Appunti di Calcolo delle probabilità

Riccardo Lo Iacono

Dipartimento di Matematica & Informatica
Università degli studi di Palermo
Sicilia
a.a. 2023-2024

Indice.

– 1 – Introduzione alla probabilità: storia e concetti base.

Storicamente il prima ad interessarsi a quella che sarebbe poi divenuto lo studio della probabilità fu il cavaliere *De Mere*. Questi accanito scommettitore, suppose che vi fosse un legame tra la frequenza con cui uscisse un certo risultato e la sua probabilità.

Esempio: Come esempio della logica di De Mere si consideri quanto segue: siano

$A =$ "Lanciando un dado a sei facce 4 volte, esce almeno un 6"

$B =$ "Lanciando due dadi a sei facce 24 volte, esce almeno una coppia (6,6)"

ci si chiede: $\Pr(A) = \Pr(B)$?

Secondo De Mere si avrebbe

$$\Pr(A) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{2}{3}$$
$$\Pr(B) = \underbrace{\frac{1}{36} + \frac{1}{36} + \cdots + \frac{1}{36}}_{\times 36} = \frac{2}{3} \implies \Pr(A) = \Pr(B)$$

Si osserva però che una valutazione più "corretta" è ottenuta sottraendo a tutti casi possibili, quelli a sfavore di A e B rispettivamente, da cui

$$\Pr(A) = \frac{6^4 - 5^4}{6^4} \approx 0.518$$
$$\Pr(B) = \frac{36^{24} - 35^{24}}{36^{24}} \approx 0.491$$

Concludendo dunque che $\Pr(A) > \Pr(B)$.

Introducendo un po di formalismo, quelli che nell'esempio di sopra sono stati indicati con A e B prendono il nome di eventi. Più in generale, fissato un certo spazio campionario Ω, dicasi ogni suo sottoinsieme E evento. Cioè un evento descrive uno dei possibili esiti di un esperimento. Casi particolari di evento sono l'evento certo Ω e quello impossibile $\varnothing$. In fine dato E un evento, si definisce $|E|$ il suo indicatore e si ha

$$|E| = \begin{cases} 1, & \text{se } E \text{ è vero;} \\ 0, & \text{se } E \text{ è falso.} \end{cases}$$

Dove con "vero" e "falso" ci si riferisce al verificarsi o meno di E.

Nota. la definizione di probabilità data da De Mere prende il nome di *criterio classico*.

– 1.1 – Partizione dell'evento certo.

Siano $E_1, \ldots, E_n$ una famiglia di eventi. Allora se

1. $\forall i, j, i \neq j \qquad E_i \wedge E_j = \varnothing$
2. $\bigcup_{i=1}^{n} (E_i) = \Omega$

si dirà che $E_1, \ldots, E_n$ formano una partizione di Ω.

– 2 – Impostazioni della probabilità.

Quella dell'impostazione classica, non è l'unica impostazione della probabilità esistente. Di fatti in aggiunta a quella sinora vista, si hanno le impostazioni *assiomatica, frequentista* e *soggettiva*, analizzate in maggior dettaglio nelle sezioni a seguire.

– 2.1 – Impostazione assiomatica.

L'impostazione assiomatica da una definizione rigorosamente matematica di probabilità, a seguito riportata.
Definizione: dato (Ω, A) uno spazio di misura: ossia Ω è uno spazio con un'algebra (o una $\sigma - algebra$) e A una sua sottofamiglia di eventi; dicasi che una funzione $\Pr : A \to [0, 1]$, dicasi probabilità se:

1. $\forall E \in A, \Pr(E) \geq 0$;
2. $\Pr(\Omega) = 1$;
3. $\Pr(E_1 \vee E_2) = \Pr(E_1) + \Pr(E_2), \forall E_1, E_2 \in A : E_1 E_2 = \varnothing$

Nel caso in cui A sia infinito, il punto 3 diventa

$$\Pr\left(\bigvee_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} \Pr(E_i)$$

– 2.2 – Impostazione frequentista.

Come evidente dall'esempio di De Mere, vi è una naturale tendenza ad associare la probabilità di un evento alla frequenza con cui lo stesso avviene. Tale concezione di probabilità è detta *impostazione frequentista.*
Definizione: considerata una successione di prove indipendenti, ripetute sotto le stesse condizioni, posto E un evento e f_N la frequenza di successo entro relativamente le prime N prove, si pone

$$\Pr(E) = \lim_{N \to \infty} f_N$$

– 2.3 – Impostazione soggettiva.

Le impostazioni sinora viste, sono generalizzabili nella *teoria soggettiva.* Con tale impostazione si ha una divisione rigorosa tra gli aspetti del certo e del probabile. Più in generale si ha quanto segue.
Definizione: sia E un evento, la probabilità $\Pr(E) = p$ rappresenta il grado di fiducia che un individuo ha del verificarsi di E.

– 2.3.1 – Criterio della scommessa.

Da quanto detto precedentemente, sia supposto $\Pr(E) = p$ per un qualche evento E, si può intendere p come la quota che, in un ipotetica scommessa, un individuo è disposto a scommettere. Cioè

$$\text{si paga } p \text{ e si riceve } \begin{cases} 1, & \text{se } E \text{ vero }, \\ 0, & \text{altrimenti} \end{cases}$$

più in generale, $\forall S \in \mathbb{R}, S \neq 0$

$$\text{si paga } pS \text{ e si riceve } \begin{cases} S, & \text{se si verifica } S, \\ 0, & \text{altrimenti }. \end{cases}$$

– 2.3.2 – Guadagno e criterio di coerenza.

Definizione: sia $\Pr(E) = p$ la probabilità che un individuo attribuisce al verificarsi di E. Dicasi la seguente quantità

$$\mathcal{G} = S\,|E| - pS = S(|E| - p)$$

guadagno.

In generale, affinché una scommessa sia sensata, questa deve essere coerente, cioè: l'assegnazione di p non deve portare ad una perdita certa, e ciò si verifica se e solo se

$$\min\{\mathcal{G}\}\max\{\mathcal{G}\} \leq 0, \forall S \in \mathbb{R}$$

– 3 – Eventi e probabilita condizionate.

Definizione: siano E e $H, H \neq 0$ eventi. Dicasi evento condizionato il seguente ente logico.

$$E \mid H = \begin{cases} \text{vero, se } E \text{ e } H \text{ veri ,} \\ \text{falso, se } E^C \text{ e } H \text{ veri ,} \\ \text{indeterminato, se } H^C \text{ vero.} \end{cases}$$

Banalmente se $H = \Omega$, si ha $E \mid H = E \mid \Omega = E$. Vale inoltre

$$E \mid H = (E \wedge \Omega) \mid H = \left[(E \wedge H) \vee (E \wedge H^C)\right] \mid H = (E \wedge H) \mid H$$

Applicando dunque il criterio di scommessa si ha che

$$\Pr(E \mid H) = \begin{cases} S, \text{ se } EH \text{ vero ,} \\ 0, \text{ se } E^C H \text{ vero} \\ pS, \text{ se } H \text{ falso.} \end{cases}$$

da cui, considerando il guadagno si ha

$$\mathcal{G} = S\,|EH| - pS\,|H| = S\,|H|\,(|E| - p)$$

posto dunque $G = \{S(1-p), -ps\}$, si deve verificare

$$\min\{G\}\max\{G\} \leq 0, \forall S \neq 0$$
$$\implies S^2(1-p)(-p) \leq 0 \iff p \in [0,\, 1]$$

affinche la condizione di coerenza sia rispettata.

– 3.1 – Teorema delle probabilità composte.

Teorema 3.1. *Dati due eventi E e $H, H \neq 0$, siano $(\Pr(H), \Pr(E \mid H), \Pr(EH))$ valutazioni di probabilità sugli eventi $\{H, E \wedge H, E \mid H\}$, allora se*

$$\Pr(EH) = \frac{\Pr(E \mid H)}{\Pr(H)}$$

l'assegnazione risulta essere coerente.

Corollario 3.1.1. *Se $\Pr(H) \neq 0$, allora*

$$\Pr(E \mid H) = \frac{\Pr(EH)}{\Pr(H)}$$

– 3.2 – Formula di disintegrazione.

Il *Teorema 3.1* può essere generalizzato a n eventi, come segue. Siano $\{H_1, \cdots, H_n\}$ una partizione di Ω e sia E un evento. Allora

$$\Pr(E) = \sum_i \Pr(E \mid H_i) \Pr(H_i)$$

– 3.3 – Teorema di Bayes.

Dalla formula di disintegrazione segue il seguente risultato, noto come *Teorema di Bayes*.
Teorema di Bayes. Siano E, H due eventi. Allora

$$\Pr(H \mid E) = \frac{\Pr(H)\Pr(E \mid H)}{\Pr(E)}$$

Questi, nella sua versione più generale stabilisce quanto segue.
Teorema di Bayes (generalizzato). Siano $\{H_1, \ldots, H_n\}$ un partizione di Ω, sia E un evento. Si ha allora

$$\Pr(H_i \mid E) = \frac{\Pr(H_i)\Pr(E \mid H_i)}{\Pr(E \mid H_1)\Pr(H_1) + \cdots + \Pr(E \mid H_n)\Pr(H_n)}$$

– 3.3.1 – Indipendenza stocastica.

Definizione: dati due eventi E e $H, H \neq 0$, si dirà che questi sono stocasticamente indipendenti se

$$\Pr(E \mid H) = \Pr(E)$$

da cui

$$\Pr(EH) = \Pr(E)\Pr(H)$$

– 4 – Numeri aleatori.

Dato Ω uno spazio di probabilità, un numero aleatorio può essere inteso come una funzione a variabili reali definita su Ω stesso. Sia dunque X un numero aleatorio, e sia $x \in \mathbb{R}$ un valore assumibile da X. Allora posto considerato l'evento $(X = x)$, a questi si associa la probabilita $\Pr(X = x)$.

In genere, si distinguono due "classi" di numeri aleatori:

- **discreti:** se il numero di valori assumibili dal numero aleatorio è finito o al più numerabile;
- **continuo:** se non discreto.

Osservazione. banalmente, si deve verificare che

$$\sum_{i=1}^{n} \Pr(X = x_i) = \sum_{i=1}^{n} p_i = 1$$

affinche l'assegnazione di probabilità risulti coerente.

– 4.1 – Numeri aleatori semplici.

Siano $E_1, \ldots, E_n$ e siano $\alpha_1, \ldots, \alpha_n$ numeri reali. Si dice che

$$X = \alpha_1 |E_1| + \cdots + \alpha_n |E_n|$$

è un numero aleatorio semplice.

Per come è costruito, il dominio di X non è definito ma lo è il codominio, codominio che può essere stabilito considerando i costituenti dell'evento. Nello specifico, se tutti gli eventi sono indipendenti, allora X può assumere 2^n valori.

– 4.1.1 – Distribuzione ipergeometrica.

Si assuma di possedere un urna con N palline, di cui pN bianche e qN nere sono note, con $qN+pN = N$. E da questa effettuare n estrazioni con restituzione. Sia posto E_i l'evento "l'i-esima pallina estratta è bianca' e sia

$$X = \sum_{i=1}^{n} |E_i|.$$

Si ha che

$$\Pr(E_i) = \frac{\binom{N-1}{pN-1}}{\binom{N}{pN}} = p$$

Cioè gli eventi sono equiprobabili.

Sia ora considerato il generico costituente di X, si ha

$$\Pr\left(E_{i_1} E_{i_2} \cdots E_{i_h} E_{i_{h+1}}^C \cdots E_{i_n}^C\right) = \cdots = \frac{\binom{N-n}{pN-h}}{\binom{N}{pN}}$$

da cui dunque

$$(X = h) = \bigvee_{\{i_1, \cdots, i_h\} \subseteq \{1, \cdots, n\}} E_{i_1} E_{i_2} \cdots E_{i_h} E_{i_{h+1}}^C \cdots E_{i_n}^C$$

implicando che

$$\Pr(X = h) = \frac{\binom{N-n}{pN-h}\binom{n}{h}}{\binom{N}{pN}}$$

poiché esistono $\binom{n}{h}$ modi per ottenere $(X = h)$.

– 4.1.2 – Distribuzione binomiale.

Siano $E_1, \ldots, E_n$ eventi indipendenti ed equiprobabili, con probabilità p. Sia $q = 1 - p$ e sia X un numero aleatorio che rappresenta il numero di successi alle n prove. Cioè

$$X = |E_1| + \cdots + |E_n|$$

Banalmente X assume valori in $\{0, \ldots, n\}$. Per stabilire la distribuzione di X è necessario calcolare $\Pr(X = h), \forall h \in \{0, \ldots, n\}$. Risulta dunque più conveniente considerare il numero di costituenti a favore ad esso. Segue che

$$\Pr(X = 0) = \Pr\left(E_1^C E_2^C \cdots E_n^C\right) = \Pr\left(E_1^C\right)\Pr\left(E_2^C\right)\cdots\Pr\left(E_n^C\right) = (1-p)^n$$
$$\Pr(X = 1) = \Pr\left(E_1 E_2^C \cdots E_n^C\right) = \Pr(E_1)\Pr\left(E_2^C\right)\cdots + \Pr\left(E_n^C\right) = p(1-p)^n$$

Si dimostra banalmente, che in generale vale

$$\Pr(X = h) = \binom{n}{h} p^h q^{n-h}$$

Dicasi che X ha distribuzione binomiale. In simboli $X \sim B_{n,p}$

– 4.1.3 – Mistura di binomiali.

Si assuma di possedere un urna con N palline, di cui r bianche, con r non noto. E da questa effettuare n estrazioni con restituzione. Siano ora $\{H_1, \ldots, H_n\}$ una partizione dell'evento certo, con

$$H_r = \text{ “Ci sono } r \text{ palline bianche”}$$

Sia $E_i =$“l'i-esima pallina estratta è bianca', per $i \in \{1, \cdots, n\}$. Si verifica che

$$\Pr(E_i \mid H_r) = \frac{r}{N}, \qquad \forall i \in \{1, \cdots, n\}$$

Da ciò

$$\begin{aligned}\Pr(E_i) &= \sum_{r=0}^{N} \Pr(E_i \mid H_r)\Pr(H_r) \\ &= \sum_{r=0}^{N} r/N \Pr(H_r)\end{aligned}$$

Da cui risulta $\Pr(E_i)$ è equiprobabile, per ogni i.

Sia ora posto X il numero aleatorio di palline estratte alle n prove. Si ha

$$\Pr(X = h \mid H_r) = \binom{n}{h}\left(\frac{r}{N}\right)^h\left(\frac{N-h}{N}\right)^{n-h}$$

da cui

$$\Pr(X = h) = \sum_{r=0}^{N}\binom{n}{h}\left(\frac{r}{N}\right)^h\left(\frac{N-h}{N}\right)^{n-h}\Pr(H_r) \tag{1}$$

Dicasi che X è una mistura di binomiali.

Nota. nel caso si effettuino estrazioni senza restituzione, la (1) diventa

$$\Pr(X = h) = \sum_{r=0}^{N}\frac{\binom{n}{h}\left(\frac{r}{N}\right)\left(\frac{N-h}{N}\right)}{\binom{N}{r}}\Pr(H_r)$$

Si dirà in questo caso che X è una mistura di iper-geometriche.

– 4.2 – Numeri aleatori discreti.

Come anticipato, se X è un numero aleatorio è il numero di valori x_i che questi può assumere è finito, o al più numerabile, si dice che X è un numero aleatorio discreto.

Sia $p_i = \Pr(X = x_i), \forall i \in \mathbb{N}$ la distribuzione di probabilità di X, sia supposto anche che

$$\sum_{i=1}^{\infty} p_i = 1 \quad \text{e} \quad \sum_{i=0}^{\infty} |x_i|\, p_i < \infty$$

allora si può definire il *valore atteso* di X, come

$$\mathbb{E}(X) = \sum_{i=1}^{\infty} x_i p_i = \mu$$

Sia ora assunto che μ esista finito, e $\mathbb{E}\left[(X-\mu)^2\right] < \infty$, si definire varianza quest'ultima quantità. Cioè

$$\mathrm{Var}(X) = \mathbb{E}\left[(X-\mu)^2\right] = \sum_{i=1}^{\infty} (X-\mu)^2 p_i$$

In fine, si definisce

$$F(x) = \Pr(X \leq x)$$

funzione di ripartizione di X. Se X è discreto si ha che

$$F(x) = \Pr(X \leq x) = \sum_{x_i \leq x} p_i, \quad x \in \mathbb{R}$$

– 4.2.1 – Distribuzione geometrica.

Definizione: dato X un numero aleatorio discreto, questi si dice avere distribuzione geometrica di parametro $p \in [0,\, 1]$ se

$$\Pr(X = n) = (1-p)^{n-1} p, \forall n \in \mathbb{N}$$

Più in generale, siano $E_1, \ldots, E_n$ eventi equiprobabili e indipendenti. Sia

$$X = \text{ “Numero di prove fino al primo successo”}$$

ossia $X = \min\{n : |E_n| = 1\}$. Segue che se X è discreto si ha

$$(X = n) = \begin{cases} E_1, n = 1 \\ {E^C}_1, \ldots, {E^C}_{n-1} E_n, \forall n \geq 2 \end{cases}$$

da cui

$$\Pr(X = n) = (1-p)^{n-1} p, \forall n \in \mathbb{N}$$

Considerando valore atteso e varianza si ha che, poiché $X \in \mathbb{N}$ e nello specifico $X = |X > 0| + |X > 1| + \cdots$, segue, posto $(1-p) = q$, che

$$\mathbb{E}(X) = \sum_{n=0}^{\infty} \Pr(X > n) = 1 + q + q^2 + \cdots = \sum_{i=0}^{\infty} q_i = \frac{1}{1-q}$$

Per calcolare la varianza, sia considerato X^2, poiché

$$X^2 = |X > 0| + 3\,|X > 1| + \cdots$$

segue $\mathbb{E}\left[X^2\right] = 1 + 3(1-p) + \cdots = \cdots = \frac{2-p}{p^2}$, da cui

$$\mathrm{Var}(X) = \mathbb{E}\left[(X-\mu)^2\right] = \frac{1-p}{p^2}$$

– 4.2.2 – Distribuzione di Poisson.

Sia X un numero aleatorio discreto, questi si dice avere distribuzione di Poisson se

1. $X \in \mathbb{N}_0 \implies X \in \{0, 1, \cdots, n, \cdots\}$
2. $\Pr(X = n) = p_n = \frac{\lambda^n}{n!} e^{-\lambda}, \forall n \in \mathbb{N}_0$

Si dimostra che

$$\mathbb{E}(X) = \mathrm{Var}(X) = \lambda$$

Nota. Tale distribuzione permette di approssimare una binomiale.

– 4.2.2.1 – Proprietà di assenza di memoria.

Definizione: sia X un numero aleatorio discreto positivo. Si dice che X gode di assenza di memoria se

$$\Pr(X > x_0 + x \mid X > x_0) = \Pr(X > x), \forall x_0, x \in \mathbb{R}_0^+$$

Osservazione. Si dimostra che se X ha distribuzione di Poisson, allora X gode di assenza di memoria.

– 4.2.3 – Distribuzione di Pascal e binomiale inversa.

Siano $E_1, \ldots, E_n$ eventi indipendenti ed equiprobabili. Sia $\Pr(E_i) = p_i$ e sia T_k il numero di prove fino al k-esimo successo. Si ha, posto $(1 - p) = q$, che

$$\Pr(T_k = n) = \binom{n-1}{k-1} p^k q^{n-k}$$

Dicasi che T_k ha distribuzione di Pascal di parametri p e k.

Sia $Z_k = T_k - k$ il numero di insuccessi prima dei k successi. Si ha

$$\Pr(Z_k = h) = \Pr(T_k = k + h) = \binom{k+h-1}{k-1} p^k q^h$$

Dicasi che Z_k ha distribuzione binomiale inversa di parametri k e p.

– 4.2.4 – Disuguaglianza di Markov.

Teorema 4.1. *Dato un numero aleatorio discreto non-negativo X con $\mathbb{E}(X) < \infty$ e un $\alpha > 0 \in \mathbb{R}$, si ha*

$$\Pr(X \geq \alpha) = \frac{\mathbb{E}(X)}{\alpha}$$

Dimostrazione: poiché $X \geq 0$, segue che $\Pr(X \geq 0) = 1$. Posto ora $\Pr(X \geq x_n) = p_n$, segue

$$\begin{aligned} \alpha \Pr(X \geq \alpha) &= \alpha \sum_{x_n \,:\, x_n \geq \alpha} p_n \\ &\leq \sum_{x_n \,:\, x_n \geq \alpha} x_n p_n \\ &\leq \sum_{x_n} x_n P_n = \mathbb{E}(X) \end{aligned}$$

da cui dunque

$$\Pr(X \geq \alpha) = \frac{\mathbb{E}(X)}{\alpha}$$

– 4.2.5 – Disuguaglianza di Chebychev.

Teorema 4.2. *Dato un numero aleatorio discreto* X *con* $\mathbb{E}(X) = \mu < \infty$ *e* $\mathrm{Var}(X) < \infty$*, si ha*

$$\Pr(|X - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}, \forall \varepsilon > 0$$

Dimostrazione: a partire da Markov (*Teorema 4.1*), si consideri $Y = (X - \mu)^2$.
Si ha che $Y \geq 0$ e $\mathbb{E}(Y) = \sigma^2$. Da Markov

$$\Pr(Y \geq \alpha) \leq \frac{\mathbb{E}(Y)}{\alpha}$$

posto allora $\alpha = \varepsilon^2$, segue

$$\Pr\left((X - \mu)^2 \geq \varepsilon^2\right) \leq \frac{\sigma^2}{\varepsilon^2}$$

ma poiché $(X - \mu)^2 \geq \varepsilon^2 \iff |X - \mu| \geq \varepsilon$, segue

$$\Pr(|X - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$$