05BQX Metodi Matematici per l'ingegneria 2011-2012

Marina Santacroce

Dipartimento di Scienze Matematiche, Politecnico di Torino

4. Variabili aleatorie

Riferimenti: S.Ross *Calcolo delle probabilità* Cap.4

Outline

Variabili aleatorie: esempi

Molto spesso si è interessati a funzioni degli esiti degli esperimenti più che agli esiti stessi.

- Consideriamo il lancio di due dadi. Si potrebbe essere interessati all'evento "somma uguale a 6", piuttosto che al *come* si è ottenuta.
- Nel caso di n lanci di una moneta, si potrebbe essere interessati al numero dei lanci che hanno dato come esito "testa", piuttosto che alla sequenza esatta.
- In una successione infinita di lanci di una moneta, si potrebbe voler sapere il lancio in cui si è ottenuta "testa" per la prima volta.
- Da un'urna con N palline, di cui m rosse e $N - m$ blu, si estraggono n palline senza reimmissione. Una funzione dell'esito dell'esperimento potrebbe essere il numero di palline rosse estratte.

Per ognuno di questi esempi si può introdurre una "funzione" degli esiti, che chiameremo *variabile aleatoria*. In particolare, le variabili associate ad ognuno di questi sono variabili aleatorie "*notevoli*", che studieremo presto nel dettaglio.

Esempio

Supponete si voglia provare a passare un esame tentando ad ogni appello con la stessa preparazione e che ogni volta la probabilità di passarlo sia p. Se si viene bocciati, si riprova all'appello successivo. Se X è il numero di volte che si ripete l'esame per passarlo.
Si ha

- $\mathbb{P}(X = 1) = \mathbb{P}(\text{Pass}) = p$,
- $\mathbb{P}(X = 2) = \mathbb{P}(\text{Bocc, Pass}) = (1 - p)p$,
- ....
- $\mathbb{P}(X = n) = \mathbb{P}(\text{Bocc,..., Bocc (n-1 volte), Pass}) = (1 - p)^{n-1}p$.

Si può vedere che

$$\sum_{k:1}^{\infty} \mathbb{P}(X = k) = \sum_{k:1}^{\infty} (1 - p)^{k-1} p = 1$$

Distribuzione di una variabile aleatoria

Def. Dato uno spazio campionario S, una famiglia di eventi e una probabilità $\mathbb{P}$, una *variabile aleatoria* X è una funzione $X : S \to \mathbb{R}$ tale che, per ogni $B \subset \mathbb{R}$, $\{X \in B\} = \{s : X(s) \in B\}$ è un evento[1].

All'evento $\{X \in B\}$ è associata una *probabilità* $\Rightarrow$ $\mathbb{P}_X(B) = \mathbb{P}(X \in B)$.

- $\mathbb{P}_X(\cdot)$ è una *misura di probabilità su* $\mathbb{R}$ e si chiama *legge di* X.

Def. La *funzione di distribuzione o di ripartizione* di una variabilie aleatoria X è una funzione $F : \mathbb{R} \to [0; 1]$ definita da

$F(x) = \mathbb{P}(X \leq x), \qquad -\infty < x < \infty.$ [2]

Da questa definizione si ricavano immediatamente le seguenti proprietà:

Prop. La funzione di distribuzione F di una variabile aleatoria X è

1. non decrescente;
2. continua a destra, i.e. se $\{x_n, n \geq 1\}$ è una successione decrescente $x_n \downarrow x$, allora $\lim_{n\to\infty} F(x_n) = F(x)$;
3. $\lim_{x\to-\infty} F(x) = 0$ e $\lim_{x\to\infty} F(x) = 1$.

[1] B deve essere nella σ-algebra di Borel di $\mathbb{R}$

[2] $F(x)$ non è altro che la legge di x calcolata per $B = (-\infty; x]$

Variabili aleatorie discrete

Def. Una variabile aleatoria X che può assumere un numero finito o al più numerabile di valori è detta **variabile aleatoria discreta**.

- L'insieme di questi valori che la variabile X assume con probabilità positiva si chiama *supporto* di X e lo indichiamo con $\mathbb{X}$.
- Ad ogni variabile aleatoria discreta X è associata una funzione di densità discreta $p : \mathbb{X} \to [0; 1]$

  $p(x_i) = \mathbb{P}(X = x_i)$, per ogni valore $x_i \in \mathbb{X}$.

Oss. Chiaramente $p(x_i) > 0$, per $x_i \in \mathbb{X}$ e $\sum_{x_i \in \mathbb{X}} p(x_i) = 1$.

$\Rightarrow$ La funzione di densità discreta p descrive completamente la *legge di X*.

La *funzione di distribuzione* di una variabile aleatoria *discreta* X è una funzione a gradini definita da

$$F(x) = \sum_{x_i \leq x} p(x_i), \qquad -\infty < x < \infty.$$

Valore atteso

Def. Sia X una variabile aleatoria discreta. Il valore atteso o media di X è dato da

$$\mathbb{E}(X) = \sum_{x_i \in \mathbb{X}} x_i p(x_i),$$

se la serie converge[3].

Esempio: Si lancia un dado equo e X è il numero uscito. Allora

- $p(i) = \frac{1}{6}; \quad i : 1, \ldots, 6.$

$$\mathbb{E}(X) = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + 3 \times \frac{1}{6} + 4 \times \frac{1}{6} + 5 \times \frac{1}{6} + 6 \times \frac{1}{6} = \frac{1+2+3+4+5+6}{6} = \frac{21}{6} = 3.5$$

[3] Supponiamo che $\mathbb{E}(|X|) < \infty$, cioè che il valor atteso esista sempre

Esempio: Tram in Melbourne

In Melbourne (Australia), non si può acquistare il biglietto a bordo del tram. Solo occasionalmente, dei controllori passano e danno una multa ai passeggeri senza biglietto.
Il biglietto costa 2\$, la multa è di 30\$ e la probabilità di incontrare un controllore è del 5%.
Si supponga di voler mimimizzare il costo medio,
$\Rightarrow$ *dovremmo comprare il biglietto o no*?

C è il costo.

- $\Rightarrow$ Comprando il biglietto, il costo ogni volta è C = 2\$ con probabilità 1.
  $\Rightarrow \mathbb{E}(C) = 2 \times 1 = 2\$$
- $\Rightarrow$ Se non lo si compra il costo è
  - $C = 0\$$ se non si incontra il controllore $\Rightarrow \mathbb{P}(C = 0) = 0.95$
  - $C = 30\$$ se si incontra il controllore $\Rightarrow \mathbb{P}(C = 30) = 0.05$.

  Quindi,
  $$\Rightarrow \mathbb{E}(C) = 0 \times 0.95 + 30 \times 0.05 = 1.50\$$$

Per minimizzare il costo, non comprare il biglietto del tram in Melbourne!

Esempio: chi vuole essere milionario?

Avete vinto $64,000 e dovete affrontare una domanda per vincerne $125,000. Quindi riceverete

- $125,000 se rispondete correttamente.
- $32,000 se la risposta è sbagliata.
- $64,000 se decidete di non rispondere.

Quale sarà la strategia migliore?

La domanda da $125,000 è: *Qual è la città natale del vostro XXX professore*?

A- Napoli B- Bolzano

C-Torino D- Roma

Se rispondete, qual è il valore atteso della vincita? È meglio rispondere o no?
Se non avete idea e tirate ad indovinare il valore atteso della vincita è

$$\frac{1}{4}125,000 + \frac{3}{4}32,000 = 55,250 < 64,000$$

mentre se non rispondete è $ 64,000.

Sembra più opportuno non rispondere anche se si potrebbe rischiare e provarci comunque.

Esempio: Chi vuole essere un milionario?(BIS)

Supponete che con l'opzione 50:50 vi liberate di due delle possibili risposte e rimanete con le due opzioni

A- Napoli D- Roma

Supponete di decidere di rispondere. Qual è il valore atteso della vincita? È meglio rispondere o no?

Il valore atteso

$$\frac{1}{2} \times 125,000 + \frac{1}{2} \times 32,000 = 78,500 > 64,000.$$

Si potrebbe rispondere ma una persona avversa al rischio sicuramente non lo farebbe!

Utilità

- Abbiamo visto degli esempi in cui il calcolo del valore atteso del costo/premio serviva per prendere decisioni: comprare o meno il biglietto, rispondere o meno alla domanda.
- ma le persone si comporterebbero in modo diverso le une dalle altre. Perchè?

$\Rightarrow$ Le persone hanno differenti atteggiamenti rispetto al richio, che possono essere rappresentati con differenti funzioni di utilità. Per prendere delle decisioni è opportuno tenerne conto massimizzando il valore atteso dell'utilità del costo/premio

$$\max \qquad \mathbb{E}(U(X))$$

- $U(x) = \log x$ è una tipica utilità per una persona avversa al rischio.
- questo spiega perchè alcune persone compreranno sempre il biglietto o perchè alcune comprano i gratta e vinci...

Come calcolare $\mathbb{E}(U(X))$?

Valore atteso di una funzione di una variabile aleatoria

Data una variabile aleatoria X, $Y = g(X)$ è una *nuova* variabile.

Dalla fuzione di densità discreta di X, si può ottenere quella di Y e utilizzarla per il calcolo del valore atteso.

Esempio: Sia X una variabile che assume i valori $\{-2, -1, 0, 1, 2\}$ e sia $Y = X^2$. Chiaramente Y assume i valori $\{0, 1, 4\}$ e

$$\begin{aligned}
\mathbb{P}(Y = 4) &= \mathbb{P}(X = 2) + \mathbb{P}(X = -2) \\
\mathbb{P}(Y = 1) &= \mathbb{P}(X = 1) + \mathbb{P}(X = -1) \\
\mathbb{P}(Y = 0) &= \mathbb{P}(X = 0)
\end{aligned}$$

Dunque

$$\mathbb{E}(Y) = 0 \cdot \mathbb{P}(Y = 0) + 1 \cdot \mathbb{P}(Y = 1) + 4 \cdot \mathbb{P}(Y = 4)$$

Prop. Sia X una variabile aleatoria discreta e g una funzione a valori reali, allora il valore atteso di $g(X)$, se esiste, è dato da

$$\mathbb{E}[g(X)] = \sum_{x_i \in \mathbb{X}} g(x_i)p(x_i).$$

Infatti,

$$\mathbb{E}(Y) = \sum_{y_j} y_j \mathbb{P}(Y = y_j) =$$

$$\sum y_j \mathbb{P}(Y = g(X) = y_j) = \sum_{y_j} y_j \sum_{x_i : g(x_i) = y_j} p_X(x_i) =$$

$$\sum_{y_j} \sum_{x_i : g(x_i) = y_j} y_j p_X(x_i) =$$

$$\sum_{x_i} g(x_i) p_X(x_i).$$

Prop. (*linearità del valore atteso*)

- $\mathbb{E}[aX + b] = a\mathbb{E}(X) + b,$ con a e $b \in \mathbb{R}$
- f e g due funzioni a valori reali

$$\mathbb{E}(f(X) + g(X)) = \mathbb{E}(f(X)) + \mathbb{E}(g(X))$$

Varianza

- Ogni variabile aleatoria X viene completamente caratterizzata dalla distribuzione ma può essere utile sintetizzare le caratteristiche con delle "*misure*".

La prima che abbiamo incontrato è il *valore atteso*. Si considerino

- $X = 0$ con probabilità 1,
- $Y = 1$ e $Y = -1$ con probabilità 0.5
- $Z = 100$ e $Z = -100$ con probabilità 0.5.

Allora

$$\mathbb{E}(X) = \mathbb{E}(Y) = \mathbb{E}(Z) = 0$$

La variabilità dei valori assunti da Y è superiore a quella di X (che è deterministica) e lo stesso dicasi per la viariabilità di Z a confronto con quella di Y.

$\Rightarrow$ Attraverso la varianza si vuole valutare la variabilità con una *misura media degli scarti dal valor medio*.

Def. Sia X una variabile di media $\mathbb{E}(X)$, allora la *varianza* di X è definita

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

che è un indice della *variabilità* della variabile aleatoria.

Sviluppando il quadrato si ricava una formula alternativa per il calcolo delle varianza:

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

$\Rightarrow$ $\mathrm{Var}(X) \geq 0$;

◇ Si verifichi che per a e $b \in \mathbb{R} \Rightarrow \mathrm{Var}(aX + b) = a^2\,\mathrm{Var}(X)$.

- $\mathbb{E}(X^n)$ si chiama momento n-esimo della variabile aleatoria X.
- $\mathrm{Var}(X)$ rappresenta il *momento secondo centrato di X*
- La quantità $\sigma_X = \sqrt{\mathrm{Var}(X)}$ si chiama *deviazione standard* di X.

$\Rightarrow$ (*Variabile aleatoria standardizzata*) Sia $Y = \frac{X - \mathbb{E}(X)}{\sigma_X}$.

◇ Si verifichi che Y ha media nulla e varianza 1.

Variabile aleatoria di Bernoulli

Supponiamo che in un esperimento si sia interessati al verificarsi di un determinato evento che indichiamo come "successo" e che avvenga con probabilità p

- la funzione $X = \mathbb{1}_{\{successo\}}$ è una v.a. discreta che vale
  - 1 (successo) con probabilità p,
  - 0 (insuccesso) con probabilità $1 - p$.

Def. La variabile di Bernoulli di parametro p

$$X \sim Be(p), \qquad 0 < p < 1$$

è una variabile aleatoria discreta che assume i valori $\mathbb{X} = \{0, 1\}$

- $p(0) = 1 - p, \qquad p(1) = p$ (*funzione di densità*).
- $\boxed{\mathbb{E}(X) = p}$ e $\boxed{\mathrm{Var}(X) = p(1 - p)}$.

Variabile aleatoria binomiale

Siamo interessati al *"numero di successi"* realizzati in n prove indipendenti, ognuna con probabilità di successo pari a p.

$\Rightarrow$ La v.a. X che "conta il numero di successi" assume i valori $\{0, 1, \ldots, n\}$.

- $\mathbb{P}(\{X = k\})=\binom{n}{k} p^k (1-p)^{n-k}, \quad k \in \{0, 1, \ldots, n\}$.

Def. La variabile binomiale di parametri n e p,

$$X \sim Bi(n, p) \qquad n \in \mathbb{N} \qquad e \qquad p \in (0; 1)$$

è una variabile aleatoria discreta che assume i valori $\mathbb{X} = \{0, 1, \ldots, n\}$

- $p(k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k \in \{0, 1, \ldots n\}$.
- $\boxed{\mathbb{E}(X) = np}$ e $\boxed{\mathrm{Var}(X) = np(1-p)}$.

Rem. Se $\{X_i, i : 1, \ldots, n\}$ sono le n Bernoulli, con X_i in corrispondenza alla prova i-esima, e $\mathbb{P}(X_i = 1) = p$, allora

$$X = \sum_{i:1}^{n} X_i \qquad X \sim Bi(n, p)$$

Esempio: Analisi del sangue

Un numero N, molto grande di persone, si deve sottoporre ad un'analisi del sangue. Sono state pensate due procedure:

(1) ogni persona viene ananlizzata separatamente $\Rightarrow$ sono necessari N test;

(2) i campioni di k persone vengono raccolti, mescolati ed analizzati insieme. Se il test è negativo, è sufficiente 1 test per k persone. Se è positivo, il test viene ripetuto per ognuna delle k persone $\Rightarrow$ sono necessari $k + 1$ test.

Supponiamo che la probabilità p di essere positivi al test sia p per ognuna delle N persone e che questi eventi siano indipendenti.

(a) Trovare la probabilità che il test sia positivo per il campione formato dal miscuglio relativo a k persone.

(b) Qual è il valore atteso e la varianza del numero di test necessari Y con la seconda procedura? (Supponiamo che N sia divisibile per k.)

Variabile aleatoria di Poisson

Def. La variabile di Poisson di parametro λ,

$$X \sim Poisson(\lambda) \qquad \lambda \in \mathbb{R}^+$$

è una variabile aleatoria discreta che assume i valori $\mathbb{X} = \{0, 1, \ldots\}$ e

- $p(k) = \frac{e^{-\lambda}\lambda^k}{k!}, \qquad k \in \mathbb{X} \qquad$ (*funzione di densità*).

◇ Si verifichi che $p(k)$, $k \in \mathbb{X}$, è una funzione di densità discreta.

◇ $\boxed{\mathbb{E}(X) = \lambda}$ e $\boxed{\mathrm{Var}(X) = \lambda}$.

Poisson come approssimazione di binomiali

Prop. (*Legge degli eventi rari*).
Sia $\{X_n, n \geq 1\}$ una successione di variabili binomiali $X_n \sim Bi(n, p_n)$, tali che

$$\lim_{n\to\infty} p_n = 0 \quad \text{and} \quad \lim_{n\to\infty} np_n = \lambda,$$

allora

$$\lim_{n\to\infty} p_n(k) = \frac{e^{-\lambda}\lambda^k}{k!}, \qquad k : 0, 1, \dots$$

dove $p_n(k) = \mathbb{P}(X_n = k)$.

$\Rightarrow$ Dimostrare con $p_n = \frac{\lambda}{n}$.

Nella pratica questo si traduce "*il numero di successi ottenuti in un grande numero n di prove indipendenti con probabilità di successo pari a p tale che np è moderato si può ben approssimare con una Poisson di parametro* $\lambda = np$."

Questa distribuzione trova applicazioni ovunque dato che molti fenomeni nella realtà sono distribuiti secondo la legge di Poisson: il numero di soldati morti per un calcio di cavallo ogni anno in ogni corpo della cavalleria prussiana, il numero di telefonate che arrivano in un call center al minuto, il numero di omicidi al giorno a Londra etc.

Esempio: Omicidi a Londra

Ogni anno la Polizia metropolitana di Londra registra circa 160 omicidi, e questo dato rimane stabile da 5 anni. Ogni omicidio è indipendente dagli altri. Sembra strano ma questa apparente casualità è in qualche modo predicibile, nel senso che il numero di omicidi giornalieri si adatta molto bene alla legge di Poisson di parametro $\lambda = 160/365 =\approx 0.44$ e quindi si può calcolare la probabilità di

a) *un giorno tranquillo*: un giorno senza omicidi,
b) *a bloody (sun)day*: 3 o più omicidi in un giorno,
c) *una settimana tranquilla*: una settimana senza omicidi.

Geometrica

In una successione di prove indipendenti, ognuna con probabilità di successo pari a p, siamo interessati alla *"prova in cui si realizza il primo successo"*,

$\Rightarrow$ la variabile aleatoria X che indica la "prova del primo successo" assume i valori $\{1, 2 \dots\}$.

- $\mathbb{P}(X = k) = \mathbb{P}(1^o$ successo alla k-esima prova$)=p(1-p)^{k-1}$, $k \in \{1, \dots\}$.

Def. La variabile geometrica di parametro p,

$$X \sim Ge(p) \qquad p \in (0; 1)$$

è una variabile aleatoria discreta che assume i valori $\mathbb{X} = \{1, \dots\}$

- $p(k) = p(1-p)^{k-1}, \qquad k \in \mathbb{X} \qquad$ (*funzione di densità*).

◇ Si verifichi che $p(k)$ con $k \in \mathbb{X}$ è una funzione di densità.

◇ $\boxed{\mathbb{E}(X) = \frac{1}{p}}$ e $\boxed{\mathrm{Var}(X) = \frac{(1-p)}{p^2}}$.

◇ Si determini la legge della variabile che indica la *prova in cui si ottiene l'r-esimo successo* $\Rightarrow$ binomiale negativa(r, p).

Esempio: Collezione Simpson

Ogni volta che tuo fratello più piccolo compra una scatola di cereali trova una delle n figurine della collezione “personaggi dei Simpson”.

Sia X_k il numero di scatole addizionali da comprare, quando si hanno $k-1$ figurine diverse, per ottenerne una nuova. Quindi $X_1 = 1$, X_2 è il numero di scatole da comprare dopo la prima per ottenere la seconda figurina, e così via.

Assumiamo che la probabilità di trovare una figurina piuttosto che un’altra in ogni scatola sia uguale e sia $1/n$.

a) Che distribuzione ha X_k ?

b) Qual è il valore atteso di Y, dove Y è il numero totale di scatole da comprare per finire la collezione?

Per la parte b) si utilizzi $\Rightarrow \quad \mathbb{E}(Y) = \sum_{k:1}^{n} \mathbb{E}(X_k)$

Ipergeometrica

Da un'urna con N palline, di cui m bianche e $N-m$ nere, se ne scelgono n a caso. Sia X il "numero di palline bianche estratte".

Def. La variabile X è una variabile ipergeometrica di parametri, N, m, n

$$X \sim IG(N, m, n) \qquad N, m, n \in \mathbb{N}, m \leq N \quad n \leq N.$$

È una variabile aleatoria discreta che assume i valori

$\mathbb{X} = \{\max(0, n-(N-m)), \ldots, \min(n, m)\}$ con *funzione di densità*

- $p(k) = \dfrac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}}, \qquad k \in \mathbb{X}.$

Ponendo $p = m/N$ si può dimostrare

- $\boxed{\mathbb{E}(X) = np}$ e $\boxed{\operatorname{Var}(X) = np(1-p)\left(1 - \dfrac{n-1}{N-1}\right)}$.

Supponendo che m e N siano molto grandi rispetto a n, è ragionevole pensare che il campionamento senza rimpiazzo non sia molto diverso da quello con rimpiazzo. In effetti si può dimostrare che la ipergeometrica può essere ben approssimata dalla binomiale di parametri $p = m/N$ e n.

Esempio: Indagine

Si supponga che per una indagine sulla povertà, 7 abitazioni vengono campionate casualmente in un condominio di 40. Secondo i dati risulta che in 5 di queste 40 abitazioni viva una famiglia sotto il livello di povertà.

Qual è la probabilità che:

(a) nessuna di queste famiglie venga campionata?

(b) esattamente 4 vengano campionate?

(c) non più di 2 vengono campionate?

(d) almeno 3 vengano campionate?

Esempio: Esperimenti di Cattura-Ricattura

Ci interessa stimare il numero N di animali che abitano una certa regione e, procediamo con un esperimento del tipo *cattura-ricattura*:

- vengono catturati m individui, marcati e rilasciati in natura;
- Successivamente, si catturano n animali; indicando con X il numero di animali marcati; X segue una distribuzione ipergeometrica di parametri N, m e n.
- Se sono stati catturati $X = k$ animali, allora si può ottenere una stima di N massimizzando rispetto ad N la probabilità $P(X = k)$.

Il valore $\widehat{N}$ si chiama stima di massima verosimiglianza (Maximum Likelihood ML) di N.

Si può dimostrare che la stima di massima verosimiglianza è la parte intera di mn/K

$$\Rightarrow \lfloor mn/K \rfloor.$$

Esempio: contiamo le trote

Supponiamo che in un lago ci siano delle trote e che se ne voglia stimare il numero N. $m = 100$ trote vengono catturate, marchiate e liberate. Qualche giorno più tardi, viene prelevato un campione di $n = 50$ trote, X dei quali sono marchiati.

Qual è la stima di massima verosimiglianza (Maximum Likelihood ML) di N se $X = 35$ e se $X = 5$?

- Se $X = 35$ allora la stima di ML di N è $\widehat{N} = \lfloor 100 \cdot 50/35 \rfloor = 142$.
- Se $X = 5$ allora la stima di ML di N è $\widehat{N} = \lfloor 100 \cdot 50/5 \rfloor = 1000$.