

Variabili aleatorie, primi passi ¹

Definizioni principali (riassunto)

1. Variabili Aleatorie Reali

$(\Omega, \mathcal{F})$ spazio di eventi con $\mathcal{F}$ σ -algebra su Ω

Una funzione $X: \Omega \rightarrow \mathbb{R}$ è una v.a. se per ogni $b \in \mathbb{R}$ si ha

$$\{X \leq b\} \equiv \{\omega \in \Omega: -\infty < X(\omega) \leq b\} \in \mathcal{F}$$

Definizione: X è una v.a. *discreta* se $X(\Omega)$ è finito o numerabile, $X(\Omega) = \{x_1, \dots, x_n, \dots\}$

In tal caso basta che $\{X = x_i\} \equiv \{\omega: X(\omega) = x_i\} \in \mathcal{F}$ per ogni x_i .

Dimostrazione: $\{X \leq b\} = \cup_{x_i \leq b} \{X = x_i\}$ per ogni $b \in \mathbb{R}$.

2. Distribuzione di X

Sia $\mathcal{B}$ la più piccola σ -algebra in $\mathbb{R}$ che contiene gli insiemi aperti ($\mathcal{B}$ da Borel)

Una v.a. $X: \Omega \rightarrow \mathbb{R}$ genera la distribuzione su $\mathcal{B}$ definita da

$$P_X(A) := P(X \in A), \quad A \in \mathcal{B}.$$

Per X discreta, ponendo $p(x_i) = P_X(\{x_i\})$, da $\{X \in A\} = \cup_{x_i \in A} \{X = x_i\}$ (unione disgiunta) otteniamo subito

$$P_X(A) = \sum_{x_i \in A} p(x_i) \quad A \in \mathcal{B}.$$

3. Una funzione $X = (X_1, \dots, X_n): \Omega \rightarrow \mathbb{R}^n$ è una v.a. se lo sono le X_i .

Fine. Adesso dettagli

1 Il modello di riferimento

Avremo come modello di riferimento uno spazio di eventi $(\Omega, \mathcal{F})$, con $\mathcal{F}$ σ -algebra su Ω (gli eventi sono gli insiemi in $\mathcal{F}$). La famiglia di *tutti* i sottoinsiemi di Ω può essere “troppo grande” (teorema di Vitali, 1905) quindi in generale $\mathcal{F}$ sarà strettamente più piccolo dell’insieme potenza di Ω .²

¹S Modica 2025.

²Nel caso particolare di $\Omega = \{\omega_1, \dots, \omega_n, \dots\}$ *numerabile* il problema non si pone; in quel caso come abbiamo visto una P su tutto $\mathcal{P}(\Omega)$ resta definita allorché prediamo una sequenza $p_n = P(\{\omega_n\})$ con $p_n \geq 0$, $\sum_n p_n = 1$, quindi possiamo prendere $\mathcal{F} = \mathcal{P}(\Omega)$. Per comodità useremo “numerabile” per dire “finito o numerabile”.

È importante capire che *la $\mathcal{F}$ rappresenta l'informazione che abbiamo su Ω* , nel senso che osservato un qualunque ω , per ogni $E \in \mathcal{F}$ sappiamo se $\omega \in E$ o $\omega \in E^c$; in altre parole gli eventi in $\mathcal{F}$ sono quelli che sappiamo se si sono verificati o no; e a quegli eventi possiamo assegnare probabilità. Per esempio se $\Omega = [0, 1)$ ed $\mathcal{F} = \{\emptyset, \Omega, [0, 1/2), [1/2, 1)\}$, se si verifica un dato ω sappiamo solo se $\omega < 1/2$ oppure $\omega \geq 1/2$. Definendo una probabilità su $\mathcal{F}$ - cioè assegnando una probabilità ad ogni evento in $\mathcal{F}$ - otteniamo uno spazio di probabilità $(\Omega, \mathcal{F}, P)$. La P deve essere definita su $\mathcal{F}$: nell'esempio di sopra non potremmo assegnare una probabilità all'intervallo $[0, 1/3)$ - perché non lo osserviamo!

2 Variabili aleatorie a valori reali

Sia $X: \Omega \rightarrow \mathbb{R}$ una funzione. Torniamo all'esempio $\Omega = [0, 1)$ con $\mathcal{F} = \{\emptyset, \Omega, [0, 1/2), [1/2, 1)\}$, e considera la funzione

$$X(\omega) = \lfloor 3\omega \rfloor = \begin{cases} 0 & 0 \leq \omega < 1/3 \\ 1 & 1/3 \leq \omega < 2/3 \\ 2 & 2/3 \leq \omega < 1. \end{cases}$$

Se l'informazione in nostro possesso è $\mathcal{F}$ non siamo in grado di osservare X ; per esempio $X = 0$ se $\omega \in [0, 1/3)$, ma in quel caso noi sapremo solo che $\omega \in [0, 1/2)$ e quindi non potremo stabilire se $X = 0$ oppure $X = 1$. Questo succede perché $X^{-1}(0) \notin \mathcal{F}$. Dato $(\Omega, \mathcal{F})$ le variabili aleatorie in pratica sono le funzioni *osservabili* con $\mathcal{F}$, cioè quelle tali che per gli insiemi di interesse $A \subseteq \mathbb{R}$ si abbia

$$X^{-1}(A) \equiv \{\omega: X(\omega) \in A\} \in \mathcal{F}.$$

In tal modo saremo in grado non solo di osservare se $X(\omega) \in A$ o meno, ma anche di assegnare probabilità agli eventi $X^{-1}(A)$.

Una funzione $X: \Omega \rightarrow \mathbb{R}$ può rappresentare per esempio un guadagno, nel qual caso una cosa che ci interessa sapere è per esempio con che probabilità $X(\omega) \geq -100$; o può essere un errore di misurazione, e vogliamo sapere con che probabilità $-\epsilon < X(\omega) < \epsilon$. Per questo vogliamo poter assegnare probabilità ad eventi del tipo $X^{-1}(A)$ con $A \subseteq \mathbb{R}$, cioè vogliamo che $X^{-1}(A) \in \mathcal{F}$. Per *tutti* gli insiemi A di $\mathbb{R}$? No, ma almeno per una famiglia di insiemi di $\mathbb{R}$ "ricca abbastanza" da includere tutti gli insiemi di interesse - che contenga per esempio tutti gli insiemi aperti, chiusi, e loro unioni, intersezioni, complementi ecc. È universalmente accettato che una famiglia appropriata di insiemi è *la σ -algebra di Borel*, che è la più piccola σ -algebra in $\mathbb{R}$ che contiene la famiglia $\mathcal{O}$ degli insiemi aperti di $\mathbb{R}$; si indica con $\mathcal{B}(\mathbb{R})$, che abbrevieremo con $\mathcal{B}$. Poiché per ogni $A \in \mathcal{B}$ vogliamo poter calcolare la probabilità che $X(\omega)$ sia in A , cioè $P(X^{-1}(A))$, una variabile aleatoria a valori reali è definita così:

Definizione 1. Una variabile aleatoria su $(\Omega, \mathcal{F})$ a valori reali è una funzione $X: \Omega \rightarrow \mathbb{R}$ tale che $X^{-1}(A) \in \mathcal{F}$ per ogni $A \in \mathcal{B}$.

È dimostrato in Appendice che affinché X sia una v.a. è sufficiente che $X^{-1}(A) \in \mathcal{F}$ per A in un insieme considerevolmente più piccolo di $\mathcal{B}$:

Proposizione 1. La funzione $X: \Omega \rightarrow \mathbb{R}$ è una variabile aleatoria (cioè $X^{-1}(A) \in \mathcal{F}$ per ogni $A \in \mathcal{B}$) se e solo se per ogni $b \in \mathbb{R}$ si ha $X^{-1}(-\infty, b] \in \mathcal{F}$.

Questo giustifica la definizione del libro (Baldi definizione 2.2).

Notazione. Per indicare l'evento $\{\omega: X(\omega) \leq b\} = X^{-1}(-\infty, b]$ scriveremo $\{X \leq b\}$ per semplificare la lettura; l'evento $X^{-1}\{x\} = \{\omega: X(\omega) = x\}$ sarà indicato con $\{X = x\}$. Come osservato nel libro, $\{X \leq b\} \in \mathcal{F}$ implica per esempio $\{X < b\} = \cup_n \{X \leq b - 1/n\} \in \mathcal{F}$, come anche $\{X = x\}$ per ogni $x \in \mathbb{R}$. In generale, l'evento $\{\omega: X(\omega) \in A\} = X^{-1}(A)$ sarà scritto come $\{X \in A\}$. Quindi $P(\{X^{-1}(A)\})$ sarà scritto $P(X \in A)$.

2.1 Variabili discrete (capitolo 2 del libro)

Una funzione $X: \Omega \rightarrow \mathbb{R}$ si dice *discreta* se la sua immagine $X(\Omega) = \{X(\omega): \omega \in \Omega\}$ è numerabile, cioè si ha $X(\Omega) = \{x_1, \dots, x_i, \dots\}$. Come abbiamo visto se X è una v.a. sarà $\{X = x_i\} \in \mathcal{F}$ per ogni i . E per una variabile discreta questo è sufficiente, perché se $\{X = x_i\} \in \mathcal{F} \forall i$ allora per $A \in \mathcal{B}$ avremo $\{X \in A\} = \cup_{x_i \in A} \{X = x_i\} \in \mathcal{F}$.

Esempio (Indicatore di un evento). La v.a. più semplice di tutte è quella che, per un evento $E \in \mathcal{F}$, vale 1 se $\omega \in E$ e zero altrimenti. La indichiamo con $\mathbf{1}_E$:

$$\mathbf{1}_E(\omega) = \begin{cases} 1 & \omega \in E \\ 0 & \omega \notin E. \end{cases}$$

Tornando alla variabile discreta generale nota che la famiglia degli $E_i \equiv \{X = x_i\}$ è una partizione di Ω con $E_i \in \mathcal{F}$, ed $X(\omega) = \sum_i x_i \mathbf{1}_{E_i}(\omega)$ - la funzione che prende valore x_i in E_i . In altre parole una variabile aleatoria discreta è della forma $X = \sum_i x_i \mathbf{1}_{E_i}$ con $\{E_i\}$ partizione di Ω tale che $E_i \in \mathcal{F} \forall i$.

Esempio. Sia $\Omega = \{1, \dots, 6\}^2$ - lancio di due dadi, $\omega = (\omega_1, \omega_2)$ con $\omega_i \in \{1, \dots, 6\}$. La X = "somma dei due numeri" è $X(\omega) = \omega_1 + \omega_2$. La possiamo scrivere come $X = \sum_{i=2}^{12} i \cdot \mathbf{1}_{E_i}$ con $E_i = \{\omega: \omega_1 + \omega_2 = i\}$.

Un'altra v.a., chiamiamola Y , è "numero di volte che esce 6". Indicando con $\lfloor x \rfloor$ il massimo intero minore o uguale ad $x \in \mathbb{R}$ (es. $\lfloor -2.5 \rfloor = -3$, $\lfloor 2.5 \rfloor = 2$) abbiamo $Y(\omega) = \lfloor \omega_1/6 \rfloor + \lfloor \omega_2/6 \rfloor$. Se proprio vogliamo scriverla nella forma $X = \sum_i x_i \mathbf{1}_{E_i}$, ponendo $S_1 = \{6\} \times$

$\{1, \dots, 6\}$ (6 al primo lancio) ed $S_2 = \{1, \dots, 6\} \times \{6\}$ (6 al secondo) abbiamo $Y = \mathbf{1}_{S_1 \setminus S_2} + \mathbf{1}_{S_2 \setminus S_1} + 2 \cdot \mathbf{1}_{S_1 \cap S_2} + 0 \cdot \mathbf{1}_{(S_1 \cup S_2)^c}$.

3 Variabili a valori in $\mathbb{R}^n$

Come vedremo, su uno spazio $(\Omega, \mathcal{F}, P)$ possono essere definite più v.a. che conviene esaminare *congiuntamente*. Per fare un esempio banale considera il modello del lancio di moneta con probabilità p di testa ripetuto n volte, cioè $\Omega = \{0, 1\}^n$, $P\{\omega\} = p^{\sum_{i=1}^n \omega_i} (1-p)^{n-\sum_{i=1}^n \omega_i}$ (qui $\mathcal{F}$ sono tutti gli insiemi perché Ω è finito). Sia $X_1(\omega) = \sum_{i=1}^n \omega_i$ ed $X_2(\omega) = n - \sum_{i=1}^n \omega_i$ (numero di successi e di insuccessi negli n lanci), e poniamo $X = (X_1, X_2)$. Sappiamo allora che $P\{\omega : X(\omega) = (k, n-k)\} = \binom{n}{k} p^k (1-p)^{n-k}$ per $k = 0, \dots, n$.

La condizione che una v.a. $X: \Omega \rightarrow \mathbb{R}^n$ deve soddisfare è la stessa che nel caso di valori in $\mathbb{R}$: devo essere in grado di calcolare $P(X^{-1}(A))$ per una famiglia ricca abbastanza di insiemi di $\mathbb{R}^n$. Come avrai forse indovinato, questa sarà la famiglia $\mathcal{B}^n$, la famiglia degli insiemi di Borel in $\mathbb{R}^n$ che è quella generata dagli aperti $\mathcal{O}$ di $\mathbb{R}^n$. Ponendo $X = (X_1, \dots, X_n)$, analogamente a quanto fatto nel caso unidimensionale si dimostra che questo è equivalente a imporre che per ogni i e b si abbia $\{\omega : X_i(\omega) \leq b\} \in \mathcal{F}$.³

Una v.a. $X: \Omega \rightarrow \mathbb{R}^n$ è discreta se $X(\Omega)$ è numerabile, e la condizione $X = \sum_i x_i \mathbf{1}_{E_i}$ è identica al caso univariato eccetto che adesso $x_i \in X(\Omega) \subseteq \mathbb{R}^n$.

4 Distribuzione di una variabile aleatoria

Come abbiamo accennato all'inizio, data su $(\Omega, \mathcal{F}, P)$ una v.a. $X: \Omega \rightarrow \mathbb{R}$ siamo interessati alla probabilità con cui X sta in certi insiemi (per es. con che probabilità $X \geq 0$). È esattamente per questo che abbiamo imposto la condizione di misurabilità $X^{-1}(A) \in \mathcal{F}$ per ogni $A \in \mathcal{B}$: l'interesse è su $P(X^{-1}(A)) = P(X \in A)$. Questo vale sia per $X: \Omega \rightarrow \mathbb{R}$ che per $X: \Omega \rightarrow \mathbb{R}^n$, e possiamo quindi considerare direttamente quest'ultimo.

Osserva che la funzione $A \mapsto P(X \in A)$ assegna probabilità ad insiemi $A \in \mathcal{B}^n$; è in effetti una probabilità su $\mathcal{B}^n$, che si indica con P_X , e si chiama *distribuzione di X* (o *densità di X*). Dunque la distribuzione di X è la probabilità su $\mathcal{B}^n$ definita da

$$P_X(A) := P(X \in A), \quad A \in \mathcal{B}^n.$$

Verifichiamo che P_X è effettivamente una probabilità: che $P_X(A) \in [0, 1]$ è evidente, e altrettanto lo è che $P_X(\mathbb{R}^n) = 1$ perché $X^{-1}(\mathbb{R}^n) = \Omega$; infine, se $\{A_i\}$ è una famiglia disgiunta

³Di fatto X è una v.a. se e solo se lo sono tutte le X_i . Questo è molto importante per accertarsi che nel quadro generale le cose siano “a posto”, ma per dimostrarlo c'è di mezzo un po' di topologia di $\mathbb{R}^n$, e qui sarebbe troppo lungo entrarci.

in $\mathcal{B}^n$ la famiglia $\{X^{-1}(A_i)\} = \{X \in A_i\}$ è disgiunta in $\mathcal{F}$, sicché $P_X(\cup A_i) = P(X \in \cup A_i) = P(\cup \{X \in A_i\}) = \sum P(X \in A_i) = \sum P_X(A_i)$.

Esempio. Se $\Omega = \{\text{sole, pioggia}\}$ con probabilità 50-50, considera le v.a. $X = \text{“vendo gelati”}$ e $Y = \text{“vendo tè caldo”}$ che danno $(100, -100)$ e $(-100, 100)$ rispettivamente (Euro per fissare le idee). Sono diverse ma hanno la stessa distribuzione su $\mathbb{R}$: $P_X(100) = P_Y(100) = 1/2$ e $P_X(-100) = P_Y(-100) = 1/2$. In pratica danno entrambe fifty-fifty 100 e -100 , quindi se siamo interessati agli Euro saremo indifferenti fra le due se dovessimo sceglierne una. E *qualunque* altra variabile Z definita su *qualunque* altro Ω che abbia distribuzione $P_Z(100) = P_Z(-100) = 1/2$ sarà equivalente ad X ed Y .

5 Distribuzioni di variabili discrete reali

Abbiamo visto che queste hanno immagine $X(\Omega) = \{x_1, \dots, x_n, \dots\}$. Poniamo $p(x_i) = P(\{X = x_i\})$. Poiché $\{X \in A\} = \cup_{x_i \in A} \{X = x_i\}$ (unione disgiunta) abbiamo

$$P_X(A) = \sum_{x_i \in A} P(X = x_i) = \sum_{x_i \in A} p(x_i).$$

La $P_X(A) = \sum_{x_i \in A} p(x_i)$ è notevole: per calcolare $P_X(A)$, per qualunque A , bastano i $p(x_i)$ - cioè i valori di P_X nei punti x_i . Questo caratterizza la semplicità delle variabili discrete (che non per niente sono anche chiamate “semplici”). A questo punto siamo riagganciati al libro, e possiamo passare ai casi più rilevanti di distribuzioni discrete.

5.1 Distribuzioni Multinomiali e Ipergeometriche

Ne abbiamo parlato nel contesto di n estrazioni, con o senza reimbussolamento, da un’urna contenente N “palline” di m tipi diversi; T è l’insieme dei tipi, $\#T = m$ e ci sono N_i palline del tipo $i \in T$. Abbiamo assegnato probabilità a risultati $\vec{k} = (k_i)_{i \in T}$ dove k_i è il numero di volte in cui è estratta una pallina di tipo i - quindi con $\sum_{i \in T} k_i = n$. Adesso vedremo che le distribuzioni viste prima sono in effetti distribuzioni delle variabili che “contano” il numero di palline di tipi diversi estratte.

5.1.1 Distribuzione Multinomiale

Qui l’esperimento consiste in n estrazioni con reimbussolamento; e $\Omega = T^n$, cioè $\omega = (\omega_1, \dots, \omega_n)$ con $\omega_i \in T$, $i = 1, \dots, n$ risultato della i -esima estrazione. Considera la variabile $X: \Omega \rightarrow \mathbb{R}^m$ con coordinate $X_i(\omega) = \#\{j: \omega_j = i\}$, $i \in T$ - cioè il numero di estrazioni in cui viene fuori una pallina di tipo i in ω (per costruzione $\sum_{i=1}^m X_i(\omega) = n$ per ogni ω , dove

$m = \#T$). Allora, con $\rho(i) = N_i/N$,

$$P(X = \vec{k}) = \mathcal{M}_{n,\rho}(\vec{k}) = \binom{n}{\vec{k}} \prod_{i \in T} \rho(i)^{k_i}$$

cioè X ha densità multinomiale.

Distribuzione Binomiale. Se $T = \{0, 1\}$ - lo “schema di Bernoulli”, con $p = \rho(1)$, $\vec{k}$ è della forma $\vec{k} = (k, n - k)$ dove k è il numero di successi. e la probabilità di k successi è $\binom{n}{k} p^k q^{n-k} = B_{n,p}(k)$. Quindi la variabile $S(\omega) = \sum \omega_i$, $\omega \in \{0, 1\}^n$ ha densità binomiale:

$$P(S = k) = \mathcal{B}_{n,p}(k), \quad k = 0, \dots, n.$$

Succintamente scriveremo $S \sim B_{n,p}$ o $S \sim \text{Binom}(n, p)$. Lo schema di Bernoulli è importante perché può essere generalizzato al caso di una sequenza *infinita* di prove, e questa estensione ha un ruolo centrale nei risultati teorici.

La binomiale su *una* prova $B_{1,p}$ si chiama *densità Bernoulliana* e si indica con $\text{Bernoulli}(p)$.

Esempio (Equità di un certo gioco). Ci sono due giocatori A e B ; il primo lancia una moneta equa n volte, il secondo $n + 1$ volte; A vince se i suoi successi (diciamo testa) sono maggiori o uguali a quelli di B . Hanno entrambi la stessa probabilità di vincere?

Il punto importante è definire bene lo spazio di riferimento $(\Omega, \mathcal{F}, P)$ in modo da poter calcolare correttamente le probabilità cercate. Prendiamo $\Omega = \{0, 1\}^{2n+1}$ - dove $2n + 1$ è il numero totale dei lanci; $\mathcal{F}$ sono tutti i suoi sottoinsiemi e P è quella che conosciamo per lo schema di Bernoulli su Ω . Definiamo le variabili $S_A = \sum_{i=1}^n \omega_i$ - numero di successi di A - ed $S_B = \sum_{i=n+1}^{2n+1} \omega_i$ - numero di successi di B . Sappiamo che $S_A \sim \text{Binom}(n, 1/2)$ ed $S_B \sim \text{Binom}(n+1, 1/2)$. L'osservazione cruciale è che dato che la probabilità di successo è $1/2$ anche gli insuccessi avranno le stesse distribuzioni: $n - S_A \sim \text{Binom}(n, 1/2)$ ed $n + 1 - S_B \sim \text{Binom}(n + 1, 1/2)$. Quindi

$$P\{S_A \geq S_B\} = P\{n - S_A \leq n - S_B\} = P\{n - S_A < n + 1 - S_B\} = P\{S_A < S_B\}$$

(l'ultima uguaglianza per quanto appena osservato sugli insuccessi). Dunque sì, le probabilità di successo sono uguali per entrambi i giocatori.

5.1.2 Distribuzione Ipergeometrica

In questo caso non c'è reimbussolamento. Il risultato ω dell'esperimento è un *insieme* $\omega \subseteq \{1, 2, \dots, N\}$ di n elementi (le n palline estratte), sicché $\Omega = \{\omega \subseteq \{1, 2, \dots, N\} : |\omega| = n\}$. Come abbiamo visto gli ω che risultano nel vettore $\vec{k}$ sono quelli tali che $\#\{\omega \cap T(i)\} = k_i, i \in T$; qui $T(i) \subseteq \{1, \dots, N\}$ è l'insieme delle palline di tipo i . In questo caso considera la

variabile $X: \Omega \rightarrow \mathbb{R}^m$ con coordinate $X_i(\omega) = \#\{\omega \cap T(i)\}$, $i \in T$ - cioè il numero di palline di tipo i estratte. Allora, con $\vec{N} = (N_i)_{i \in T}$,

$$P(X = \vec{k}) = \mathcal{H}_{n, \vec{N}}(\vec{k}) = \frac{\prod_{i \in T} \binom{N_i}{k_i}}{\binom{N}{n}}$$

cioè X ha distribuzione ipergeometrica. Analogamente al caso precedente, se $T = \{0, 1\}$, per il numero di “successi” troviamo

$$P\left(\sum_i \omega_i = k\right) = \frac{\binom{N_1}{k} \binom{N-N_1}{n-k}}{\binom{N}{n}}.$$

To do now. Guarda gli esempi 2.6, 2.8 e 2.18 nel libro.

5.2 Distribuzione Geometrica

Considera un processo di Bernoulli - prove ripetute indipendenti con outcome binario con probabilità di successo p - con numero di ripetizioni *infinito*.⁴ Quindi $\Omega = \{0, 1\}^{\mathbb{N}} \ni \omega = (\omega_1, \omega_2, \dots, \omega_n, \dots)$, con $\omega_i \in \{0, 1\} \forall i$: un elemento di Ω è una sequenza infinita di zeri e uni. Su questo spazio definiamo la variabile

$$T(\omega) = \inf \{n: \omega_n = 1\} \in \mathbb{N} \cup \{+\infty\}$$

cioè *il numero di prove al primo successo* (T è spesso interpretato come “istante del primo successo”). Se $\omega_n = 0$ per ogni n è $T(\omega) = +\infty$ (perché $\inf \{\emptyset\} = +\infty$). Il primo successo arriva alla prova k se ci sono $k - 1$ insuccessi seguiti da un successo; dall’indipendenza delle prove segue dunque che

$$P(T = k) = (1 - p)^{k-1}p, \quad k = 1, 2, \dots$$

È questa la *distribuzione geometrica*; scriveremo $T \sim \text{Geom}(p)$ se servirà.⁵ Osserva che

$$P(T < \infty) = \sum_{i=1}^{\infty} (1 - p)^{i-1}p = 1.$$

Qui vediamo per la prima volta la differenza fra “impossibilità” e probabilità zero: $P(T = \infty) = 0$, ma una sequenza di zeri esiste in Ω . Per la cronaca, un’altra dimostrazione che $P(T = \infty) = 0$ è questa: $\{T = \infty\} = \bigcap_k \{T > k\}$ che è una sequenza decrescente; quindi per la continuità di P abbiamo $P(T = \infty) = \lim_{k \rightarrow \infty} P(T > k) = \lim_{k \rightarrow \infty} (1 - p)^k = 0$.

⁴L’esempio 2.9 del libro è in realtà di questo tipo, con $p = 1/6$.

⁵Il libro la chiama “geometrica modificata”, ma è una voce fuori dal coro.

Una variabile simile a T è il *numero di insuccessi* prima del primo successo, che possiamo chiamare $\hat{T}$; ovviamente $T = k \iff \hat{T} = k - 1$ cioè $\hat{T} = T - 1$, quindi

$$P(\hat{T} = k) = P(T = k + 1) = (1 - p)^k p.$$

Osservazione. Lo spazio $\Omega = \{0, 1\}^{\mathbb{N}}$ non lo possiamo studiare veramente. Come forse sapete è in corrispondenza biunivoca con l'intervallo $[0, 1]$, che è un insieme *molto* più complesso di un numerabile. La σ -algebra $\mathcal{F}$ non può più essere quella di tutti i sottoinsiemi di Ω , e sulla P ... accettiamo che sia determinata univocamente dal suo comportamento su sequenze finite dove è definita come sappiamo, cioè: dato $x \in \{0, 1\}^n$ è $P\{\omega: (\omega_1, \dots, \omega_n) = x\} = p^{\sum_{i=1}^n \omega_i} (1 - p)^{n - \sum_{i=1}^n \omega_i}$.

5.2.1 Mancanza di memoria

L'importante proprietà in questione è questa: $P(T = k + m \mid T > k) = P(T = m)$.

La dimostrazione è banale: $P(T = k + m \mid T > k) = (1 - p)^{k+m-1} p / (1 - p)^k = P(T = m)$. L'errore tipico che si fa non essendone consapevoli è pensare, per esempio nel gioco del lotto (esempio 2.11 del libro), cose del tipo “Il 32 non esce da un sacco di tempo, ora deve uscire”. No. Il fatto che non esce da un sacco di tempo non influenza per niente la probabilità che esca in futuro.

Si può dimostrare (dimostrazione accessibile) che soltanto la distribuzione geometrica ha mancanza di memoria.

5.3 Distribuzione di Poisson

Si dice che una variabile X a valori interi non negativi ha *distribuzione di Poisson* se

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k = 0, 1, \dots$$

Si scrive in questo caso $X \sim \text{Poisson}_\lambda$. Osserva che $P(X < \infty) = \sum_{k \geq 0} P(X = k) = 1$. Possiamo guardare sul libro la discussione di questa densità e la sua interpretazione naturale.

La convergenza $\text{Binom}(n, \lambda/n) \rightarrow \text{Poisson}(\lambda)$ se $n \rightarrow \infty$ sfrutta il fatto già visto nel caso multinomiale/ipergeometrica che $\binom{n}{k} \sim \frac{n^k}{k!}$ per $n \rightarrow \infty$ (notazione dal relativo file). Quindi⁶

$$B_{n, \lambda/n}(k) = \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \sim \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^{n-k} \sim e^{-\lambda} \frac{\lambda^k}{k!} = \text{Poisson}_\lambda(k).$$

⁶Nota che qui $p_n = \lambda/n$ quindi $np_n = \lambda$ per ogni n . Il risultato (con dimostrazione quasi invariata) vale anche nel caso in cui $np_n \rightarrow \lambda$ per $n \rightarrow \infty$.

L'utilizzo pratico (vedi in particolare l'esempio 2.15 nel libro) è che una binomiale con n grande e p piccolo si può approssimare con una Poisson con $\lambda = np$. In generale l'approssimazione è accettabile se $np \leq 5$.

Esempio. Esempio simile al 2.15: Un'azienda vende scorze di cannoli in scatole di 250, e in media l'1% sono difettose. Qual'è la probabilità di trovarne meno di 4 in una scatola? Qui siamo in presenza di una binomiale con $n = 250$ e $p = 0.01$ (successo = cannolo difettoso) quindi $np = 2.5$, e possiamo usare Poisson per approssimare la binomiale, assumendo che i pezzi difettosi sono $X \sim \text{Poisson}(2.5)$. E dalle tavole troviamo $P(X \leq 3) = .7576$. Con la binomiale avremmo trovato .7581: l'errore relativo è $|.7576 - .7581| / .7581 = .00066$ - molto basso.

La Poisson emerge in una situazione che si verifica spesso:

Proposizione. Sia $N \sim \text{Poisson}(\lambda)$ e supponi che dato $N = n$ si ha $X \sim \text{Binomial}(n, p)$. Allora la distribuzione di X è $\text{Poisson}(\lambda p)$.

Dimostrazione. Per vedere che $X \sim \text{Poisson}(\lambda p)$ calcoliamo $P(X = k)$ - tenendo presente che dato n il numero di successi sarà $k = 0, \dots, n$:

$$\begin{aligned} P(X = k) &= \sum_{n \geq k} P(N = n)P(X = k | N = n) \\ &= \sum_{n \geq k} e^{-\lambda} \frac{\lambda^n}{n!} \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} = \frac{e^{-\lambda} (\lambda p)^k}{k!} \sum_{n \geq k} \frac{[\lambda(1-p)]^{n-k}}{(n-k)!} \\ &= \frac{e^{-\lambda} (\lambda p)^k}{k!} e^{\lambda(1-p)} = e^{-\lambda p} \frac{(\lambda p)^k}{k!}. \end{aligned} \quad \square$$

Esempio. In radiologia, i raggi X producono rotture di cromosomi e il loro numero è Poisson con parametro λ noto (che dipende da vari fattori); si sa inoltre che ogni rottura guarisce con probabilità p nota, indipendentemente dalle altre. Il risultato di sopra implica che il numero dei cromosomi rotti che on guariscono è Poisson con parametro $\lambda(1-p)$, noto. Un altro esempio concreto è il 2.32 del libro.

6 Un problema di optimal stopping

Questo è un problema famoso, conosciuto come “il problema della segretaria”, la cui analisi richiede condizionare in modo intelligente.⁷ Si applica poi la distribuzione uniforme, e in effetti anche l'approssimazione

$$\int_{x=a}^{b+1} f(x)dx \leq \sum_{i=a}^b f(i) \leq \int_{x=a-1}^b f(x)dx,$$

⁷Ci appoggiamo a Ross *First Course in Probability*

valida per f integrabile *decreciente* (se disegni la capisci).

Il problema è il seguente: ci vengono presentati n oggetti *in sequenza* e ne dobbiamo scegliere uno; può essere un collaboratore, o un'offerta di lavoro, o un marito... li chiameremo "premi". A ogni passo $k < n$ possiamo accettare o rifiutare e passare al successivo. Al passo k sappiamo valutare il valore relativo dei k premi che abbiamo visto ma non conosciamo il valore di quelli che restano - non sappiamo se fra gli $n - k$ che non abbiamo visto ce n'è di migliori. L'obiettivo è massimizzare la probabilità di scegliere *il migliore* degli n premi.

Non risolveremo il problema interamente; considereremo un insieme circoscritto di strategie e fra queste sceglieremo la migliore. La classe di strategie in questione - indicizzata da k - è la seguente: rifiuta i primi k premi, e poi accetta il primo che sia migliore dei k che hai rifiutato (se non ne vengono fuori dovrai scegliere l'ennesimo). Sia $P_k(\text{best})$ la probabilità che il premio migliore venga selezionato quando si impiega questa strategia. Vogliamo scegliere k per massimizzare P_k . Per calcolarlo *condizioniamo su X , la posizione del premio migliore*. Assumeremo che tutti gli $n!$ ordinamenti del valore degli oggetti siano ugualmente probabili, quindi $P(X = i) = 1/n$ per ogni i . Condizionando otteniamo

$$P_k(\text{best}) = \sum_{i=1}^n P_k(\text{best} \mid X = i)P(X = i) = \frac{1}{n} \sum_{i=1}^n P_k(\text{best} \mid X = i).$$

Se il premio migliore è tra i primi k siamo fregati: $\forall i \leq k P_k(\text{best} \mid X = i) = 0$. D'altro canto se $i > k$ il premio migliore sarà selezionato se il migliore tra i primi $i - 1$ premi si trova tra i primi k (altrimenti un premio in posizione $< i$ li batterà e sarà scelto). Ma con $X = i$, data l'assunzione di probabilità uniforme sugli ordinamenti la probabilità di questo evento è semplicemente $k/(i - 1)$. Quindi se $i > k$ avremo $P_k(\text{best} \mid X = i) = k/(i - 1)$.

Mettendo insieme i due casi otteniamo (approssimazione usando la formula di sopra)

$$\begin{aligned} P_k(\text{best}) &= \frac{1}{n} \sum_{i=k+1}^n \frac{k}{i-1} \approx \frac{k}{n} \int_{k+1}^n \frac{1}{x-1} dx = -\frac{k}{n} \log \left(\frac{k}{n-1} \right) \\ &\approx \frac{k}{n} \log \left(\frac{n}{k} \right) \equiv f(k). \end{aligned}$$

A questo punto approssimiamo supponendo k continuo; massimizzando la funzione (concava) $f(x)$, da $f'(x) = \frac{1}{n} \log \left(\frac{n}{x} \right) - \frac{1}{n} = 0$ otteniamo $x = n/e$. Quindi il k ottimo è $\approx n/e$ (un po' più di $1/3$ di n). E la probabilità di selezionare il premio migliore è approssimativamente $f(n/e) = 1/e \approx 0.36788$. Sorprendentemente alto in effetti.

Appendice

Osserviamo intanto che in generale dato un qualunque insieme E e una famiglia di insiemi $\mathcal{G}$ in E , la più piccola σ -algebra in E contenente $\mathcal{G}$ non è altro che l'intersezione di tutte le

σ -algebre di E che contengono $\mathcal{G}$. Si indica con $\sigma(\mathcal{G})$ e si dice *generata* da $\mathcal{G}$.⁸ $\sigma(\mathcal{G})$ contiene tutte le unioni e intersezioni numerabili e complementi di elementi di $\mathcal{G}$, unioni intersezioni e complementi di questi, ecc. ecc. Nel nostro caso $\mathcal{B} = \sigma(\mathcal{O})$.

Può essere istruttivo verificare che $\mathcal{B}$ è generata anche dalla famiglia $\mathcal{C}$ degli insiemi chiusi (dettagli in nota).⁹ La cosa interessante è che la stessa $\mathcal{B}$ è generata da una famiglia molto più piccola della famiglia di tutti i chiusi: la famiglia degli intervalli $(-\infty, b]$ con $b \in \mathbb{R}$:

Proposizione 2. $\mathcal{B} = \sigma(\mathcal{J})$ con $\mathcal{J}$ famiglia degli intervalli $(-\infty, b]$ con $b \in \mathbb{R}$.¹⁰

Dimostrazione. Dimostriamo (i) $\mathcal{B} = \sigma(\mathcal{I})$ con $\mathcal{I}$ famiglia degli intervalli (a, b) ; e (ii) $\sigma(\mathcal{I}) = \sigma(\mathcal{J})$. (i) Da $\mathcal{I} \subseteq \mathcal{B}$ segue che $\mathcal{B}$ è una σ -algebra che contiene $\mathcal{I}$, quindi $\sigma(\mathcal{I}) \subseteq \mathcal{B}$; d'altra parte (ricorda da Analisi) ogni aperto in $\mathbb{R}$ è unione numerabile di intervalli aperti (a, b) , quindi $\mathcal{O} \subseteq \sigma(\mathcal{I})$; e da questo segue $\mathcal{B} = \sigma(\mathcal{O}) \subseteq \sigma(\mathcal{I})$. Quindi $\mathcal{B} = \sigma(\mathcal{I})$. (ii) Possiamo scrivere $(-\infty, b] = \cup_n \cap_m (a_n, b + 1/m)$ con $a_n \rightarrow -\infty$, quindi $\mathcal{J} \subseteq \sigma(\mathcal{I})$ il che implica $\sigma(\mathcal{J}) \subseteq \sigma(\mathcal{I})$; d'altra parte $(a, b] = (-\infty, b] \setminus (-\infty, a] \in \sigma(\mathcal{J})$, e quindi anche $(a, b) = \cup_n (a, b - 1/n] \in \sigma(\mathcal{J})$; ma $\mathcal{I} \subseteq \sigma(\mathcal{J})$ implica $\sigma(\mathcal{I}) \subseteq \sigma(\mathcal{J})$. $\square$

Un altro fatto rilevante è che affinché sia $X^{-1}(A) \in \mathcal{F}$ per ogni $A \in \mathcal{B} = \sigma(\mathcal{J})$ basta che $X^{-1}(J) \in \mathcal{F}$ per ogni $J \in \mathcal{J}$. Questo vale in generale:

Proposizione 3. *Dati due spazi di eventi $(\Omega, \mathcal{F})$ ed $(E, \mathcal{E})$, sia $X : \Omega \rightarrow E$, e supponi che $\mathcal{E} = \sigma(\mathcal{G})$. Se $X^{-1}(A) \in \mathcal{F}$ per ogni $A \in \mathcal{G}$ allora $X^{-1}(A) \in \mathcal{F}$ per ogni $A \in \sigma(\mathcal{G}) = \mathcal{E}$.*

Dimostrazione. La famiglia $\mathcal{A} := \{A \subseteq E : X^{-1}(A) \in \mathcal{F}\}$ è una σ -algebra (verifica!) che contiene $\mathcal{G}$ e quindi anche $\sigma(\mathcal{G})$. $\square$

Corollario (Proposizione 1 nel testo). *Considera i due spazi di eventi $(\Omega, \mathcal{F})$ ed $(\mathbb{R}, \mathcal{B})$ e la funzione $X : \Omega \rightarrow \mathbb{R}$. X è una variabile aleatoria (cioè $X^{-1}(A) \in \mathcal{F}$ per ogni $A \in \mathcal{B}$) se per ogni $b \in \mathbb{R}$ si ha $X^{-1}((-\infty, b]) \equiv \{\omega \in \Omega : -\infty < X(\omega) \leq b\} \in \mathcal{F}$.*

⁸Questa intersezione non è vuota perché contiene l'insieme potenza di E ; è facile verificare che è una σ -algebra; e per costruzione è la più piccola che contiene $\mathcal{G}$.

⁹Un chiuso è complemento di un aperto quindi $\mathcal{C} \subseteq \sigma(\mathcal{O})$ che implica $\sigma(\mathcal{C}) \subseteq \sigma(\mathcal{O})$; e analogamente poiché un aperto è complemento di un chiuso $\mathcal{O} \subseteq \sigma(\mathcal{C})$ che implica $\sigma(\mathcal{O}) \subseteq \sigma(\mathcal{C})$.

¹⁰in realtà basta $b \in \mathbb{Q}$, ma non importa