32 Speranza matematica e varianza

La speranza matematica di una v.a. essenzialmente è un numero che congloba
la probabilità degli eventi
col loro
valore numerico (dato da una v.a.) pensabile magari come un costo ovvero guadagno.

La varianza di una v.a. estende il concetto già visto di varianza di un dataset numerico ovvero di una popolazione: è una misura della variabilità.

Consideriamo in questa Lezione variabili aleatorie discrete, e nella Lezione 34 le continue.

32.1 Speranza matematica (o valore atteso, o media)

EN: expected value, expectation, mathematical expectation, mean, average, first moment

Un esperimento ideale

Consideriamo questo esperimento ideale: estrarremo da un bicchier d'acqua distillata un atomo a caso, e sia X la variabile aleatoria che conta i suoi elettroni. L'atomo sarà di idrogeno o di ossigeno, con 1 elettrone o 8 elettroni rispettivamente.
In un'estrazione quanti elettroni possiamo attenderci mediamente? Questo è il significato di speranza matematica della variabile aleatoria X.
La speranza matematica di X è da intendersi come la media di numerosissime sue determinazioni ovvero realizzazioni. Per esempio potremmo trovare questi valori, in estrazioni successive:

1 1 1 8 1 8 1 1 8 1 1 1 8 1 1....

Naturalmente l'1 tenderà a ricorrere doppiamente dell'8 perchè l'idrogeno è doppiamente frequente nel bicchiere di H_2O rispetto all'ossigeno.
È ovvio che la media tenderà a

$$\frac{2}{3} \cdot 1 + \frac{1}{3} \cdot 8$$

cioè $\frac{10}{3} \approx 3.33$.
Ecco, questa è la speranza matematica del numero di elettroni di

un'estrazione; ovvero, il *valore atteso*, termine però fuorviante, perchè non possiamo affatto attenderci di estrarre 3.333... elettroni: ne verranno 8 oppure 1.

Un esempio con un gioco

Supponiamo che ora lanceranno un dado, e io posso scegliere, a costo 0,

— o di ricevere 2 euro se viene pari

— o di ricevere 5 euro se viene 3;

cosa mi conviene scegliere?

• Se spero nel pari, mediamente vincerei metà volte, e allora posso considerare che mediamente vincerò metà dei 2 euro in palio, cioè 1 euro;

• se spero nel numero 3, mediamente vincerei 5 euro 1 volta su 6, e allora posso considerare che mediamente vincerò 5/6 dell'euro euro in palio.

Allora conviene la scommessa sul pari perché 1 è più di 5/6.

Il caso generale

Il concetto è immensamente più generale dei giochi d'azzardo, perché non solo con le assicurazioni, ma anche coi fatti dell'economia e della salute pubblica e individuale, in ultima analisi, la situazione con

spese certe

e ricavi incerti

è analoga ad un gioco d'azzardo.

Anche i farmaci, che pure salvano tante persone e magari in casi acuti (come infarto e incidente stradale) gli regalano decenni di vita ulteriore in passato impossibili, talvolta fanno pure danni, e anche quelli vanno messi in conto, in un bilancio costi-benefici.

Si pensi, per esempio, al rofecoxib, cioè il Vioxx, che tra il 1998, e il 2004, quando fu ritirato, si stima abbia causato morti cardiache in eccesso, quasi 100.000 nell'ipotesi migliore.

Un'attenta considerazione del caso numerico soprastante ci induce a definire come **speranza matematica** (o *media* o *valore atteso*) di una variabile aleatoria discreta $X := \binom{x_1\, x_2\, x_3\, ...\, x_k\, [...]}{p_1\, p_2\, p_3\, ...\, p_k\, [...]}$

$$\boxed{\mu = E(X) := \sum_k x_k{\cdot}p_k \qquad \textit{somma su tutti i valori } x_k \textit{ della v.a.}} \tag{103}$$

e questa è una somma finita o una serie, e in quest'ultimo caso a livello teorico si richiede anche una condizione di regolarità, $\sum_k |x_k| p_k < +\infty$ che in questa trattazione elementare diamo ma mai calcoleremo, e comunque è sempre verificata per le variabili aleatorie che considereremo.

32.2 Costi e benefici

La speranza matematica matematizza il concetto di costi e benefici. Un intervento farmaceutico o più in generale medico su una persona ha, semplificando, un costo con una certa probabilità, e un beneficio con un'altra. Entrambi sono difficili da quantificare; in particolare: dopo quanto tempo si esegue la misura? È ovvio che un farmaco per una malattia cronica che dà un certo beneficio per un anno, e magari 0 danni, potrebbe presentare ben peggiori risultati dopo 10 anni.

Ma anche supponendo una misura perfetta, un singolo potrebbe avere preferenze ben diverse dal sistema sanitario. Consideriamo per esempio 2 terapie per una malattia che causa atroce sofferenza:

Terapia 1:

20% morte

80% guarigione

Terapia 2:

10% morte

80% passa da atroce sofferenza a grave sofferenza

10% guarisce

Il sistema potrebbe preferire la Terapia 2, che ha metà rischio di morte e di solito (90%) migliora la situazione del paziente, ma qualcuno preferirebbe la 1, che gli dà una probabilità dell'80% di guarigione, invece che di passare da sofferenza atroce a grave.

Ancora più complessa la situazione quando il destinatario dell'intervento non è un singolo individuo ma la collettività stratificata in classi che pagano costi diversi a fronte di vantaggi diversi. Per esempio il distanziamento sociale e più in generale tutta la covidizzazione della società fa pagare un prezzo altissimo ai giovani, senza nessun apprezzabile diminuzione di rischio di morte per loro, che in pratica hanno solo costi; l'ipotetico vantaggio è praticamente solo per altre classi di età, più avanzate.

È ben evidente che poi la questione si intreccia coi vantaggi non sanitari, bensì economici (ed eventualmente pure di altro genere), che particolari segmenti della società (numericamente minuscoli), possono ottenere da grandi interventi sanitari.

In generale per tutti gli interventi sulla società in ogni caso ci sono costi e benefici, ma quasi mai i benefici sono esattamente per le stesse persone che pagano i costi.
A questo mondo tutto si fa per un presunto beneficio superiore al danno previsto, ma i soggetti destinatari sono sempre diversi.
Coi media quelli che hanno soldi e potere possono convincere la maggioranza che quello che si fa è vantaggioso per tutti, e infatti certamente è vantaggioso – ma più per loro.

Un esempio eclatante, collegato, è il grande fenomeno umano della guerra, che tanto incide sulle cause di morte a livello mondiale. Si fanno delle stime di possibili guadagni e perdite, più o meno probabili, e infine si decide:

"Armiamoci e partite!"

Nessuna legge al mondo vieta di sognare che i fatti della politica, e delle politiche sanitarie, avvengano in base a criteri diversi.

32.3 Varianza

Se esiste la speranza matematica, si definisce la **varianza** di una v.a. discreta X in 2 modi equivalenti

$$\sigma^2 = Var(X) := \sum_k (x_k - E(X))^2 \cdot p_k = E(X^2) - E(X)^2 \qquad (104)$$

e la radice quadrata si dice **deviazione standard**, σ o SD, che ha la stessa unità di misura della variabile aleatoria.

Per esempio le giocate sul pari e sul 3, prima considerate, producono guadagni medi rispettivi

$$E(U) = \frac{1}{2} \cdot 2 = 1 \qquad E(V) = \frac{1}{6} \cdot 5 = \frac{5}{6} \approx 0.833.$$

Ma con diversa varianza, che in qualche modo esprime la tendenza a rischiare molto per avere molto: la si calcoli nei 2 casi con le formule soprastanti.

Teorema 1.

Legge	$E(X)$ μ	$Var(x)$ σ^2
$X \sim \mathbb{U}\{a, b\}$	$\frac{a+b}{2}$	$\frac{(b-a+1)^2-1}{12}$
$X \sim B(n, p)$	$n \cdot p$	$np(1-p)$
X geometrica iniziante da 0	$\frac{1-p}{p}$	$\frac{1-p}{p^2}$
X geometrica iniziante da 1	$\frac{1}{p}$	$\frac{1-p}{p^2}$

Esercizio$^{f}_{\mu}$ Qual è il valore atteso del punteggio di un dado? E la varianza?
Usando la definizione di speranza matematica si ha

$$X := \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{pmatrix} \quad E(X) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} =$$

$$= \frac{1}{6} \cdot (1 + 2 + 3 + 4 + 5 + 6) = \frac{21}{6} = \frac{7}{2} = 3.5.$$

Usando il teorema sulla speranza matematica di $\mathbb{U}\{a, b\}$

$$X \sim \mathbb{U}\{1, 6\} \Rightarrow E(X) = \frac{1+6}{2} = \frac{7}{2} = 3.5 \; (come\ sopra).$$

Con le 2 definizioni di varianza si ha ugualmente

$$Var(X) = \left(1 - \frac{7}{2}\right)^2 \cdot \frac{1}{6} + \left(2 - \frac{7}{2}\right)^2 \cdot \frac{1}{6} + \ldots + \left(6 - \frac{7}{2}\right)^2 \cdot \frac{1}{6} = \ldots = \frac{35}{12}$$

$$X^2 = \begin{pmatrix} 1 & 4 & 9 & 16 & 25 & 36 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{pmatrix} \quad Var(X) = 1^2 \cdot \frac{1}{6} + \ldots + 6^2 \cdot \frac{1}{2} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}$$

e con il teorema sulla varianza di $\mathbb{U}\{a, b\}$ si ha ugualmente

$$X \sim \mathbb{U}\{1, 6\} \Rightarrow Var(X) = \frac{(6-1+1)^2 - 1}{12} = \frac{36-1}{12} = \frac{35}{12}.$$

32.4 Note su varianza e istogramma della densità

$$\boxed{\text{piccola varianza: valori tendenzialmente vicini alla media}} \tag{105}$$

Grande varianza ovvero deviazione standard: valori dispersi.

Piccola varianza ovvero deviazione standard: valori addensati presso la media.

Per la varianza, esistono varianze più piccole (valori più addensati) o più grandi di altre (valori più dispersi) ma non ha molto senso considerare grande o piccola una singola varianza.
Quanto sopra detto vale per qualunque densità discreta (e in effetti, poi, varrà anche per le continue). (Anche se “bimodale").

Se l'istogramma di una densità discreta (o, in seguito, il grafico di una densità continua) ha una forma più o meno a campana, com'è per le variabili aleatorie binomiali,

a grande varianza corrisponde campana bassa e larga

a piccola varianza corrisponde campana alta e stretta.

Quest'osservazione vale anche per l'uniforme discreta e le geometriche, che, con qualche elasticità, si può pure dire che abbiano densità a forma più o meno a campana – seppure molto vagamente. E similmente varrà per l'uniforme continua, l'esponenziale, e molte altre.

32.5 L'aspettativa di vita o speranza di vita

Consideriamo un esserino robotizzato che viene creato il 1 gennaio, e il 31 dicembre lancia al suo interno un dado e se viene 1 si spegne, muore. Se è sopravvissuto, il 31 dicembre dell'anno dopo fa la stessa cosa e così via. La durata della sua vita è una variabile aleatoria di cui possiamo considerare la speranza matematica. In questo caso semplice, è una variabile aleatoria geometrica iniziante da1 di parametro $\frac{1}{6}$ e allora con speranza matematica 6, che è l'aspettativa di vita dell'esserino, in anni.

Possiamo ipotizzare un esserino automatico che lanci due dadi e una moneta e cessa di vivere solo con 1, 1, testa, che ha probabilità $\frac{1}{72}$, e allora l'esserino ha vita media 72 anni, e l'aspettativa di vita nel mondo è più o meno quella.

È un modello della situazione relativa agli esseri umani, ma con varie semplificazioni:

– l'essere umano può morire ogni giorno, ovviamente in generale con probabilità molto minore di $\frac{1}{72}$;

– la probabilità di morte non è costante negli anni: è molto maggiore nel primo anno che nel secondo, e in vecchiaia aumenta enormemente: un centenario in generale ha probabilità molto minore di $\frac{71}{72}$ di vivere un altro anno;

– la probabilità di morte ad un determinato anno di età varia

globalmente nel tempo: un diciottenne del 2018 aveva molta più probabilità di arrivare a 19 anni dei diciottenni del 1918, soggetti a Prima Guerra Mondiale – e altre difficoltà di quel tempo.
Con riferimento all'ultimo punto, come si fa a calcolare la speranza di vita, se non sappiamo quali guerre eventualmente ci saranno, e altre catastrofi, o più in generale cambiamenti – anche migliorativi, ovvio – della sopravvivenza?
È assolutamente impossibile calcolarla, e non ha neppure alcun senso reale la probabilità che una persona vivente oggi sia viva fra 1 anno, senza sapere neppure se il mondo esisterà ancora. (Come sarebbe definita una tale probabilità? Non ha alcun senso). Quello che viene chiamato aspettativa o speranza di vita, detto molto semplificatamente, è il risultato di un calcolo in cui si considera tutto “a bocce ferme", cioè come se le cose dovessero andare come sono andate finora. (Per i ventenni, i ventunenni, i ventiduenni, eccetera, per tutte le classi di età). *Non è una vera speranza matematica* in senso stretto, ma in qualche modo la “imita".

Oltre all'aspettativa di vita alla nascita, si può considerare l'aspettativa di vita residua, ad una determinata età. Per gli esserini robotizzati, è uguale a quella alla nascita, perché la variabile aleatoria geometrica gode dell'*assenza di memoria*: restano sempre da vivere mediamente 6 o 72 anni nei due casi considerati. (E similmente costante è il tempo da attendere ancora un determinato numero, o ambo, eccetera, del lotto, anche se “in ritardo", contrariamente a quello che pensano ingenui giocatori). Per l'essere umano ovviamente non è così, per il variare della mortalità al progredire dell'età. In particolare in Italia l'aspettativa di vita a 65 anni ha un importante valore legale e infine sociale perché a quel valore, determinato dall'ISTAT, Istituto Nazionale di Statistica, è per legge legata la cosiddetta *età pensionabile.*

Sulla Gazzetta Ufficiale della Repubblica Italiana, Serie Generale - Anno 160, Numero 267, del 14 novembre 2019, nel Decreto 5 novem-

bre 2019 del Ministero dell'Economia e delle Finanze, leggiamo: "Vista la nota del Presidente dell'Istituto nazionale di statistica (ISTAT) n. 2768968/19 del 16 ottobre 2019, con cui si comunica che la variazione della speranza di vita all'età di sessantacinque anni e relativa alla media della popolazione residente in Italia ai fini dell'adeguamento dei requisiti di accesso al pensionamento con decorrenza 1° gennaio 2021 corrispondente alla differenza tra la media dei valori registrati negli anni 2017 e 2018 e il valore registrato nell'anno 2016 è pari a 0,021 decimi di anno".
(Si tratta, ovviamente, di questione connessa anche a fattori medici e farmaceutici, oltre che sociali, politici e geopolitici; la spesa farmaceutica statale è in enorme diminuzione negli ultimi anni, pare ridotta del 70 percento in un decennio).

32.6 Passeggiate aleatorie

Traiamo da Wikipedia, l'enciclopedia libera, versione inglese alla voce "Random walk", in italiano *passeggiata aleatoria"*:

> A random walk is a mathematical object, known as a stochastic or random process, that describes a path that consists of a succession of random steps on some mathematical space such as the integers. An elementary example of a random walk is the random walk on the integer number line, $\mathbb{Z}$, which starts at 0 and at each step moves +1 or −1 with equal probability. Other examples include the path traced by a molecule as it travels in a liquid or a gas, the search path of a foraging animal, the price of a fluctuating stock and the financial status of a gambler can all be approximated by random walk models, even though they may not be truly random in reality.
> (...) If a and b are positive integers, then the expected number of steps until a one-dimensional simple random walk starting at 0 first hits b or $-a$ is ab.

Cioè detto T il numero di passi fino al raggiungimento di b o $-a$,

$$\boxed{E(T) = a \cdot b} \tag{106}$$

Traiamo da https://commons.wikimedia.org/wiki/File:Random_Walk_example.svg la seguente figura, che rappresenta (la parte iniziale di) 8 cammini aleatori. ???FIGURA VEDO TEX

Ecco su WolframAlpha una passeggiata aleatoria di 100 passi, nuova ad ogni richiamo del LINK->

Su queste basi ipotizziamo ora un "parametro vitale complessivo", che non riusciamo in questo esempio semplificato a specificare nei dettagli, ma che comunque non deve superare il valore soglia 7, nè scendere sotto il valore soglia -7, variando di anno in anno nell'individuo, salendo o scendendo di 1 con uguale probabilità 50%. Possiamo proprio immaginare un robottino che ogni 31 dicembre lancia una moneta e somma 1 se viene testa o sottrae 1 se viene croce ad un suo "capitale virtuale", partito da 0, e muore ovvero si spegne se sfora i limiti prestabiliti. In base a quanto sopra detto, con $a = b = 8$, la speranza matematica della vita è 64 anni, cioè $8 \cdot 8$. Ma se ora (come in un videogioco si danno le vite supplementari) supponiamo che la prima volta che una soglia (8 o -8) viene raggiunta, un farmaco modifica di 1 o -1 il parametro vitale nella direzione giusta, ecco che al nostro ipotetico soggetto la Farmacia regala altri anni di vita! Superato il momento critico, quando rischiava di andare a 8 o -8 e morire, ricomincia una nuova vita; certo la ricomincia in posizione rischiosa, a 7 o -7, e al passo successivo può morire (e non è detto che gli sia concessa una seconda salvezza), ma non è detto, magari si allontana dal limite invalicabile e sopravvive ancora a lungo.
Questo può essere visto come uno dei significati della Farmacia: la cura urgente del caso acuto (con qualche speranza che il soggetto si allontani poi dalla situazione limite).

Oltre all'aspettativa di vita alla nascita, la mortalità nel primo anno è l'altro grande parametro usato per valutare lo stato di salute complessivo di una popolazione.
Leggiamo su Wikipedia, l'enciclopedia libera:

> Of the 27 most developed countries, the U.S. has the highest Infant Mortality Rate, despite spending much more on health care per capita.

La mortalità infantile in sostanza è una probabilità, in senso frequentista, mentre l'aspettativa di vita alla nascita è una sorta di speranza matematica.

32.7 Esercizi sulla speranza matematica di v.a. discrete

Esercizio$^{f}_{\mu}$ Qual è la vincita netta attesa giocando 1 euro su una cinquina del lotto? (Con la parola "netta" intendiamo che conteggiamo la perdita sicura di 1 euro, il costo della giocata).
Detta V la v.a. "vincita netta", ricordando che si vince 6 milioni di volte la giocata, con probabilità $1/\binom{90}{5} = 1/43\,949\,268$,

$$V = \begin{pmatrix} 5\,999\,999 & -1 \\ \frac{1}{43\,949\,268} & \frac{43\,949\,267}{43\,949\,268} \end{pmatrix}$$

$$E(V) = 5\,999\,99 \cdot \frac{1}{43\,949\,268} - \frac{43\,949\,267}{43\,949\,268} = -\frac{37\,949\,268}{43\,949\,268} =$$

adesso magari semplifichiamo per 2, poi ancora per 2, poi per 3

$$= -\frac{3\,162\,439}{3\,662\,439} \approx -0.8635\,euro$$

cioè una perdita media di circa 86 centesimi di euro. (In generale i giochi d'azzardo sono molto meno svantaggiosi di questo).

Esercizio$^{f}_{\mu}$ Qual è la vincita netta attesa giocando 1 euro su un numero della roulette europea?
La vincita avviene con probabilità $1/37$ perché i numeri equiprobabili sono 37, da 0 a 36, e dà 36 volte la giocata:

$$V = \begin{pmatrix} 35 & -1 \\ \frac{1}{37} & \frac{36}{37} \end{pmatrix}$$

$$E(V) = 35 \cdot \frac{1}{37} - \frac{36}{37} = -\frac{1}{37} \approx 0.027\, euro$$

cioè mediamente perdiamo, molto approssimativamente, 3 centesimo ad ogni giocata di 1 euro.

Nota. Di questa Lezione ci si dovrebbe aspettare che lo studente conosca a memoria le formule *contemporaneamente* numerate *et* riquadrate, e le 4 formule della speranza matematica del Teorema 1, oltre a tutto l'inquadramento teorico della questione.

IX – Variabili aleatorie continue