandoniamo adesso le densità binomiali.

31.3 Densità in qualche modo bimodali

Diremo *bimodale* una densità “come” quella in Fig. 31.3. (102)

Al livello di questo testo elementare non possiamo definire rigorosamente cosa intendiamo per densità discreta con 2 (o eventualmente più, ma qua non ce ne occuperemo) punti di massimo relativo, o addirittura assoluto: è ovvio che in qualche modo bisognerà prescindere da singoli valori eccezionali, ma definirlo matematicamente non sarebbe facile. Comunque in molti casi semplici la *bimodalità* si riconosce a prima vista.

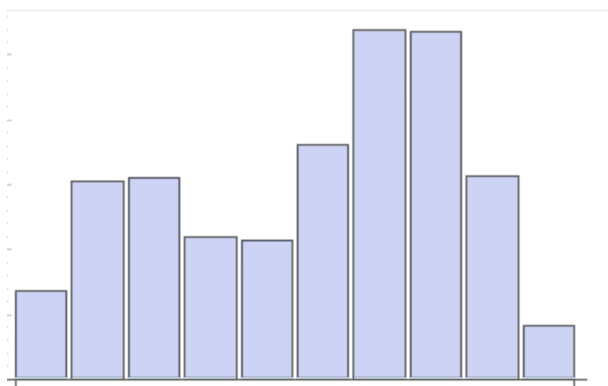


Figure 45: Densità bimodale. Screenshot da WolframAlpha

Distribuzioni "bimodali" fanno ritenere che "dietro" vi siano 2 fenomeni distinti.

Per esempio se misuriamo i tempi di impegno ad uno sportello, e dopo aver (sapientemente) riunito i tempi misurati in classi facciamo l'istogramma, e questo presenta 2 massimi nettamente distinti, lontani, è plausibile che allo sportello i clienti stiano facendo 2 operazioni diverse: chi un'operazione più rapida, chi una più laboriosa. (Nella zona intermedia fra i 2 massimi sono i veloci a fare l'operazione laboriosa e i lenti a fare quella semplice).

Densità bimodali possono rivelare fenomeni interessanti. Per esempio in certi tempi e certi luoghi, il tempo che intercorre fra il matrimonio e la nascita del primo figlio si distribuisce con una tipica forma "a due gobbe", con un massimo assoluto intorno ai 9 mesi, corrispondente ai figli concepiti nel viaggio di nozze, e un massimo relativo vagamente alquanto prima (intorno ai 6-7 mesi nella statistica in Figura)

Nati vivi in ciascun mese successivo alla celebrazione del matrimonio.
(Australia 1911-1931)

Mesi	Maschi	Femmine	Maschi % Femmine
fino a 1	4528	4174	108,48
1- 2	5621	5211	107,87
2- 3	7933	7509	105,65
3- 4	11067	10146	109,08
4- 5	14223	12780	111,29
5- 6	18459	17139	107,70
6- 7	23804	22611	105,28
7- 8	21290	20778	102,46
8- 9	16538	15529	106,50
9-10	36075	33878	106,49
10-11	30457	29238	104,17
11-12	22348	21020	106,32

Figure 46: Da: *Il rapporto nei sessi nei nati da concepimento antenuziale*, Diego De Castro; Giornale degli Economisti e Rivista di Statistica, Serie quarta, Vol. 73 (Anno 48), No. 10 (ottobre 1933), p. 733. Online in <https://www.jstor.org/stable/23229506>

In un articolo⁽¹⁷¹⁾ scientifico si è studiato perchè le valutazioni soggettive di benessere si distribuiscano – o paiono distribuirsi, la questione è complessa e viene studiata nell'articolo – in modo bimodale, con tante persone che danno una variamente bassa valutazione della loro salute, e tante altre che danno una variamente alta valutazione:

The EQ-5D is internationally one of the most widely used health-related quality-of-life instruments.¹ It is therefore very important to obtain a good understanding of the characteristics of EQ-5D data. Many studies of the original 3-level version of the EQ-5D (now known as the EQ-5D-3L) have found that its index scores typically have a nonnormal distribution divided into 2 distinct groups. (This is often described as "bimodal," but we will argue that this label is misleading.) This phenomenon has been found in both patient and general populations, for many conditions, and using different country-specific value sets.

¹⁷¹Parkin D, Devlin N, Feng Y. What Determines the Shape of an EQ-5D Index Distribution? Medical Decision Making. 2016;36(8):941-951. doi:10.1177/0272989X16645581

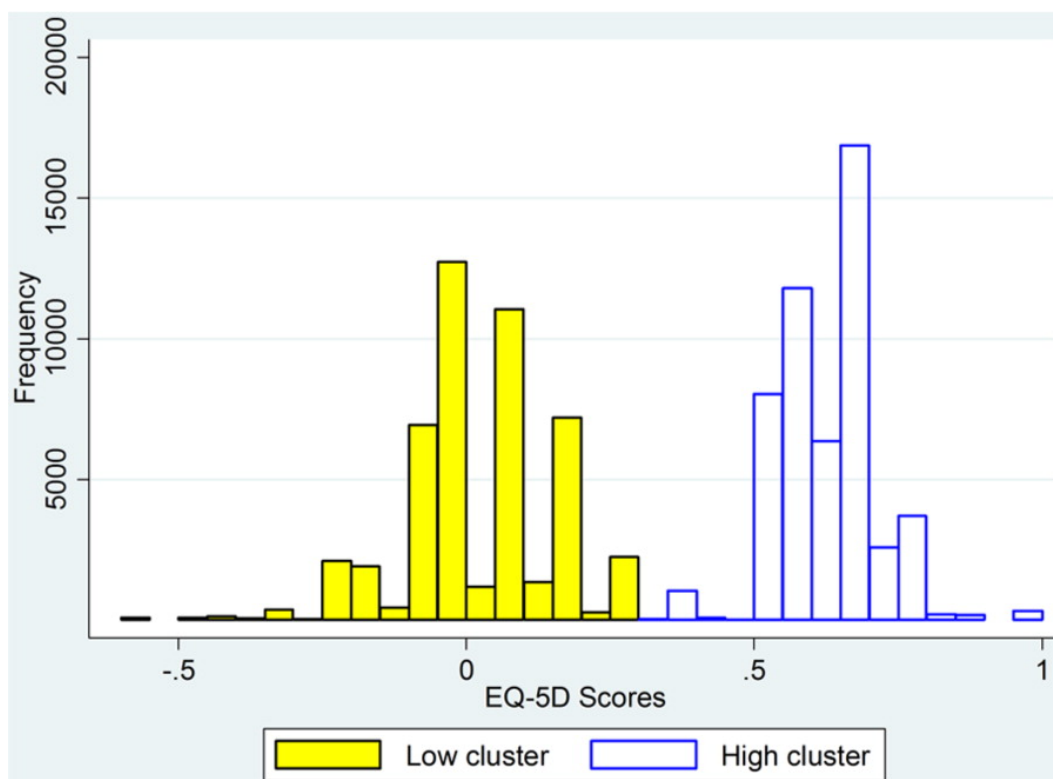


Figure 47: Figura 2 del citato articolo: "Distribution of presurgery EQ-5D index scores for hip replacement patients, using UK-weighted index."

Ancora, leggiamo⁽¹⁷²⁾ in https://en.wikipedia.org/wiki/Multimodal_distribution

Examples of variables with bimodal distributions include the time between eruptions of certain geysers, the color of galaxies, the size of worker weaver ants, the age of incidence of Hodgkin's lymphoma, the speed of inactivation of the drug isoniazid in US adults, the absolute magnitude of novae, and the circadian activity patterns of those crepuscular animals that are active both in morning and evening twilight.

Si noti la – in qualche modo sorprendente – *age of incidence of Hodgkin's lymphom*.

¹⁷²Retrieved 24 november 2022

31.4 La natura empirica delle densità bimodali

Si noti che le distribuzioni bimodali tendono a ricorrere da rilevazioni di dati nella realtà sensibile, e quindi le frequenze corrispondono a probabilità in senso frequentista. Invece non è bimodale nessuno dei pur numerosi tipi di densità discrete e continue che più classicamente si definiscono a livello teorico, con formule (e parecchi sono esposti in questo testo elementare). Leggiamo ancora nella sopra citata pagina di Wikipedia:

Bimodal distributions, despite their frequent occurrence in data sets, have only rarely been studied

Bottom line. Tuttavia, non solo a livello teorico ma anche empirico ovvero sperimentale, le distribuzioni unimodali, e più specificamente più o meno a campana, sono immensamente più frequenti di quelle multimodali, e dopo questa breve parentesi in questa Lezione, quasi di quelle sole ci occuperemo.



Nota. Di questa Lezione ci si dovrebbe aspettare che lo studente conosca a memoria le formule *contemporaneamente* numerate *et* riquadrate, oltre a tutto l'inquadramento teorico della questione.

31.5 ESERCIZI SULLA LEZIONE 39

31.5.1 Esercizio risolto a - Dado regolare a 8 facce

f
 μ Che probabilità c'è di ottenere almeno 2 volte il numero 7 con 5 lanci di un dado regolare a 8 facce?

Potremmo sommare le 4 probabilità

$$P(k=2) + P(k=3) + P(k=4) + P(k=5)$$

ma è ovvio che conviene considerare l'evento complementare e cioè calcolare

$$\begin{aligned} & 1 - (P(k=0) + P(k=1)) = \\ &= 1 - \binom{5}{0} \left(\frac{1}{8}\right)^0 \left(1 - \frac{1}{8}\right)^{5-0} - \binom{5}{1} \left(\frac{1}{8}\right)^1 \left(1 - \frac{1}{8}\right)^{5-1} = \\ &= 1 - 1 \cdot 1 \cdot \left(\frac{7}{8}\right)^5 - 5 \cdot \left(\frac{1}{8}\right)^1 \left(\frac{7}{8}\right)^4 = \\ &= 1 - \left(\frac{7}{8}\right)^4 \cdot \left(\frac{7}{8} - \frac{5}{8}\right) = 1 - \left(\frac{7}{8}\right)^4 \cdot \frac{2}{8} = 1 - \frac{2401}{16384} = \\ &= \frac{13983}{16384} \approx 0.853 = 85.3\%. \end{aligned}$$

Esercizio μ Che probabilità c'è di ottenere 3 volte il numero 4 con 5 lanci di un dado regolare a 6 facce?

Esercizio μ Che probabilità c'è di ottenere almeno 2 volte il numero 3 con 5 lanci di un dado regolare a 6 facce?

Esercizio μ Che probabilità c'è di ottenere almeno 3 volte il numero 3 con 5 lanci di un dado regolare a 6 facce?

Esercizio μ Che probabilità c'è di ottenere almeno 2 volte numeri primi con 4 lanci di un dado regolare a 6 facce?

Esercizio μ Che probabilità c'è di ottenere un numero primo di volte testa con 5 lanci di una moneta regolare?

Esercizio μ Che probabilità c'è di ottenere un numero primo di volte croce con 6 lanci di una moneta regolare?

Esercizio μ Che probabilità c'è di ottenere un numero primo di volte testa con 7 lanci di una moneta regolare?

Esercizio μ Che probabilità c'è di ottenere un numero dispari di volte testa con 5 lanci di una moneta regolare?