Aspettative condizionate e disintegrazione

Abbiamo trovato utile usare la disintegrazione per calcolare probabilità: data una partizione $\{A_i\}_{i\geq 1}$ di Ω con $P(A_i) > 0\ \forall i$ sappiamo che $P(A) = \sum_i P(A_i)P(A \mid A_i)$. Ma - parlando di variabili reali, $X\colon \Omega \to \mathbb{R}$ - vorremmo anche poter disintegrare per calcolare valori attesi, scrivendo $E(X) = \sum_i P(A_i)E(X \mid A_i)$.[1] Qui vedremo che lo possiamo fare, ma $E(X \mid A)$ non l'abbiamo ancora definita... Lo dobbiamo fare, altrimenti non sappiamo cosa mettere dentro la somma.

Oltre a confermare la formula di sopra, per Y continue vedremo che analogamente possiamo scrivere $EX = E\left[E(X \mid Y = y)\right]$ (dove $E\left[\cdot\right]$ è rispetto ad Y). Il termine inglese per queste operazioni è "iterated expectations".

Seguono dettagli. L'essenziale è afferrare le definizioni per poterle applicare negli esercizi. Considereremo variabili reali. Dobbiamo anche ricordare che per definizione $\int_A f(x)dx = \int_{\mathbb{R}} 1_A(x)f(x)dx$.

1 Disintegrare rispetto a una variabile Y

Se Y è discreta non c'è niente di nuovo: $E(X \mid Y = y_j)$ è già stata definita, con $A = \{Y = y_j\}$. Consideriamo Y continua. Densità e media condizionata sono definite ponendo

$$f_{X|Y}(x \mid y) = \frac{f(x,y)}{f_Y(y)} \quad ed \quad E(X \mid Y = y) = \int_{\mathbb{R}} x f_{X|Y}(x \mid y)dx.$$

Scriveremo $E(X \mid Y)$ per indicare la variabile aleatoria che prende valore $E(X \mid Y = y)$ per $Y = y$. Anche in questo caso possiamo disintegrare per calcolare EX:

Proposizione. $EX = E\left[E(X \mid Y)\right]$.

Dimostrazione. Nel libro c'è un errore di stampa quindi riscriviamo:

$$\begin{aligned} E\left[E(X \mid Y)\right] &= \int_{\mathbb{R}} E(X \mid Y = y) f_Y(y)dy = \int_{\mathbb{R}} \left[\int_{\mathbb{R}} x f_{X|Y}(x \mid y)dx\right] f_Y(y)dy \\ &= \int_{\mathbb{R}} x \left[\int_{\mathbb{R}} f(x,y)dy\right] dx = EX. \end{aligned}$$

□

Esempio 1. Una bacchetta di lunghezza 1 è spezzata in un punto a caso $Y \sim \mathcal{U}(0,1)$. Dato $Y = y$ è scelto a caso un altro punto di rottura X fra 0 e y. Vogliamo $E(X)$ che usando il risultato di sopra è uguale ad $E(E(X \mid Y))$; quindi serve $E(X \mid Y)$, che si costruisce a partire da $E(X \mid Y = y)$. In questo caso $f_{X|Y}(x \mid y) = \frac{1}{y}1_{(0,y)}$ perché dato y, la X è per costruzione uniforme su $(0,y)$. Quindi $E(X \mid Y = y) = \int_{\mathbb{R}} x\frac{1}{y}1_{(0,y)}dx = \frac{1}{y}\int_0^y x dx = y/2$; perciò $E(X \mid Y) = Y/2$; concludiamo che $EX = E(E(X \mid Y)) = E(Y/2) = 1/4$.

[1] Per esempio nell'esercizio 3.26 di Baldi volevamo ET dove T era il tempo di vita di uno strumento che poteva venire da una linea 1, evento A di probabilità $P(A)$, e allora era $T_1 \sim \text{Esp}(\lambda)$, oppure da una linea 2, evento A^c, nel qual caso era $T_2 \sim \text{Esp}(\mu)$. E ci era venuto di scrivere $ET = P(A)E(T \mid A) + P(A^c)E(T \mid A^c)$, indovinando che $E(T \mid A) = ET_1$ ed $E(T \mid A^c) = ET_2$.

Esempio 2. Problema BT-3.23.

> **Problem 23.** Let the random variables X and Y have a joint PDF which is uniform over the triangle with vertices at $(0,0)$, $(0,1)$, and (1.0).
>
> (a) Find the joint PDF of X and Y.
>
> (b) Find the marginal PDF of Y.
>
> (c) Find the conditional PDF of X given Y.
>
> (d) Find $\mathbf{E}[X \mid Y = y]$, and use the total expectation theorem to find $\mathbf{E}[X]$ in terms of $\mathbf{E}[Y]$.
>
> (e) Use the symmetry of the problem to find the value of $\mathbf{E}[X]$.

La domanda interessante è l'ultima: quanto ci aspettiamo che sia EX? O ragionando in 2D, dove sarà il vettore (EX, EY)? Lo faremo in due modi: il primo è più diretto; il secondo lo facciamo per applicare il risultato $EX = E(E(X \mid Y))$. Ma andiamo con ordine.

Come sappiamo la distribuzione uniforme su un insieme B ha densità $f(x) = 1_B(x)/\ Vol(B)$ cosicché per $A \subseteq B$ è $P(A) = \int_A f = Vol(A)/\ Vol(B)$. Quindi in questo esempio, con B il triangolo dato abbiamo $Vol(B) = 1/2$; perciò la densità congiunta di (X, Y) è $f(x, y) = 2 \cdot 1_B$.

Per calcolare EX basta trovare la marginale $f_X(x)$. Ma $f(x) = 2\int_0^{1-x} dy = 2(1-x)$, da cui $EX = 2\int_0^1 x(1-x)dx = 2\left[x^2/2 - x^3/3\right]_0^1 = 1/3$. Analogamente troviamo $EY = 1/3$; dunque il vettore delle medie si trova nel "il centro" del triangolo, cioè il punto dove si incontrano le tre mediane (centro di massa credo che si chiami).

Adesso facciamo una strada più lunga per calcolare EX, per esercitarci a condizionare. Troviamo prima la marginale $f_Y(y)$; da questa ricaviamo $f_{X|Y}(x) = f(x,y)/f_Y(y)$, che serve per trovare $E(X \mid Y = y)$; e per ultimo applichiamo $EX = E(E(X \mid Y))$.

La marginale su Y è $f_Y(y) = \int_0^{1-y} 2dx = 2(1-y)$. Per la condizionata, per $0 \le y < 1$ abbiamo $f_{X|Y}(x) = \frac{1}{1-y} \cdot 1_{(0,1-y)}(x)$; dobbiamo escludere $y = 1$ perché $f_Y(1) = 0$. A questo punto, per $y < 1$,

$$E(X \mid Y = y) = \int_0^{1-y} x f_{X|Y}(x)dx = \frac{1-y}{2}$$

mentre per $y = 1$ abbiamo $E(X \mid Y = y) = 0$ perché se $y = 1$ deve essere $X = 0$. Da questo ricaviamo

$$EX = E(E(X \mid Y)) = \int_0^1 \frac{1-y}{2} \cdot 2(1-y)dy = \left[-\frac{(1-y)^3}{3}\right]_0^1 = \frac{1}{3}.$$

Esempio 3. Da Ross *First Course in Probability*, Capitolo 7. Sia $U_i \sim \mathcal{U}(0,1)$, $i \ge 1$ una sequenza di uniformi indipendenti. Sia $N = \min\{n\colon \sum_{i=1}^n U_i \ge 1\}$ il numero minimo di variabili che si devono sommare per superare 1. Vogliamo $E(N)$.

In questo problema è interessante che la soluzione viene fuori da quella di un problema più generale: per $x \in [0,1]$ definisci $N(x) = \min\{n\colon \sum_{i=1}^n U_i \ge x\}$ e poni $m(x) = EN(x)$. Nota che

$m(0) = 1$. Condizioniamo adesso su U_1: troviamo $m(x) = \int_0^1 E(N(x) \mid U_1 = y)dy$. Ma

$$E(N(x) \mid U_1 = y) = \begin{cases} 1 & se\ y > x \\ 1 + m(x-y) & se\ y \leq x, \end{cases}$$

e sostituendo otteniamo $m(x) = 1 + \int_0^x m(x-y)dy = 1 + \int_0^x m(u)du$. Sicché, derivando, $m'(x) = m(x)$ cioè $m(x) = ke^x$. Da $m(0) = 1$ segue $k = 1$, perciò concludiamo che $m(x) = e^x$. In particolare $m(1) = e$. In media bisogna sommare circa 3 variabili uniformi per superare 1.

2 Disintegrazione su una partizione di Ω

Le definizioni rilevanti sono le seguenti.

Variabili discrete. Sappiamo che una variabile discreta X che prende valore x_j in E_j con $\{E_j\}$ partizione di Ω si può scrivere come $X = \sum_j x_j 1_{E_j}$, ed $EX = \sum_j x_j P(X = x_j) = \sum_j x_j P(E_j)$. Con $A \subseteq \Omega$ definiamo

$$E(X \mid A) = \sum\nolimits_j x_j P(E_j \mid A).$$

Variabili continue. In questo caso dobbiamo per prima cosa definire la densità condizionale $f_{X|A}(x)$ per eventi condizionanti $A \in \mathcal{F}$. Ricorda che la densità $f(x)$ è definita come una funzione tale che $P(X \in B) = \int_B f(x)dx$ per B insiemi di Borel. Analogamente, definiamo *densità condizionale* una funzione (non-negativa) $f_{X|A}(x)$ tale che

$$P(X \in B \mid A) = \int_B f_{X|A}(x)dx. \tag{1}$$

Nei casi concreti questa si trova facilmente. Nota che $\int_{\mathbb{R}} f_{X|A}(x)dx = 1$ come deve essere (perché $P(X \in \mathbb{R} \mid A) = P(\Omega \cap A)/P(A) = 1$). La $E(X \mid A)$ è definita nel modo naturale:

$$E(X \mid A) = \int_{\mathbb{R}} x f_{X|A}(x)dx.$$

Con le definizioni nero su bianco è abbastanza agevole dimostrare la seguente

Proposizione 4 (Disintegrazione del valore atteso). *Con $\{A_i\}$ partizione di Ω abbiamo $EX = \sum_i P(A_i)E(X \mid A_i)$, per X sia discrete che continue.*

Dimostrazione. Nel caso discreto abbiamo

$$\begin{aligned} \sum\nolimits_i P(A_i)E(X \mid A_i) &= \sum\nolimits_i \Big(\sum\nolimits_j x_j P(E_j \mid A_i)\Big) P(A_i) = \sum\nolimits_j x_j \sum\nolimits_i P(E_j A_i) \\ &= \sum\nolimits_j x_j P(E_j) = E(X). \end{aligned}$$

Per il caso continuo, intanto osserviamo che

$$f(x) = \sum\nolimits_i P(A_i) f_{X|A_i}(x). \tag{2}$$

Verifica: disintegrando probabilità abbiamo $P(X \le x) = \sum_i P(A_i) P(X \le x \mid A_i)$, che usando la 1 equivale a $\int_{-\infty}^{x} f(t)dt = \sum_i P(A_i) \int_{-\infty}^{x} f_{X|A_i}(t)dt$; derivando rispetto ad x viene fuori la (2). Usando questa il resto è diretto:

$$\begin{aligned} EX &= \int_{\mathbb{R}} x f(x) dx = \int_{\mathbb{R}} x \left[\sum\nolimits_i P(A_i) f_{X|A_i}(x) \right] dx = \sum\nolimits_i P(A_i) \int_{\mathbb{R}} x f_{X|A_i}(x) dx \\ &= \sum\nolimits_i P(A_i) E(X \mid A_i) \end{aligned}$$

che è quanto volevamo.[2] □

Esempio 5. (questo è più che altro un rompicapo, serve solo ad applicare le definizioni). Considera 3 lanci di un dado. Siano X il numero dei 6 usciti ed Y il numero degli 1. Nota che $X \sim \text{Binom}_{3,1/6}$ quindi $EX = 1/2$. Conferma il risultato calcolando $\sum_{m=0}^{3} P(Y = m) E(X \mid Y = m)$. Nota che (con ovvia notazione) $(X \mid Y = m) \sim \text{Binom}_{3-m,1/5}$ perché nei rimanenti $3 - m$ lanci ci sono solo 5 possibilità (non ci sono 1). Dettagli in nota.[3]

3 Usare $E(X \mid \{X \in A\})$

Con questo condizionante, siccome $\{X \in B\} \cap \{X \in A\} = \{X \in A \cap B\}$ abbiamo

$$P(X \in B \mid X \in A) = \frac{P(X \in A \cap B)}{P(X \in A)} = \frac{1}{P(X \in A)} \int_{A \cap B} f(x) dx = \frac{1}{P(X \in A)} \int_B f(x) \cdot 1_A(x)\, dx$$

quindi la (1) dà

$$f_{X|\{X \in A\}}(x) = \frac{f(x)}{P(X \in A)} \cdot 1_A(x), \tag{3}$$

e da questo a sua volta ricaviamo

$$E(X \mid X \in A) = \int_{\mathbb{R}} x f_{X|\{X \in A\}}(x) dx = \frac{1}{P(X \in A)} \int_A x f(x) dx.$$

In particolare con $A = [t, \infty)$ abbiamo $\{X \in A\} = \{X > t\}$ quindi la (3) si legge

$$f_{X|\{X > t\}}(x) = \frac{f(x)}{P(X > t)} \cdot 1_{(t,\infty)}(x).$$

[2] Per essere precisi, se la partizione $\{A_i\}$ non è finita assumiamo che la serie dentro l'integrale è assolutamente convergente, nel qual caso è legittimo portare la somma fuori dall'integrale.

[3] $Y \sim \text{Binom}_{3,1/6}$ quindi $EX = \sum_{m=0}^{3} \binom{3}{m} (1/6)^m (5/6)^{3-m} \cdot (3 - m)/5 = 1/2$.

Esempio 6. Sia $X \sim \text{Esp}(\lambda)$. Allora, considerando l'evento $\{X > t\}$, poiché

$$\frac{f(x)}{P(X>t)} = \frac{\lambda e^{-\lambda x}}{e^{-\lambda t}} = f(x-t)$$

ricaviamo

$$\begin{aligned} E(X \mid X>t) &= \frac{1}{P(X>t)} \int_t^\infty x f(x) dx = t + \frac{1}{P(X>t)} \int_t^\infty (x-t) f(x) dx \\ &= t + \int_t^\infty (x-t) f(x-t) d(x) = t + \int_0^\infty z f(z) d(z) = t + EX, \end{aligned}$$

che è un'altra espressione del fatto che l'esponenziale ha mancanza di memoria.

Esempio 7 (The winner's curse). C'è un premio di valore $V \sim \mathcal{U}(0,1)$, e tu puoi offrire $b \geq 0$ prima che il valore V sia realizzato. La regola è: se $b < \frac{2}{3}V$ l'offerta è rigettata; se $b \geq \frac{2}{3}V$ l'offerta è accettata, e guadagni $V - b$. Quanto devi offrire? La chiave per capire il problema qui è che l'evento che condiziona l'accettazione è $V \leq (3/2)b$ - cioè vinci se V è relativamente basso (anche se maggiore di b), e questo ti deve mettere in guardia. Indicando con G il guadagno netto abbiamo (applicando l'esempio (1)

$$\begin{aligned} EG &= P(V > (3/2)b) \cdot 0 + P(V \leq (3/2)b) \cdot E(V - b \mid V \leq (3/2)b) \\ &= \begin{cases} 1 \cdot (1/2 - b) & b \geq 2/3 \\ (3/2)b \cdot [(3/4)b - b] & b < 2/3 \end{cases} < 0 \ \forall b > 0 \end{aligned}$$

sicché in valore atteso il guadagno è negativo per ogni $b > 0$! Meglio non offrire niente.

Esempio 8. Questo è il Problema BT-3.20.[4]

> **Problem 20.** **An absent-minded professor schedules two student appointments for the same time. The appointment durations are independent and exponentially distributed with mean thirty minutes. The first student arrives on time, but the second student arrives five minutes late. What is the expected value of the time between the arrival of the first student and the departure of the second student?**

Indichiamo con T la variabile di cui vogliamo il valore atteso, e con T_1 e T_2 la durata del ricevimento dei due studenti; sappiamo che $T_1, T_2 \sim \text{Esp}(\lambda = 1/30)$. Usando il risultato nell'esempio 6 e la Proposizione 4 troviamo

$$\begin{aligned} ET &= P(T_1 \leq 5) * (5 + ET_2) + P(T_1 > 5) * (E(T_1 \mid T_1 > 5) + ET_2) \\ &= (1 - e^{-5/30}) * 35 + e^{-5/30} * (35 + 30) = 35 + 30 * e^{-5/30} = 60.39. \end{aligned}$$

[4] BT è Bertsekas-Tsitsiklis *Introduction to Probability.*

Esempio 9. L'entità di un danno automobilistico (in migliaia di euro) sia $X \sim \text{Esp}(1/2)$, dunque valore atteso 2. La compagnia applica una franchigia di 3 (mila euro). Vogliamo calcolare $E(X \mid X > 3)$, valore atteso dei danni da pagare.

Intanto $P(X > 3) = e^{-3/2} \approx 22.3\%$, quindi la compagnia *non* pagherà circa l'80% dei sinistri. Con $A = \{X > 3\}$ abbiamo $f_{X|A}(x) = e^{3/2} \cdot \frac{1}{2} e^{-x/2} \cdot 1_A = \frac{1}{2} e^{-(x-3)/2} \cdot 1_A$ quindi (ponendo $x = u+3$)

$$\begin{aligned} E(X \mid X > 3) &= \frac{1}{2} \int_3^{+\infty} x e^{-\frac{1}{2}(x-3)} dx = \frac{1}{2} \int_0^{+\infty} (u+3) e^{-\frac{1}{2}u} du \\ &= \frac{1}{2} \int_0^{+\infty} (u+3) e^{-\frac{1}{2}u} du = EX + 3 = 5. \end{aligned}$$

La compagnia pagherà in media $5 - 3 = 2$ mila euro per incidente rimborsabile. Nota che anche $EX = 2$ - ma con la franchigia la compagnia paga (in valore atteso) 2 solo sul 22% dei sinistri!

Esempio 10. Problema BT-3.28. Calcola la media della variabile X con densità

$$f(x) = \begin{cases} p\lambda e^{-\lambda x} \equiv p \cdot f_\lambda(x) & x \geq 0 \\ (1-p)\lambda e^{\lambda x} \equiv (1-p) \cdot f_\lambda(-x) & x < 0 \end{cases}$$

dove $\lambda > 0$ e $0 < p < 1$, ed $f_\lambda(x) = \lambda e^{-\lambda x}$.

Osserva che $P(X \geq 0) = \int_0^\infty f = p$, quindi le densità condizionali, dalla (3), sono $f_{X|\{X\geq 0\}}(x) = p^{-1} f(x) \cdot 1_{x\geq 0}(x) = f_\lambda(x) 1_{x \geq 0}$ ed $f_{X|\{X<0\}}(x) = f_\lambda(-x) 1_{x<0}$ - cioè le densità di Esp_λ e $-\text{Esp}_\lambda$. Disintegrando sull'insieme $\{X \geq 0\}$ otteniamo allora

$$\begin{aligned} EX &= pE(X \mid X \geq 0) + (1-p)E(X \mid X < 0) \\ &= pE(\text{Esp}_\lambda) + (1-p)E(-\text{Esp}_\lambda) \\ &= p/\lambda - (1-p)/\lambda = (2p-1)/\lambda. \end{aligned}$$

4 $E(X \mid Y)$come stima di X

Lo facciamo dal libro, Sezione 3.11