APPUNTI SULLA PROBABILITÀ CONTINUA

DARIO TREVISAN

Indice

1. Introduzione informale

Vogliamo [1] studiare quantità aleatorie che possono assumere un infinito "continuo" di valori. Ad esempio, la temperatura di una CPU oppure la frequenza di trasmissione di un segnale sono *idealmente* quantità continue. Tuttavia, teorie fisiche moderne direbbero che sono comunque discrete (o meglio quantizzate) oppure più semplicemente saremo sempre limitati a quantità discrete per via della sensibilità degli strumenti o il modo in cui i dati vengono effettivamente raccolti e conservati (precisione di macchina): per noi quindi lo studio di quantità continue non ha quindi lo scopo di rappresentare in modo più realistico fenomeni naturali, ma piuttosto trovarne una "approssimazione" continua che permetta di usare strumenti di calcolo (integrali e derivate).

Per dare un esempio di riferimento in cui una quantità continua è una approssimazione di quantità discrete (o viceversa!), pensiamo alla seguente situazione. Dato $n \geq 1$, supponiamo che X^n sia una variabile aleatoria a valori in

$$E^n := \left\{ 0, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n} \right\} \subseteq [0,1] \subseteq \mathbb{R}.$$

e supponiamo inoltre che, rispetto all'informazione iniziale Ω, la legge di X^n sia uniforme (sugli n punti descritti sopra), ossia

$$P\left(X^n = \frac{i}{n} \middle| \Omega \right) = \frac{1}{n} \quad \text{per ogni } i \in \{0, 1, \ldots, n-1\}.$$

[1] Queste note sono state scritte un po' in fretta, se trovate errori comunicatemelo pure via mail. Grazie intanto a tutti gli studenti che finora hanno segnalato imprecisioni.

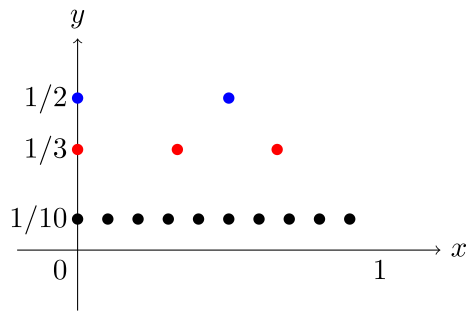

Figura 1. Grafici della densità di X^n per $n = 2$ (blu), $n = 3$ (rosso), $n = 10$, (nero).

Per n via via più grande, se rappresentiamo la densità di X^n (Figura 1), notiamo che una sequenza di punti (n punti) che si "infittisce" (ognuno dista dal precedente/successivo $1/n$) ma contemporaneamente si "appiattiscono" sull'asse delle ascisse.

Questo ci fa pensare che la densità (discreta) di X^n tende a zero, e in effetti dato $x \in [0,1]$, abbiamo

$$P\left(X^n = x \,|\, \Omega\right) = \begin{cases} \frac{1}{n} & \text{se } x = \frac{i}{n} \text{ per qualche } i \in \{0, 1, \ldots, n\}. \\ 0 & \text{altrimenti.} \end{cases}$$

e quindi $\lim_{n\to\infty} P\left(X^n = x \,|\, \Omega = 0\right)$. Sembrerebbe quindi che non ci sia possibilità di definire una variabile "limite" X, perché dovrebbe valere $P(X = x \,|\, \Omega) = 0$.

Un altro ragionamento tuttavia mostra che questo è in realtà possibile. Data una funzione continua $f : [0,1]$, possiamo calcolare il valore atteso di $f(X^n)$ (rispetto alla informazione Ω):

$$\begin{aligned} \mathbb{E}\left[f(X^n)|\Omega\right] &= \sum_{e \in E^n} f(e) P(X^n = e \,|\, \Omega) = \sum_{i=0}^{n-1} f\left(\frac{i}{n}\right) P(X^n = i/n \,|\, \Omega) \\ &= \sum_{i=0}^{n-1} f\left(\frac{i}{n}\right) \frac{1}{n}. \end{aligned}$$

L'espressione che abbiamo trovato ha una interpretazione geometrica molto semplice (se supponiamo per semplicità che f sia positiva): si tratta infatti di valutare f in ciascuno dei punti i/n e moltiplicarne il valore per $1/n$: quindi stiamo calcolando aree di un rettangoli di lati $f(i/n)$ e $1/n$ e sommando per $i \in \{0, 1, \ldots, n\}$ (Figura 2).

Dopo aver visto queste figure, non abbiamo dubbi che, al tendere di n all'infinito, stiamo calcolando l'integrale di Riemann della funzione f nell'intervallo $[0,1]$ (questo vale anche nel caso in cui f assuma valori negativi, perché l'integrale di Riemann calcola aree "con segno").

Abbiamo trovato che

$$\lim_{n\to\infty} \mathbb{E}\left[f(X^n)|\Omega\right] = \int_0^1 f(x) \mathrm{d}x. \tag{1}$$

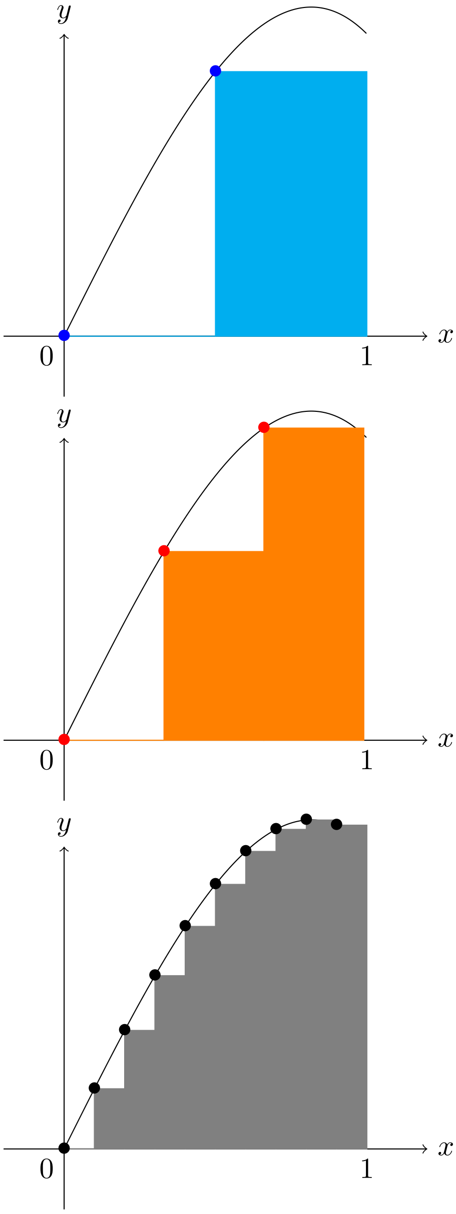

FIGURA 2. $\mathbb{E}\left[f(X^n)|\Omega\right]$, per $f(x) = 2x - x^3$, $n = 2$ (azzurro) $n = 3$ (arancio) ed $n = 10$ (grigio).

Quindi, anche se la densità $P(X^n = x|\Omega)$ tende a zero (e vale $1/n$) quando sommiamo su tutti (gli n) valori, otteniamo nel limite qualcosa di non nullo.

È possibile anche trovare un argomento (sempre intuitivo) che spiega meglio questo passaggio al limite: fissato $x \in [0, 1)$, consideriamo un "piccolo intervallo" $\delta x = [x, x + 1/n)$ e calcoliamo, invece di $P(X^n = x|\Omega)$, la probabilità che X^n cada nell'intervallo δx. Abbiamo

$$P(X^n \in \delta x|\Omega) = \frac{1}{n}$$

se $x \in [0, 1 - 1/n]$ (altrimenti vale 0). Infatti, la *lunghezza* dell'intervallo δx

è uguale a $1/n$ e quindi è tale per cui cade un punto (e uno solo) della forma i/n, ovunque sia $x \in [0, 1 - 1/n]$ (c'è un problema quando $x \in (1 - 1/n, 1]$ perché non vi sono punti lì, ma in compenso sono pochi valori di x). Notiamo quindi che vale

$$P(X^n \in \delta x|\Omega) = \ell(\delta x),$$

dove scriviamo $\ell(I) = b - a$ per la lunghezza di un intervallo $I = [a, b)$. Questo ci fa pensare che, per definire la legge di una ipotetica variabile limite X^∞, possiamo chiedere invece di una condizione su $P(X^\infty = x|\Omega)$ (che abbiamo visto è nullo), una condizione su $P(X^\infty \in \delta x|\Omega)$, dove δx è un (qualunque) piccolo intervallo intorno ad x, di lunghezza $\ell(\delta x)$. Nell'esempio, la condizione sembrerebbe

$$P(X^\infty \in \delta x|\Omega) \approx \ell(\delta x),$$

dove $\approx$ in realtà sta a a significare un passaggio al limite per intervalli di lunghezza che tende a zero:

$$\lim_{\ell(\delta x) \to 0} \frac{P(X^\infty \in \delta x|\Omega)}{\ell(\delta x)} = 1.$$

In analogia con sviluppi più familiari, potremmo pensare che la legge di X^∞ intorno ad ogni punto x presenta un comportamento che può essere ben approssimato "all'ordine zero" con la funzione 0 (infatti vale $P(X^\infty = x|\Omega) = 0$, e al "primo ordine" con la funzione "lunghezza dell'intervallo intorno ad x". In questo modo (sempre non rigoroso!) recuperiamo la validità di (1) direttamente con X^∞: basta infatti decomporre l'intervallo $[0, 1]$ in piccoli intervalli δx disgiunti e calcolare

(2)
$$\begin{aligned}
\mathbb{E}[f(X^\infty)|\Omega] &= \sum_x \mathbb{E}[f(X^\infty)|\Omega \cap \{X^\infty \in \delta x\}]\, P(X^\infty \in \delta x|\Omega) \\
&\approx \sum_x \mathbb{E}[f(X^\infty)|\Omega \cap \{X^\infty \in \delta x\}]\, \ell(\delta x) \\
&\approx \sum_x f(x)\ell(\delta x) \quad (f \text{ continua} \Rightarrow \mathbb{E}[f(X^\infty)|\Omega \cap \{X^\infty \in \delta x\}] \approx f(x)) \\
&\approx \int_0^1 f(x)\mathrm{d}x.
\end{aligned}$$

Ovviamente il nostro esempio era molto semplice, in generale si potrebbe pensare ad altri esempi[2] in cui si trova, per una qualche funzione $\varrho(y|\Omega) : [0, 1] \to [0, \infty)$ la relazione

$$P(Y^\infty \in \delta y|\Omega) \approx \varrho(y|\Omega)\ell(\delta y), \tag{3}$$

e di conseguenza (1) e formalmente, seguendo gli stessi passaggi di (2),

$$\mathbb{E}[f(Y^\infty)|\Omega] = \int_0^1 f(y)\varrho(y|\Omega)\mathrm{d}y. \tag{4}$$

Chiameremo $\varrho(\cdot|\Omega)$ la *densità continua* della variabile Y^∞ (rispetto all'informazione Ω).

[2] ad esempio ponendo $P(Y^n = i/n|\Omega) = 2i/n(n-1)$, per $i \in \{0, 1, \ldots, n-1\}$, al limite si trova $\varrho(y|\Omega) = 2y$.

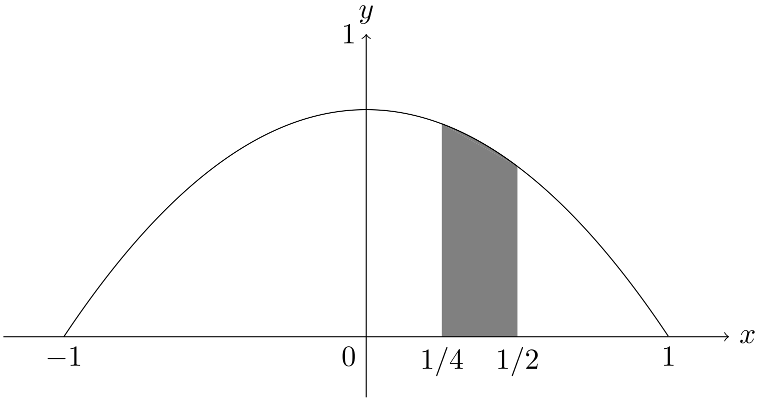

FIGURA 3. Esempio di densità di una variabile continua X con $\varrho(x|I) = \frac{3}{4}(1-x^2)$ e $P(X \in [\frac{1}{4}, \frac{1}{2}]|I) = \int_{\frac{1}{4}}^{\frac{1}{2}} \varrho(x|I)\mathrm{d}x$ (area in grigio).

2. Variabili aleatorie (assolutamente) continue

Diamo ora la definizione rigorosa di una variabile aleatoria continua X a valori in un intervallo $E \subseteq \mathbb{R}$.

Definizione 1. Sia $E \subseteq \mathbb{R}$ un intervallo. Diciamo che una variabile aleatoria $X \in E$ (ossia, nella teoria di Kolmogorov $X : \Omega \to E$) è *assolutamente continua* rispetto all'informazione I, se esiste $\varrho(\cdot|I) : E \mapsto [0, \infty)$ (Riemann) integrabile tale che

$$P(X \in A\,|I) = \int_A \varrho(x|I)\mathrm{d}x \quad \text{per ogni intervallo } A \subseteq E. \tag{5}$$

La funzione $\varrho(\cdot|I)$ è detta *densità* (continua) della variabile X (rispetto all'informazione I).

Osservazione 2 (tipo di intervalli). In realtà si potrebbe dimostrare che l'identità (5) vale per insiemi più complicati (ad esempio unioni di intervalli), e pure che sarebbe sufficiente verificarla o per intervalli chiusi $[a, b]$, o solamente per intervalli aperti (a, b) o solamente intervalli della forma $[a, b)$ o infine del tipo $(a, b]$ (per ogni scelta di a, $b \in \mathbb{R}$, $a \leq b$). Questo perché una importante proprietà delle variabili continue $X \in E$ è che, per ogni $x \in E$,

$$P(X = x|I) = 0.$$

Notiamo che nella definizione data non vi è traccia di "approssimazioni" discrete, ma la condizione approssimativa (3) viene "integrata" e diventa esatta in (5). In termini geometrici, stiamo chiedendo che la probabilità che $X \in A$ sia data dall'area del sottografico della funzione densità $\varrho(\cdot|I)$ nell'intervallo A (Figura 3).

Il termine *assolutamente* si riferisce al fatto che stiamo chiedendo (in un certo senso) che lo sviluppo (3) valga, ma potremmo immaginare (e in realtà esistono, ad esempio certi insiemi "frattali") altre leggi per cui abbiamo sviluppi diversi, ad esempio

$$P(Y \in \delta y|\Omega) \approx \varrho(y|\Omega)\,(\ell(\delta y))^{\alpha}\,,$$

dove $\alpha \in (0,1)$. Noi non tratteremo queste variabili, e quindi useremo in seguito semplicemente il termine "variabili continue".

Osservazione 3 (Notazione). Quando abbiamo a che fare con più di una variabile continua $(X, Y \dots)$, è bene specificare di quale variabile ϱ è densità, scrivendo quindi

$$\varrho(X = \cdot|I), \quad \varrho(Y = \cdot|I)\dots$$

Attenzione però a trattare $=$ come "vera" uguaglianza e cadere in errori del tipo: per $\lambda \neq 0$,

$$\varrho((\lambda X) = x|I) = \varrho\left(X = \frac{x}{\lambda}|I\right) \quad \Leftarrow \text{FALSA in generale.}$$

Infatti vedremo (Lemma 11) che la densità si trasforma in un modo diverso (ma comunque naturale).

Osservazione 4 (Integrale di Riemann). Abbiamo deciso di definire (5) in termini di un integrale di Riemann (eventualmente improprio). Ci sono altre teorie dell'integrazione (Lebesgue) ma per i nostri scopi Riemann è più che sufficiente. Ovviamente bisogna sembre assicurarsi che gli integrali che si scrivono (anche negli esercizi) siano ben definiti!

Osservazione 5 (Proprietà della densità). Abbiamo richiesto nella definizione di $\varrho(\cdot|I)$ che $\varrho(X = x|I) \geq 0$. Questo è naturale, perché le probabilità sono nonnegative.

Notiamo però che non è necessario chiedere $\varrho(\cdot|I) \leq 1$, anzi è facile costruire variabili aleatorie con densità più grande di 1 in alcuni punti.

Una condizione importante (e non scritta nella definizione) da avere sempre ben presente è quella che si ottiene ponendo $A = E$ in (3):

$$1 = P(X \in E|I) = \int_E \varrho(X = x|I)\mathrm{d}x,$$

perché $X \in E$ è un evento certo (abbiamo richiesto $X \in E$).

Vediamo ora come definire valore atteso e varianza di variabili aleatorie continue. Più in generale, definiamo direttamente $\mathbb{E}[f(X)|I]$ dove $X \in E$ è una variabile continua ed $f : E \to \mathbb{R}$ è una funzione (deterministica) continua. Si tratta semplicemente di trasformare (2) in una definizione!

Definizione 6 (Valore atteso)**.** Sia $X \in E$ una variabile continua (rispetto all'informazione I) con densità $\varrho(X = \cdot|I)$, e sia $f : E \to \mathbb{R}$ una funzione continua. Definiamo

$$\mathbb{E}[f(X)|I] = \int_E f(x)\varrho(X = x|I)\mathrm{d}x,$$

purché l'integrale a destra esista (eventualmente come integrale di Riemann improprio).

Dovremo assicurarci in tutti i casi (qualora non fosse ovvio) che l'integrale che definisce $\mathbb{E}[f(X)|I]$ esista. Una condizione sufficiente è

$$\int_E |f(x)|\, \varrho(X = x|I)\mathrm{d}x < \infty.$$

Ponendo $f(x) = x$, otteniamo in particolare

$$\mathbb{E}[X|I] = \int_E x\varrho(X = x|I)\mathrm{d}x.$$

Definizione 7 (Varianza)**.** Sia $X \in E$ una variabile continua (rispetto all'informazione I) con densità $\varrho(X = \cdot|I)$, tale che $\mathbb{E}[X|I]$ sia ben definito. Definiamo allora

$$\mathrm{Var}(X|I) = \mathbb{E}\left[(X - \mathbb{E}[X|I])^2 |I\right] \in [0, +\infty].$$

Come nel caso discreto, non ci sono problemi a definire $\mathrm{Var}(X|I)$ (se $\mathbb{E}[X|I]$ è definita) eventualmente la porremo uguale a $+\infty$. Non è difficile verificare, come nel caso discreto, che vale l'identità (se $\mathbb{E}[X|I]$ è definita)

$$\begin{aligned}\mathrm{Var}(X|I) &= \mathbb{E}\left[X^2|I\right] - (\mathbb{E}[X|I])^2 \\ &= \int_E x^2\varrho(X = x|I)\mathrm{d}x - \left(\int_E x\varrho(X = x|I)\mathrm{d}x\right)^2.\end{aligned}$$

Definizione 8 (Funzioni di ripartizione e sopravvivenza)**.** Sia $X \in E$ una variabile continua (rispetto all'informazione I) con densità $\varrho(X = \cdot|I)$. Definiamo, per $t \in E$, la funzione di ripartizione

$$t \mapsto P(X \leq t|I) = P(X \in (-\infty, t]|I) = \int_{E\cap(-\infty,t]} \varrho(X = x|I)\mathrm{d}x,$$

e la funzione di sopravvivenza

$$t \mapsto P(X > t|I) = P(X \in (t, +\infty)|I) = \int_{E\cap(t,+\infty)} \varrho(X = x|I)\mathrm{d}x.$$

Notiamo che la funzione di ripartizione e di sopravvivenza (come nel caso discreto) sono collegate dall'identità

$$P(X \leq t|I) + P(X > t|I) = 1.$$

A volte è più naturale pensare in termini di funzione di sopravvivenza piuttosto che di ripartizione (ad esempio con i tempi di vita di apparecchiature ecc.)

Lemma 9 (Dalla funzione di ripartizione alla densità)**.** *Se di una variabile $X \in E$ conosciamo la funzione di ripartizione $t \mapsto P(X \leq t|I)$, e osserviamo che si scrive nella forma*

$$P(X \leq t|I) = \int_{E\cap(-\infty,t]} q(x)\mathrm{d}x, \quad \textit{(per ogni } t \in E\textit{)}$$

allora necessariamente X è continua e vale $q(x) = \varrho(X = x|I)$.

In particolare, la densità si ottiene prendendo la *derivata* della funzione di ripartizione:

$$\left(\frac{d}{dt}P(X \leq t|I)\right)\Big|_{t=x} = \varrho(X = x|I).$$

Se invece abbiamo la funzione di sopravvivenza, basta prendere *l'opposto della derivata* (ossia cambiare di segno):

$$-\left(\frac{d}{dt}P(X>t|I)\right)\Big|_{t=x} = \varrho(X=x|I).$$

Questo perché semplicemente $P(X>t|I) = 1 - P(X \le t|I)$.

Dimostrazione. Infatti, se $A = (a,b] \subseteq E$ è un intervallo, consideriamo il sistema di alternative

$$\{X \le a\}, \quad \{a < X \le b\} = \{X \in A\}, \quad \{X > b\}.$$

Di conseguenza

$$\begin{aligned}1 &= P(X \le a|I) + P(X \in (a,b]|I) + P(X > b|I)\\ &= P(X \le a|I) + P(X \in A|I) + (1 - P(X \le b|I)),\end{aligned}$$

e quindi

$$\begin{aligned}P(X \in (a,b]|I) &= P(X \le b|I) - P(X \le a|I)\\ &= \int_{E\cap(-\infty,b]} q(x)\mathrm{d}x - \int_{E\cap(-\infty,a]} q(x)\mathrm{d}x\\ &= \int_{E\cap(a,b]} q(x)\mathrm{d}x.\end{aligned}$$

Ne segue dalla definizione di variabile aleatoria assolutamente continua che $q(x) = \varrho(X=x|I)$. □

Rimandiamo alla Sezione 5 per ulteriori proprietà del valore atteso e della varianza (e delle funzioni di sopravvivenza e ripartizione) di variabili continue. Concludiamo invece la sezione con due risultati generali di "trasformazione" di variabili continue: il primo mostra come si trasforma la legge di $X \in E$ se aggiungiamo l'informazione $X \in B$, dove $B \subseteq \mathbb{R}$ è un intervallo.

Lemma 10. *Sia $X \in E$ una variabile aleatoria continua con densità $\varrho(X = x|I)$, e sia $B \subseteq \mathbb{R}$ un intervallo con $P(X \in B|I) \neq 0$. Allora X è continua anche rispetto a $I \cap \{X \in B\}$, e vale*

$$\varrho(X=x|I\cap\{X\in B\}) = \begin{cases} \frac{\varrho(X=x|I)}{P(X\in B|I)} & \text{se } x \in E\cap B,\\ 0 & \text{altrimenti.}\end{cases} \tag{6}$$

Notiamo che la formula (6) può essere scritta brevemente usando la funzione *indicatrice* dell'intervallo B, definita semplicemente ponendo

$$\chi_B(x) := \begin{cases} 1 & \text{se } x \in B,\\ 0 & \text{altrimenti.}\end{cases}$$

a questo punto la tesi diventa: per $x \in E$,

$$\varrho(X=x|I\cap\{X\in B\}) = \chi_B(x)\frac{\varrho(X=x|I)}{P(X\in B|I)}.$$

Diamone una interpretazione intuitiva, basandoci su (3): se δx è un piccolo intervallo attorno ad $x \in E$, vale per la formula di Bayes

$$\begin{aligned} P(X \in \delta x | I \cap \{X \in B\}) &= P(X \in B | I \cap \{X \in \delta x\}) \frac{P(X \in \delta x | I)}{P(X \in B | I)} \\ &\approx P(X \in B | I \cap \{X \in \delta x\}) \frac{\varrho(X = x | I)\ell(\delta x)}{P(X \in B | I)} \\ &= P(X \in B | I \cap \{X \in \delta x\}) \frac{\varrho(X = x | I)}{P(X \in B | I)} \ell(\delta x). \end{aligned}$$

A questo punto, notiamo che

$$P(X \in B | I \cap \{X \in \delta x\}) \approx \chi_B(x),$$

infatti sarà molto grande (al limite 1) se $x \in B$ mentre se $x \notin B$, sarà piccola (al limite 0). Di conseguenza troviamo la formula cercata.

Dimostrazione rigorosa. Dato un intervallo $A \subseteq E$, possiamo scrivere

$$\begin{aligned} P\left(X \in A | I \cap \{X \in B\}\right) &= \frac{P\left(\{X \in A\} \cap \{X \in B\} | I\right)}{P\left(X \in B | I\right)} \\ &= \frac{P\left(X \in (A \cap B) | I\right)}{P\left(X \in B | I\right)} \\ &= \frac{\int_{A \cap B} \varrho(X = x | I) \mathrm{d}x}{P\left(X \in B | I\right)} \\ &= \int_A \chi_B(x) \frac{\varrho(X = x | I)}{P\left(X \in B | I\right)} \mathrm{d}x. \end{aligned}$$

Ricordando la definizione di densità continua (5), abbiamo dimostrato quindi che X è continua con densità (6). □

Il secondo mostra come si può ottenere la densità di $\lambda X + c$ (se $\lambda \neq 0$, $c \in \mathbb{R}$) dalla densità di X: esistono anche risultati più generali (Teoremi di cambi di variabile), ma noi non ne faremo uso.

Lemma 11. *Sia $X \in \mathbb{R}$ una variabile aleatoria continua, con densità $\varrho(X = \cdot | I)$, siano λ, $c \in \mathbb{R}$, $\lambda \neq 0$. Allora $\lambda X + c$ è una variabile aleatoria continua (rispetto ad I) con densità, per $z \in \mathbb{R}$,*

$$\varrho((\lambda X + c) = z | I) = \varrho\left(X = \frac{z - c}{\lambda} \,\Big|\, I\right) \frac{1}{|\lambda|}.$$

Intuitivamente, l'apparizione del fattore $|\lambda|^{-1}$ si può spiegare nel seguente modo: se X ha una unità di misura, ad esempio, X è una lunghezza misurata in metri, la densità di probabilità $\varrho(X = \cdot | I)$ è misurata in (probabilità)/(metri): questo è evidente dalla (3). Perciò l'effettò di moltiplicare λX può essere pensato come un *cambio di unità di misura* e di conseguenza anche la densità va espressa nelle nuove unità di misura.

Un'altra giustificazione formale è sempre di usare piccoli intervalli δz intorno a $z \in \mathbb{R}$. Se $\lambda X + c$ appartiene all'intervallo $\delta z = [z, z + \ell(\delta z)]$, allora X deve appartenere all'intervallo $\delta x = [\frac{z-c}{\lambda}, \frac{z-c}{\lambda} + \frac{\ell(\delta z)}{\lambda}]$ (se $\lambda > 0$) che è un piccolo intervallo intorno a $x = \frac{z-c}{\lambda}$, di lunghezza $\ell(\delta z)/|\lambda|$. Di conseguenza, usando la (3),

$$P((\lambda X + c) \in \delta z | I) = P(X \in \delta x | I) \approx \varrho \left(X = \frac{z-c}{\lambda} \,\Big|\, I \right) \ell(\delta x)$$
$$\approx \varrho \left(X = \frac{z-c}{\lambda} \,\Big|\, I \right) \frac{1}{|\lambda|} \ell(\delta z).$$

Dimostrazione rigorosa. Supponiamo prima $\lambda > 0$. $A = [a, b] \subseteq \mathbb{R}$ intervallo, calcoliamo

$$\begin{aligned} P\left((\lambda X + c) \in [a, b] | I\right) &= P\left(a \le \lambda X + c \le b | I\right) \\ &= P\left(\frac{a-c}{\lambda} \le X \le \frac{b-c}{\lambda} \Big| I\right) \\ &= \int_{\frac{a-c}{\lambda}}^{\frac{b-c}{\lambda}} \varrho(X = x|I) \mathrm{d}x \quad \text{(perché } X \text{ è continua)} \\ &= \int_a^b \varrho\left(X = \frac{z-c}{\lambda} \Big| I\right) \frac{1}{\lambda} \mathrm{d}z, \end{aligned}$$

dove nell'ultima identità abbiamo usato il cambio di variabile nell'integrale: $z = \lambda x + c$ (e quindi $\mathrm{d}z = \lambda \mathrm{d}x$). Ricordando la definizione di densità continua (5), abbiamo quindi mostrato che $\lambda X + c$ è continua e ne abbiamo dato la densità. Il caso $\lambda < 0$ è analogo, basta però fare attenzione al passaggio

$$P\left(a \le \lambda X + c \le b | I\right) = P\left(\frac{b-c}{\lambda} \le X \le \frac{a-c}{\lambda} | I\right),$$

perché dividiamo per $\lambda < 0$. □

3. Esempi

In questa sezione studiamo due esempi fondamentali di densità continue. In realtà si tratta di *famiglie* di densità in cui sono presenti alcuni parametri di cui discuteremo il significato (come nel caso delle variabili discrete: uniformi, bernoulli, binomiali, geometriche, ecc.)

3.1. **Variabili uniformi.**

Dato un intervallo $E = [a, b]$ (con $a < b$) diciamo che una variabile $X \in E$ ha legge (continua) uniforme su E (rispetto a I) se è continua con densità $\varrho(X = x|I)$ *costante* per $x \in E$. Dalla condizione

$$1 = \int_E \varrho(X = x|I) \mathrm{d}x = (b - a) \varrho(X = a|I),$$

ne segue che la costante vale $1/(b - a) = 1/\ell(E)$. Notiamo anche che, se $\ell(E) < 1$, la densità assume valori maggiori di 1 (ma questo, dicevamo nell'Osservazione 5, non è un problema).

Calcoliamo il valore atteso di una variabile X uniforme sull'intervallo $[a, b]$:

$$\mathbb{E}\left[X | X \text{ unif. su } [a, b]\right] = \int_a^b x \frac{\mathrm{d}x}{b-a} = \frac{1}{b-a} \frac{x^2}{2} \Big|_a^b = \frac{b^2 - a^2}{2(b-a)} = \frac{a+b}{2},$$

che corrisponde geometricamente al *punto medio* dell'intervallo $[a, b]$.

Per calcolare la varianza di X, calcoliamo prima

$$\mathbb{E}\left[X^2|X \text{ unif. su } [a,b]\right] = \int_a^b x^2 \frac{\mathrm{d}x}{b-a} = \frac{1}{b-a}\frac{x^3}{3}\Big|_a^b = \frac{b^3-a^3}{3(b-a)} = \frac{a^2+b^2+ab}{3},$$

e poi otteniamo con semplici passaggi algebrici

$$\mathrm{Var}\,(X|X \text{ unif. su } [a,b]) = \frac{a^2+b^2+ab}{3} - \left(\frac{a+b}{2}\right)^2 = \frac{1}{12}(b-a)^2.$$

Otteniamo quindi che la varianza è proporzionale al quadrato della lunghezza di $E = [a,b]$. La deviazione standard vale

$$\sigma(X|X \text{ unif. su } [a,b]) = \frac{1}{\sqrt{12}}(b-a) \approx 0.3(b-a).$$

Calcoliamo infine la funzione di ripartizione di una variabile uniforme: per $t \in E = [a,b]$,

$$P(X \le t|X \text{ unif. su } [a,b]) = \int_a^t \frac{\mathrm{d}x}{b-a} = \frac{t-a}{b-a},$$

e la funzione di sopravvivenza di consequenza vale, per $t \in [a,b]$,

$$P(X > t|X \text{ unif. su } [a,b]) = \frac{b-t}{b-a}.$$

Esercizio 12. Sia X una variabile uniforme (rispetto all'informazione I) sull'intervallo A e sia $B \subseteq \mathbb{R}$ tale che $A \cap B$ è un intervallo di lunghezza strettamente positiva. Mostrare che X, rispetto all'informazione $I \cap \{X \in B\}$ è uniforme sull'intervallo $A \cap B$. *(Sugg: usare il Lemma 10.)*

Esercizio 13. Sia X una variabile uniforme sull'intervallo $[a,b]$, siano λ, $c \in \mathbb{R}$ con $\lambda \neq 0$. Mostrare che $\lambda X + c$ è uniforme (determinare su quale intervallo). *(Sugg: usare il Lemma 11.)*

3.2. **Variabili esponenziali.** Le variabili esponenziali sono per certi aspetti la "approssimazione continua" delle variabili discrete geometriche. Se le variabili geometriche indicano il numero di tentativi prima che, in una successione di esperimenti indipendenti avvenga un "successo", le variabili esponenziali possono rappresentare il tempo (continuo) prima che avvenga un "successo" (anche se di solito si pensa ad esempio alla rottura di un oggetto, quindi il termine successo è un po' fuorviante). Una proprietà importante, comune ad esponenziali e geometriche, è l'assenza di memoria (Lemma 14).

Diamo la definizione precisa, usiamo la lettera T perché vogliamo pensare a una variabile "tempo". Dato $\lambda > 0$, diciamo che una variabile $T \in [0,+\infty)$ ha legge (continua) esponenziale di parametro λ (rispetto a I) se è continua con densità (per $t \in [0,\infty)$)

$$\varrho(T = t|I) = e^{-\lambda t}\lambda.$$

Scriviamo brevemente che T è $\mathcal{E}(\lambda)$. Il termine moltiplicativo λ assicura che

$$\int_0^\infty \varrho(T = t|I)\mathrm{d}t = \int_0^\infty e^{-\lambda t}\lambda \mathrm{d}t = 1,$$

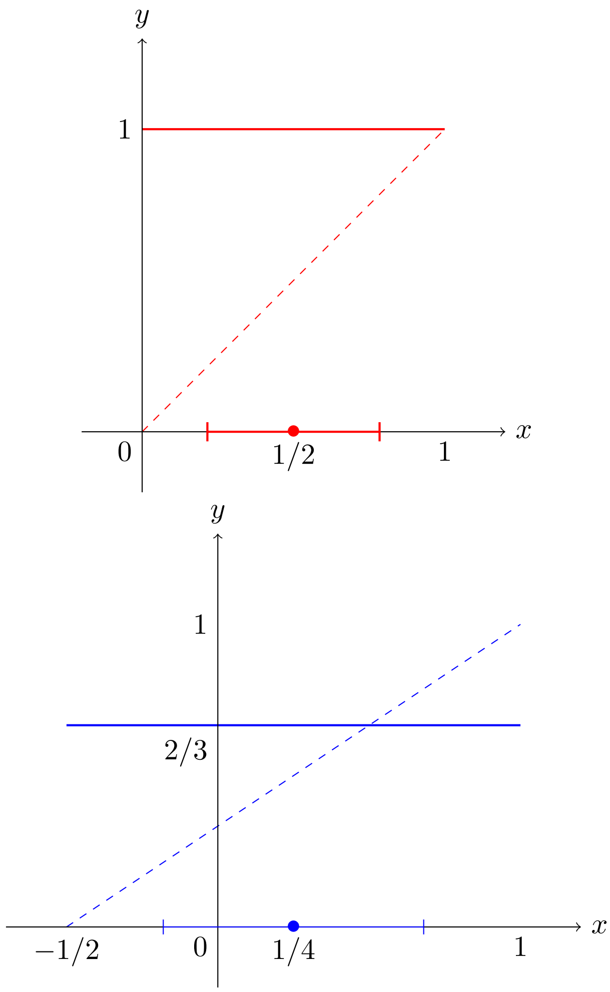

Figura 4. Esempi di densità di una variabile continua X con legge uniforme su $[0,1]$ (linea continua in rosso) e con legge uniforme su $[-\frac{1}{2},1]$ (linea continua in blu), con i rispettivi valori attesi $\pm\sigma$ (punti e intervalli colorati sulle ascisse) e funzioni di ripartizione (linee tratteggiate)

quindi in teoria un potrebbe anche dimenticarselo e ricavarlo ogni volta da questa condizione. Il termine $e^{-\lambda t}$ è di tipo esponenziale e necessariamente dobbiamo avere $\lambda > 0$, altrimenti l'integrale improprio su $[0,\infty)$ non esisterebbe. Notiamo (Figura 5) che, per λ piccolo, la densità si "allarga" (e quindi abbassa) su $[0,\infty)$ mentre per λ grande si "restringe" ad un "picco" vicino a $t=0$.

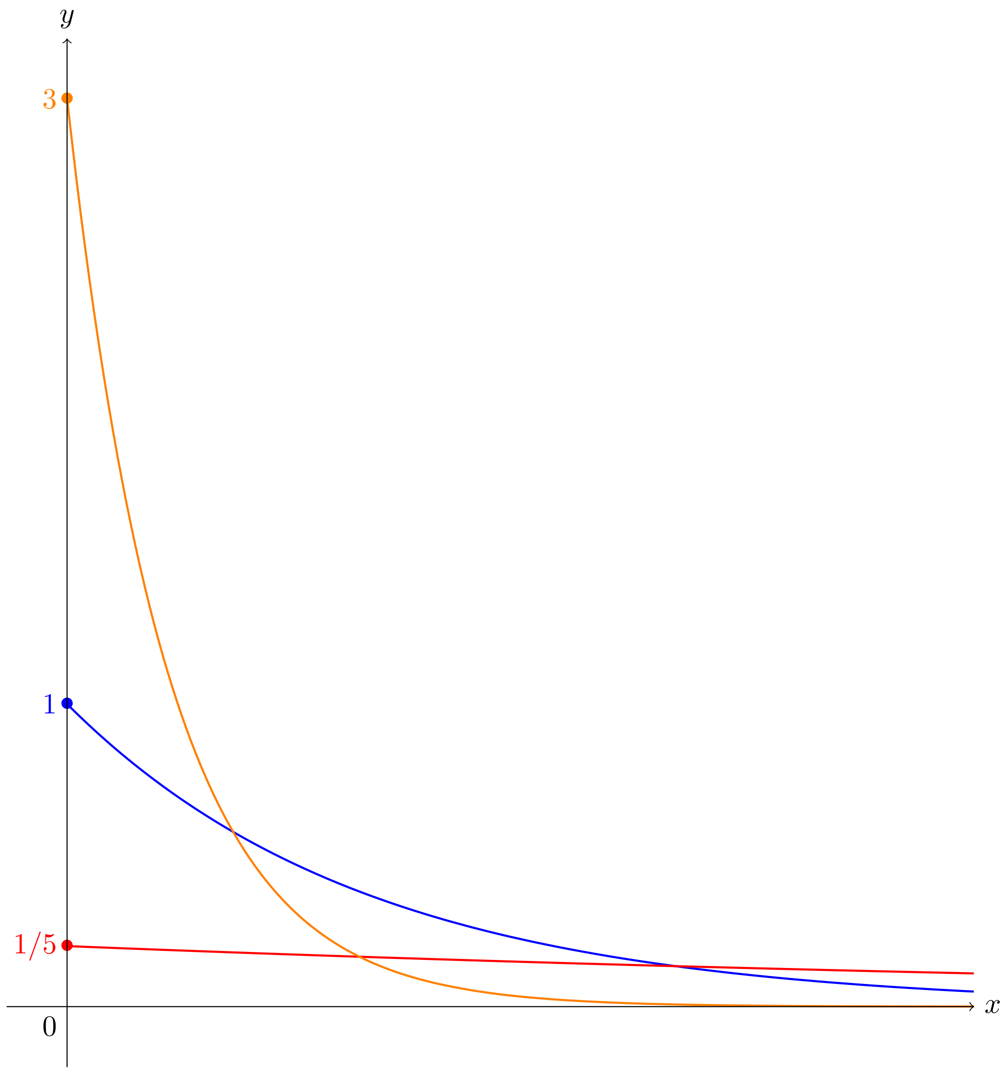

FIGURA 5. Esempi di densità di una variabile continua T con legge esponenziale di parametri $\lambda = 1$ (blu), $\lambda = 1/5$ (rossa), $\lambda = 3$ (arancio).

Calcoliamo il valore atteso di una variabile T con densità $\mathcal{E}(\lambda)$.

$$\begin{aligned}
\mathbb{E}\left[T|T \text{ è } \mathcal{E}(\lambda)\ \right] &= \int_0^\infty te^{-\lambda t}\lambda \mathrm{d}t \\
&= \frac{1}{\lambda}\int_0^\infty \lambda te^{-\lambda t}\lambda \mathrm{d}t \\
&= \frac{1}{\lambda}\int_0^\infty se^{-s}\mathrm{d}s \quad \text{(cambio di variabile } \lambda t = s) \\
&= \frac{1}{\lambda}\left(-se^{-s}\right)\Big|_0^\infty + \frac{1}{\lambda}\int_0^\infty e^{-s}\mathrm{d}s \quad \text{(integrando per parti)} \\
&= \frac{1}{\lambda}\left(-e^{-s}\right)\Big|_0^\infty \quad \text{(integrando ancora per parti)} \\
&= \frac{1}{\lambda}.
\end{aligned}$$

Il cambio di variabile $\lambda t = s$ appare frequentemente: una possibile giustificazione della sua utilità è data nell'Esercizio 15.

Per calcolare la varianza di T, calcoliamo prima

$$\begin{aligned}
\mathbb{E}\left[T^2 | T \text{ è } \mathcal{E}(\lambda)\ \right] &= \int_0^\infty t^2 e^{-\lambda t} \lambda \mathrm{d}t \\
&= \frac{1}{\lambda^2} \int_0^\infty \lambda^2 t^2 e^{-\lambda t} \lambda \mathrm{d}t \\
&= \frac{1}{\lambda^2} \int_0^\infty s^2 e^{-s} \mathrm{d}s \quad \text{(cambio di variabile } \lambda t = s) \\
&= \frac{1}{\lambda^2} \left(-s^2 e^{-s}\right)\Big|_0^\infty + \frac{2}{\lambda^2} \int_0^\infty s e^{-s} \mathrm{d}s \quad \text{(integrando per parti)} \\
&= \frac{2}{\lambda^2} \int_0^\infty s e^{-s} \mathrm{d}s = \frac{2}{\lambda^2}
\end{aligned}$$

integrando ancora per parti oppure usando gli integrali calcolati sopra.

Otteniamo allora che la varianza vale

$$\mathrm{Var}\left(T | T \text{ è } \mathcal{E}(\lambda)\ \right) = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

e quindi la deviazione standard

$$\sigma(T | T \text{ è } \mathcal{E}(\lambda)\ ) = \frac{1}{\lambda}.$$

Alla luce di questi risultati, in alcuni ambiti si preferisce usare $\beta = 1/\lambda$ come parametro fondamentale per le esponenziali.

Calcoliamo infine la funzione di sopravvivenza (più naturale di quella di ripartizione, trattandosi spesso di tempi di vita di apparecchiature ecc.). Per $t \in [0, \infty)$,

$$\begin{aligned}
P(T > t | T \text{ è } \mathcal{E}(\lambda)\ ) &= \int_t^\infty e^{-\lambda s} \lambda \mathrm{d}s \quad \text{(non ripetere } t \text{ nell'integrando!)} \\
&= -e^{-\lambda s}\Big|_t^\infty = e^{-\lambda t}.
\end{aligned}$$

Mentre la funzione di ripartizione vale, per $t \in [0, \infty)$,

$$P(T \leq t | T \text{ è } \mathcal{E}(\lambda)\ ) = 1 - e^{-\lambda t}.$$

Lemma 14 (Assenza di memoria)**.** *Sia T una variabile esponenziale di parametro $\lambda > 0$ (rispetto all'informazione I) e sia $t_0 \in [0, \infty)$. Allora, sapendo $I \cap \{T \geq t_0\}$, la variabile $T - t_0$ è esponenziale di parametro $\lambda > 0$. In simboli: (usando la densità) per $t \in [0, \infty)$,*

$$\varrho\left((T - t_0) = t | I \cap \{T \geq t_0\}\right) = e^{-\lambda t} \lambda,$$

(usando la funzione di sopravvivenza) per $t \in [0, \infty)$,

$$P(T - t_0 > t | I \cap \{T \geq t_0\}) = e^{-\lambda t}.$$

Questo risultato può essere anche espresso in modo efficace pensando a T come un tempo di vita di un apparecchiatura: sapendo che il tempo di vita è maggiore di un certo t_0, allora il tempo di vita "rimamente", $T - t_0$, è pure una variabile esponenziale, con lo stesso parametro λ. Ovviamente

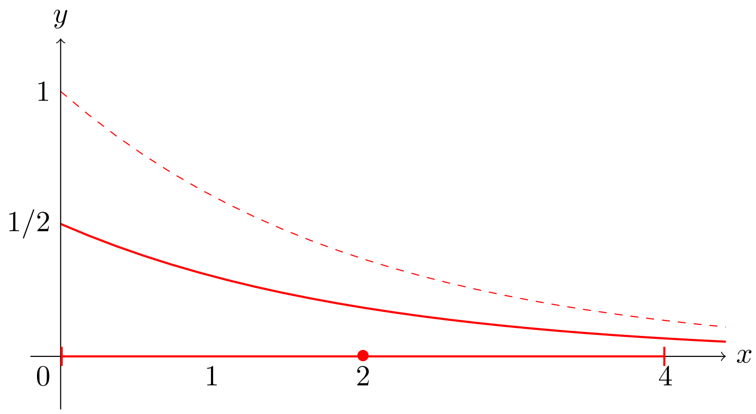

FIGURA 6. Esempio di densità di una variabile X con legge $\mathcal{E}(0.5)$ (linea continua in rosso) e con il rispettivo valori attesi $\pm\sigma$ (punto e intervalli colorati sulle ascisse) e funzione di sopravvivenza (linea tratteggiata)

questo non significa che $T - t_0$ è uguale a T, semplicemente hanno la *stessa densità* (ma rispetto a informazioni diverse!).

Dimostrazione. Diamo due dimostrazioni di questo fatto. Nella prima, mostriamo che la funzione di sopravvivenza di $T - t_0$ (rispetto all'informazione $I \cap \{T \geq t_0\}$) coincide con quella di una esponenziale (questo è sufficiente per identificare la densità di $T - t_0$). Notiamo che $T - t_0$ assume valori in $[0, \infty)$, se assumiamo l'informazione $\{T - t_0 \geq 0\}$. Quindi $T - t_0$ può essere vista come variabile aleatoria a valori in $[0, \infty)$ (rispetto a $I \cap \{T - t_0 \geq 0\}$). Dato $t \geq 0$, calcoliamo

$$
\begin{aligned}
P((T - t_0) \geq t | I \cap \{T \geq t_0\}) &= P(T \geq t + t_0 | I \cap \{T \geq t_0\}) \\
&= P(T \geq t_0 | I \cap \{T \geq t + t_0 \geq 0\}) \frac{P(T \geq t + t_0 | I)}{P(T \geq t_0 | I)} \quad \text{(Bayes)} \\
&= 1 \cdot \frac{P(T \geq t + t_0 | I)}{P(T \geq t_0 | I)} \quad (\text{se } T \geq t + t_0 \text{ allora } T \geq t_0) \\
&= \frac{e^{-\lambda(t + t_0)}}{e^{-\lambda t_0}} \quad (\text{funzione di sopravvivenza di } T \text{ rispetto ad } I) \\
&= e^{-\lambda t}.
\end{aligned}
$$

Una seconda dimostrazione calcola direttamente la densità di $T - t_0$ rispetto a $I \cap \{T \geq t_0\}$, usando il Lemma 11 e il Lemma 10. Dato $t \in [0, \infty)$, vale

$$
\begin{aligned}
\varrho((T - t_0) = t | I \cap \{T \geq t_0\}) &= \varrho((T - t_0) = t | I \cap \{T - t_0 \geq 0\}) \\
&= \frac{\varrho((T - t_0) = t | I)}{P(T - t_0 \geq 0 | I)} \quad (\text{Lemma 10) con } X = T - t_0 \text{ e } B = [0, \infty)) \\
&= \frac{\varrho(T = t + t_0 | I)}{P(T \geq t_0 | I)} \quad (\text{Lemma 11 con } X = T,\ \lambda = 1,\ c = t_0) \\
&= \lambda e^{-\lambda(t + t_0)} e^{\lambda t_0} = \lambda e^{-\lambda t}.
\end{aligned}
$$

□

Esercizio 15. Sia X una variabile aleatoria esponenziale (rispetto all'informazione I) di parametro λ. Sia $\mu > 0$. Mostrare che μX è una variabile esponenziale di parametro λ/μ. *(Sugg: usare il Lemma 11.)*

4. Probabilità condizionale

Abbiamo visto come definire in modo preciso variabili aleatorie che assumono (rispetto a una certa informazione I) un infinito "continuo di valori", introducendo il concetto di densità continua. In questa sezione ci occupiamo di estendere in modo preciso la validità di identità del tipo

$$P(B|I) = \sum_{e \in E} P(B|I \cap \{X = e\}) P(X = e|I), \tag{7}$$

per un evento B e una variable discreta $X \in E$, rimpiazzando però il sistema di alternativo "discreto" $(\{X = e\})_{e \in E}$, se $X \in E$ è una variabile discreta, con l'analogo sistema di alternative, ma nel caso "continuo". Se ragioniamo in termini formali, ossia prendendo piccoli intervalli δx_i intorno a un numero finito di punti $(x_i)_{i=1,\ldots,n}$ in modo da ricoprire tutto E, otteniamo l'identità

$$P(B|I) = \sum_{i=1}^{n} P(B|I \cap \{X \in \delta x_i\}) P(X \in \delta x_i|I),$$

Usando (3), approssimiamo

$$P(X \in \delta x_i|I) \approx \varrho(X = x_i|I)\ell(\delta x_i),$$

quindi (7) diventa

$$P(B|I) \approx \sum_{i=1}^{n} P(B|I \cap \{X \in \delta x_i\}) \varrho(X = x_i|I)\ell(\delta x_i).$$

Rimane però il problema di calcolare (o approssimare) i termini del tipo

$$P(B|I \cap \{X \in \delta x\}).$$

Se usiamo la definizione di probabilità condizionata (di Kolmogorov) abbiamo

$$P(B|I \cap \{X \in \delta x\}) = \frac{P(B \cap \{X \in \delta x\} |I)}{P(X \in \delta x|I)}.$$

Al tendere della la lunghezza dell'intervallo δx a zero, abbiamo un denominatore che tende a zero, ma pure il numeratore (perché comunque $P(B \cap \{X \in \delta x\} |I) \leq P(X \in \delta x|I)$), quindi abbiamo una forma indeterminata $0/0$. Intuitivamente però, potremmo pensare che, se $\ell(\delta x) \to 0$, l'informazione si fa via via più precisa e quindi dovrebbe esistere,

$$\lim_{\ell(\delta x) \to 0} P(B|I \cap \{X \in \delta x\}) \text{“} = P(B|I \cap \{X = x\})\text{''},$$

o più formalmente

$$P(B|I \cap \{X \in \delta x\}) \approx \text{“}P(B|I \cap \{X = x\})\text{''}, \tag{8}$$

e portarci alla versione continua di (7):

$$\begin{aligned} P(B|I) &\approx \sum_{i=1}^{n} P(B|I \cap \{X \in \delta x_i\}) \varrho(X = x_i|I) \ell(\delta x_i) \\ &= \int_E P(B|I \cap \{X = x\}) \varrho(X = x|I) \mathrm{d}x. \end{aligned}$$

Come nel caso della densità, diamo una definizione rigorosa della quantità $P(A|I \cap \{X = x\})$, quando $X \in E$ è una variabile continua, trascurando tutto il discorso delle "approssimazioni discrete" ma imponendo direttamente la validità della versione continua di (7).

Definizione 16 (Probabilità condizionale)**.** Sia $E \subseteq \mathbb{R}$ un intervallo, sia A un evento e sia $X \in E$ una variabile continua con densità $\varrho(X = \cdot|I)$. La probabilità condizionale di B dato $I \cap \{X = \cdot\}$ è una funzione (Riemann integrabile) definita su E,

$$E \ni x \mapsto P\left(B|I \cap \{X = x\}\right) \in [0, 1],$$

tale che valga l'identità

$$P(B|I) = \int_E P(B|I \cap \{X = x\}) \varrho(X = x|I) \mathrm{d}x. \tag{9}$$

Più in generale, richiediamo che valga anche l'identità, per ogni $A \subseteq E$ intervallo,

$$P(B|I \cap \{X \in A\}) P(X \in A|I) = \int_A P(B|I \cap \{X = x\}) \varrho(X = x|I) \mathrm{d}x. \tag{10}$$

Osservazione 17 (Sull'identità (10))*.* Se dividiamo per $P(X \in A|I)$ nella (10) e ricordiamo la formula della densità di X rispetto a $I \cap \{X \in A\}$ (Lemma 10 con A invece di B), questa diventa

$$P(B|I \cap \{X \in A\}) = \int_A P(B|I \cap \{X = x\}) \varrho(X = x|I \cap \{X \in A\}) \mathrm{d}x.$$

Nella maggior parte delle applicazioni, basterà usare (9), ma la definizione matematica *rigorosa* richiede (10) per ogni $B \subseteq E$ intervallo.

Osservazione 18 (Problemi matematici della probabilità condizionale)*.* L'esistenza e l'unicità della probabilità condizionale di un evento, data $I \cap \{X \in \cdot\}$ è un problema matematico non banale, ma per i problemi che considereremo, si potrebbe dimostrare che esiste sempre ed è (essenzialmente) unica. L'unicità però non è garantita per tutti i punti (infatti ad esempio se uno cambia $P(B|I \cap \{X = x\})$ in un solo punto, l'integrale (10) non cambia!), ma se è *continua* (come troveremo noi nei nostri esempi), allora è proprio unica.

Esercizio 19. Sia $P(B|I \cap \{X = \cdot\})$ la probabilità condizionale di B rispetto a $I \cap \{X = \cdot\}$ e sia $A \subseteq E$ intervallo con $P(X \in A|I) > 0$. Mostrare che la probabilità condizionale di B rispetto a $I \cap \{X \in A\} \cap \{X = \cdot\}$ coincide con $P(B|I \cap \{X = x\})$ se $x \in A$, e vale zero altrimenti.

5. Estensione del calcolo ad informazione condizionale

Per gli scopi pratici che ci proponiamo, tratteremo, per $x \in E$, la probabilità condizionale

$$P\left(B|I \cap \{X = x\}\right)$$

come una vera e propria *estensione* del concetto di probabilità condizionata, sfruttando sistematicamente la validità di (9) (e (10)). Abbiamo quindi la possibilità di estendere il calcolo delle probabilità rispetto ad *informazione condizionale*, ossia possiamo supporre che la nostra I (che finora abbiamo lasciato intendere essere un evento "normale" di probabilità non nulla) comprenda in realtà eventi del tipo $\{X = x\}$, dove X è una variabile aleatoria continua.

Le regole *intuitive* che dobbiamo tenere sempre in mente nel calcolo così esteso sono (3),

$$P(X \in \delta x|I) \approx \varrho\left(X = x|I\right)\ell(\delta x),$$

e (8),

$$P(B|I \cap \{X = x\}) \approx P(B|I \cap \{X \in \delta x\}),$$

dove δx è un piccolo intervallo intorno ad $x \in E$.

Le regole *rigorose* (che in realtà sono nelle definizioni) sono date da (5), cioè

$$P(X \in A|I) = \int_A \varrho\left(X = x|I\right) \mathrm{d}x,$$

e (10), cioè

$$P(B|I \cap \{X \in A\}) = \int_A P(B|I \cap \{X = x\})\varrho\left(X = x|I \cap \{X \in A\}\right) \mathrm{d}x$$

per ogni $A \subseteq E$ intervallo.

Possiamo *iterare* la nostra definizione di variabili assolutamente continue e probabilità condizionale, quando I è della forma $I' = I \cap \{X = x\}$, e X è una variabile continua rispetto a I (chiamiamola informazione "condizionale"). Ad esempio, diciamo che $Y \in F \subseteq \mathbb{R}$ (F intervallo) è assolutamente continua rispetto a $I' = I \cap \{X = x\}$ se esiste una funzione *densità* (a volte detta densità condizionale)

$$\varrho\left(Y = y|I \cap \{X = x\}\right)$$

tale che, per ogni intervallo $A \subseteq \mathbb{R}$, si abbia

$$P(Y \in F|I \cap \{X = x\}) = \int_A \varrho\left(Y = y|I \cap \{X = x\}\right) \mathrm{d}y.$$

Esempio 20. Siano $X \in E$, $Y \in F \subseteq \mathbb{R}$ due variabili aleatorie e supponiamo che $X \in E$ sia una variabile continua rispetto ad I con densità $\varrho(X = \cdot|I)$, e che, per ogni $x \in E$, la variabile $Y \in F'$ sia continua rispetto a $I \cap \{X = x\}$ con densità

$$\varrho\left(Y = y|I \cap \{X = x\}\right).$$

Allora, dato $B \subseteq F$ intervallo, vale

$$\begin{aligned} P(Y \in B|I) &= \int_E P(Y \in B|I \cap \{X = x\})\mathrm{d}x \\ &= \int_E \left(\int_B \varrho\left(Y = y|I \cap \{X = x\}\right) \mathrm{d}y\right) \mathrm{d}x. \end{aligned}$$

Se anche $A \subseteq E$ è un intervallo, vale allora (usando l'Esercizio 19)

$$P(Y \in B | I \cap \{X \in A\}) = \int_A \left(\int_B \varrho\,(Y = y | I \cap \{X = x\})\, \mathrm{d}y \right) \mathrm{d}x.$$

Il concetto di *indipendenza* tra due eventi o tra due variabili aleatorie, rispetto all'informazione I si estende in modo identico quando I è informazione condizionale: in particolare A, B eventi sono indipendenti se

$$P(A \cap B | I) = P(A|I) P(B|I)$$

o, se $P(B|I) > 0$,

$$P(A | I \cap B) = P(A|I).$$

Similmente, X e Y sono variabili indipendenti se, dati intervalli (o anche insiemi più generali) A, $B \subseteq \mathbb{R}$, gli eventi $\{X \in A\}$ e $\{Y \in B\}$ sono indipendenti.

Osservazione 21 (Indipendenza e variabili continue). Quando però (almeno) una delle variabili in questione è continua, allora può essere utile usare il fatto (non lo dimostriamo) che l'indipendenza permette di "trascurare" l'informazione anche all'interno della probabilità condizionale: Se X e Y sono indipendenti (rispetto ad I) allora

$$P(Y \in A | I \cap \{X = x\}) = P(Y \in A | I)$$

e nel caso in cui Y abbia densità (rispetto ad I) allora ha pure densità rispetto a $I \cap \{X = x\}$ e vale semplicemente

$$\varrho(Y = y | I \cap \{X = x\}) = \varrho(Y = y | I).$$

Pure il calcolo di valori attesi, varianze ecc. si estende a variabili discrete o continue rispetto a informazione condizionale I (che quindi contiene eventi della forma $\{X = x\}$ per X variabile continua). Abbiamo quindi, per $X \in E$ variabile continua e $f : E \to \mathbb{R}$ funzione (deterministica) continua,

$$\mathbb{E}\left[f(X) | I \right] = \int_E f(x) \varrho(X = x | I) \mathrm{d}x$$

e definizioni analoghe per la varianza $\mathrm{Var}\left(f(X) | I \right)$, e la deviazione standard $\sigma(f(X)|I) = \sqrt{\mathrm{Var}\left(f(X) | I \right)}$...

Raccogliamo le proprietà principali di queste operazioni nel seguente teorema, che non dimostriamo. Confrontatele con le proprietà enunciate (e dimostrate) nel caso di variabili discrete.

Teorema 22. *Date variabili aleatorie $X \in F$, $Y \in G \subseteq \mathbb{R}$ discrete o continue rispetto ad I (anche condizionale), date $f : F \to \mathbb{R}$, $g : G \to \mathbb{R}$ funzioni continue, tali che $\mathbb{E}\left[f(X) | I \right]$, $\mathbb{E}\left[g(Y) | I \right]$ siano ben definiti (cioè gli integrali esistano), valgono le seguenti proprietà:*

(1) *(linearità)* $\mathbb{E}\left[f(X) + g(Y) | I \right] = \mathbb{E}\left[f(X) | I \right] + \mathbb{E}\left[g(Y) | I \right]$, $\mathbb{E}\left[\lambda f(X) | I \right] = \lambda \mathbb{E}\left[f(X) | I \right]$, *se* $\lambda \in \mathbb{R}$ *è costante.*
(2) *(monotonia) se* $P(f(X) \leq g(Y) | I) = 1$, *allora* $\mathbb{E}\left[f(X) | I \right] \leq \mathbb{E}\left[g(Y) | I \right]$.
(3) *per ogni* $\lambda > 0$, *vale*

$$P\left(|f(X)| \geq \lambda | I \right) \leq \frac{\mathbb{E}\left[|f(X)| \right]}{\lambda} \quad \text{(Markov)}$$

$$P\left(|f(X) - \mathbb{E}\left[f(X) | I \right]| \geq \lambda | I \right) \leq \frac{\mathrm{Var}\left(f(X) | I \right)}{\lambda^2} \quad \text{(Chebychev)}$$

(4) Se X e Y sono indipendenti, allora

$$\mathbb{E}\left[f(X)g(Y)|I\right] = \mathbb{E}\left[f(X)|I\right]\mathbb{E}\left[g(Y)|I\right].$$

(5) Se X e Y sono indipendenti, allora

$$\operatorname{Var}\left(f(X)+g(Y)|I\right) = \operatorname{Var}\left(g(Y)|I\right) + \operatorname{Var}\left(g(Y)|I\right).$$

Possiamo enunciare delle formule di Bayes “estese” nel caso di densità e probabilità condizionali.

Lemma 23 (Bayes: discreto/continuo)**.** *Siano $X \in E$, $Y \in F$ variabili aleatorie tali che*

(1) Y è discreta rispetto ad I (informazione eventualmente condizionale);
(2) per ogni $y \in F$, X è continua rispetto ad $I \cap \{Y = y\}$ con densità

$$\varrho(X = \cdot|I \cap \{Y = y\}).$$

Allora X è continua rispetto ad I, con densità

$$\varrho(X = \cdot|I) = \sum_{y \in F} \varrho(X = \cdot|I \cap \{Y = y\})P(Y = y|I)$$

e vale la formula di Bayes, per $x \in E$, $y \in F$,

$$P(Y = y|I \cap \{X = x\}) = \varrho(X = x|I \cap \{Y = y\})\frac{P(Y = y|I)}{\varrho(X = x|I)}. \tag{11}$$

Possiamo enunciare la formula (11) anche “rovesciando” i ruoli di X e Y. Si ottiene allora

$$\varrho(X = x|I \cap \{Y = y\}) = P(Y = y|I \cap \{X = x\})\frac{\varrho(X = x|I)}{P(Y = y|I)}.$$

Ricordiamo che possiamo sempre calcolare

$$P(Y = y|I) = \int_E P(Y = y|I \cap \{X = x\})\varrho(X = x|I)\mathrm{d}x.$$

Osservazione 24 (caso discreto con due alternative). Un caso molto semplice ma frequente è quello in cui $F = \{0, 1\}$ ossia Y è una variabile indicatrice di un evento $A = \{Y = 1\}$. In tal caso, le ipotesi diventano: X è continua sia rispetto ad $I \cap A$ e rispetto a $I \cap A^c$, con densità

$$\varrho(X = \cdot|I \cap A), \quad \varrho(X = \cdot|I \cap A^c).$$

E la tesi è che la densità di X rispetto ad I vale

$$\varrho(X = \cdot|I) = \varrho(X = \cdot|I \cap A)P(A|I) + \varrho(X = \cdot|I \cap A^c)P(A^c|I)$$

e la formula di Bayes è, per $x \in E$,

$$P(A|I \cap \{X = x\}) = \varrho(X = x|I \cap A)\frac{P(A|I)}{\varrho(X = x|I)}.$$

Teorema 25 (Bayes: continuo/continuo)**.** *Siano $X \in E$, $Y \in F$ variabili aleatorie tali che*

(1) Y è continua rispetto ad I (informazione eventualmente condizionale), con densità $\varrho(Y = \cdot|I)$.

(2) per ogni $y \in F$, X è continua rispetto ad $I \cap \{Y = y\}$ con densità

$$\varrho(X = \cdot | I \cap \{Y = y\}).$$

Allora X è continua rispetto ad I, con densità

$$\varrho(X = \cdot | I) = \int_{y \in F} \varrho(X = \cdot | I \cap \{Y = y\}) \varrho(Y = y|I) \mathrm{d}y$$

e vale la formula di Bayes, per $x \in E$, $y \in F$,

$$\varrho(Y = y|I \cap \{X = x\}) = \varrho(X = x|I \cap \{Y = y\}) \frac{\varrho(Y = y|I)}{\varrho(X = x|I)}. \tag{12}$$

Osservazione 26 (Come ricordare le formule di Bayes)*.* Per ricordare queste formule, conviene pensare sempre al caso discreto, in cui si ha

$$P(Y = y|I \cap \{X = x\}) = P(X = x|I \cap \{Y = y\}) \frac{P(Y = y|I)}{P(X = x|I)}$$

(applicate la formula per eventi al caso $A = \{X = x\}$, $B = \{Y = y\}$). Nel caso continuo, si vede che alcuni termini devono diventare densità e altri probabilità condizionali. Notate però che le "unità di misura" devono essere le stesse in ambo i membri.

Concludiamo questa sezione enunciando (senza dimostrazione) una formula di "convoluzione" per la somma di variabili aleatorie continue, del tutto analoga al caso discreto.

Lemma 27 (Densità della somma di variabili aleatorie)**.** *Siano X, $Y \in \mathbb{R}$ variabili aleatorie tali che*

(1) Y è continua rispetto ad I (informazione eventualmente condizionale), con densità $\varrho(Y = \cdot|I)$.
(2) per ogni $y \in F$, X è continua rispetto ad $I \cap \{Y = y\}$ con densità

$$\varrho(X = \cdot | I \cap \{Y = y\}).$$

Allora la variabile $X + Y$ è continua (rispetto ad I) con densità, per $z \in \mathbb{R}$,

$$\varrho\left((X + Y) = z|I\right) = \int_{\mathbb{R}} \varrho\left(X = z - y|I \cap \{Y = y\}\right) \varrho\left(Y = y|I\right) \mathrm{d}y.$$

Nel caso in cui X e Y siano indipendenti, si può trascurare il condizionamento rispetto a $\{Y = y\}$ e quindi si trova

$$\varrho\left((X + Y) = z|I\right) = \int_{\mathbb{R}} \varrho\left(X = z - y|I\right) \varrho\left(Y = y|I\right) \mathrm{d}y. \tag{13}$$

6. Densità gaussiana (o normale)

In questa sezione studiamo una famiglia fondamentale di densità continue, dette gaussiane (dal matematico Gauss) o normali. Si potrebbe pensare a questa famiglia come ad una controparte "continua" delle variabili binomiali (più eventualmente una costante), ma il suo ruolo nella probabilità è ben più "centrale", e in effetti vedremo un teorema (del limite centrale 35) che ne giustifica l'apparizione in molte situazioni.

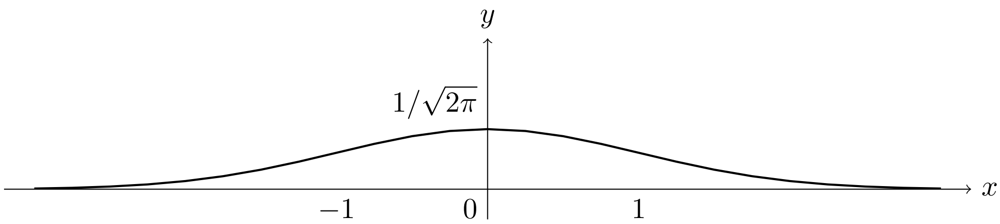

FIGURA 7. Grafico della densità di una variabile X normale standard $\mathcal{N}(0,1)$ nell'intervallo $(-3,3)$.

Definizione 28. Fissati $m \in \mathbb{R}$, $\sigma^2 \in (0,\infty)$, diciamo che $X \in \mathbb{R}$ è una variabile con legge gaussiana (o normale) di parametri (m,σ^2) (rispetto all'informazione I) se X è continua con densità

$$\varrho(X = x|I) = \exp\left[-\frac{1}{2}\frac{(x-m)^2}{\sigma^2}\right]\frac{1}{\sqrt{2\pi\sigma^2}},$$

e scriviamo X è $\mathcal{N}(m,\sigma^2)$.

Vedremo tra poco (Lemma 33) che m è il valore atteso di X e σ^2 è la varianza (quindi $|\sigma|$ è la deviazione standard di X).

Osservazione 29 (normale standard). Nel caso in cui $m = 0$, $\sigma^2 = 1$, ossia X è $\mathcal{N}(0,1)$ diciamo che X ha legge normale *standard*. La densità di una variabile normale standard diventa semplicemente

$$\varrho(X = x|X \text{ è } \mathcal{N}(0,1)) = e^{-\frac{x^2}{2}}\frac{1}{\sqrt{2\pi}} \tag{14}$$

L'importanza della normale standard, oltre alla semplicità della formula di densità, è che molti calcoli per variabili normali generali si possono ricondurre al caso particolare della normale standard, grazie al Lemma 31. Se ne tracciamo il grafico (Figura 7), notiamo che pur essendo sempre positiva, la densità tende a zero molto rapidamente (infatti, permette di integrarci contro qualunque polinomio), inoltre è una funzione *pari* (simmetrica rispetto all'asse $x = 0$), quindi ne deduciamo

$$P(X > 0|X \text{ è } \mathcal{N}(0,1)) = P(X < 0|X \text{ è } \mathcal{N}(0,1)) = \int_0^{+\infty} e^{-\frac{x^2}{2}}\frac{1}{\sqrt{2\pi}}\mathrm{d}x = \frac{1}{2}. \tag{15}$$

Osservazione 30 (costante $1/\sqrt{2\pi}$). La costante $1/\sqrt{2\pi}$ nella densità gaussiana appare affinché la densità integrata su tutto $\mathbb{R}$ sia 1:

$$\int_{\mathbb{R}} \exp\left[-\frac{1}{2}\frac{(x-m)^2}{\sigma^2}\right]\frac{1}{\sqrt{2\pi\sigma^2}}\mathrm{d}x = 1.$$

Una dimostrazione di questo fatto però non è banale (ad esempio può richiedere integrali multipli) e quindi lo assumiamo senza dimostrarlo. Tuttavia molti altri calcoli (integrali), una volta noto questo fatto, sono alla nostra portata.

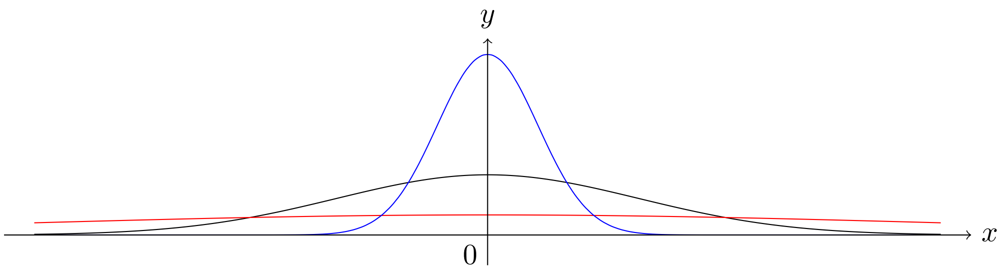

FIGURA 8. Esempi di densità gaussiane: in nero $\mathcal{N}(0,1)$, in rosso $\mathcal{N}(0,9)$, in blu $\mathcal{N}(0,1/9)$.

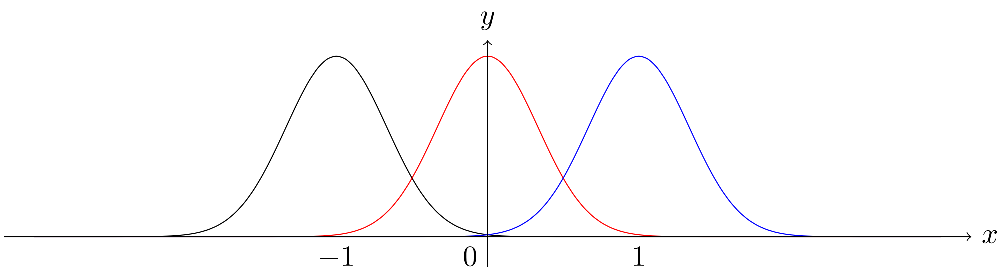

FIGURA 9. Esempi di densità gaussiane: in nero $\mathcal{N}(-1,1/9)$, in rosso $\mathcal{N}(0,1/9)$, in blu $\mathcal{N}(1,1/9)$.

Lemma 31 (trasformazioni affini di gaussiane)**.** *Siano a, $b \in \mathbb{R}$, con $a \neq 0$, siano $m \in \mathbb{R}$, $\sigma^2 \in (0,\infty)$ e sia X una variabile $\mathcal{N}(m,\sigma^2)$ (rispetta all'informazione I). Allora la variabile $aX+b$ è $\mathcal{N}(am+b, a^2\sigma^2)$ (sempre rispetto all'informazione I).*

Osservazione 32. Per ricordare come si trasformano i parametri, basta ricordare (Lemma 33) che m corrisponde al valore atteso di una gaussiana, quindi $\mathbb{E}[aX+b|I] = a\mathbb{E}[X|I]+b = am+b$ è il parametro m "trasformato", mentre σ^2 corrisponde alla varianza, quindi $\operatorname{Var}(aX+b) = \operatorname{Var}(aX) = a^2 \operatorname{Var}(X) = a^2\sigma^2$ è il parametro σ^2 "trasformato".

Dimostrazione. Si tratta di applicare il Lemma 11, con a al posto di λ, b al posto di c. Otteniamo che la densità di $aX+b$ è

$$
\begin{aligned}
\varrho((aX+b)=yI) &= \varrho(X=(y-b)/a|I)\,\frac{1}{|a|} \\
&= \exp\left[-\frac{1}{2\sigma^2}\left(\frac{y-b}{a}-m\right)^2\right]\frac{1}{\sqrt{2\pi\sigma^2}}\frac{1}{|a|} \\
&= \exp\left[-\frac{1}{2\sigma^2}\left(\frac{y-b-am}{a}\right)^2\right]\frac{1}{\sqrt{2\pi\sigma^2 a^2}} \\
&= \exp\left[-\frac{1}{2}\frac{(y-(am+b))^2}{(a\sigma)^2}\right]\frac{1}{\sqrt{2\pi\,(a\sigma)^2}},
\end{aligned}
$$

che è la densità di una legge $\mathcal{N}(am+b,\sigma^2)$. $\square$

Lemma 33. *Sia $X \in \mathbb{R}$ una variabile aleatoria con legge $\mathcal{N}(m, \sigma^2)$ (rispetto all'informazione I). Allora vale*

$$\mathbb{E}[X|I] = m, \quad \mathrm{Var}(X|I) = \sigma^2.$$

Dimostrazione. Dimostriamolo prima nel caso in cui X sia $\mathcal{N}(0,1)$, ossia normale standard. Il fatto che la densità decresca molto rapidamente per $x \to \pm\infty$ e sia pari, ci permette di concludere che

$$\begin{aligned}
\mathbb{E}[X|X \text{ è } \mathcal{N}(0,1)] &= \int_{\mathbb{R}} x e^{-\frac{x^2}{2}} \frac{1}{\sqrt{2\pi}} \mathrm{d}x \\
&= \int_0^{+\infty} x e^{-\frac{x^2}{2}} \frac{1}{\sqrt{2\pi}} \mathrm{d}x + \int_{-\infty}^{0} x e^{-\frac{x^2}{2}} \frac{1}{\sqrt{2\pi}} \mathrm{d}x \\
&= \int_0^{+\infty} x e^{-\frac{x^2}{2}} \frac{1}{\sqrt{2\pi}} \mathrm{d}x - \int_0^{+\infty} y e^{-\frac{y^2}{2}} \frac{1}{\sqrt{2\pi}} \mathrm{d}y \quad \text{(cambio di variabile } y = -x\text{)} \\
&= 0,
\end{aligned}$$

brevemente: perché $x e^{-\frac{x^2}{2}} \frac{1}{\sqrt{2\pi}}$ è dispari, quindi i contributi all'integrale simmetrici rispetto all'asse $x = 0$ si cancellano (Figura 10). In modo simile, ma sfruttando il fatto che $x^2 e^{-\frac{x^2}{2}} \frac{1}{\sqrt{2\pi}}$ è pari, abbiamo invece

$$\begin{aligned}
\mathbb{E}\left[X^2|X \text{ è } \mathcal{N}(0,1)\right] &= \int_{\mathbb{R}} x^2 e^{-\frac{x^2}{2}} \frac{1}{\sqrt{2\pi}} \mathrm{d}x \\
&= \int_0^{+\infty} x^2 e^{-\frac{x^2}{2}} \frac{1}{\sqrt{2\pi}} \mathrm{d}x + \int_{-\infty}^{0} x^2 e^{-\frac{x^2}{2}} \frac{1}{\sqrt{2\pi}} \mathrm{d}x \\
&= \int_0^{+\infty} x^2 e^{-\frac{x^2}{2}} \frac{1}{\sqrt{2\pi}} \mathrm{d}x - \int_0^{+\infty} y^2 e^{-\frac{y^2}{2}} \frac{1}{\sqrt{2\pi}} \mathrm{d}y \quad \text{(cambio di variabile } y = -x\text{)} \\
&= -2 \int_0^{+\infty} x \left(-x e^{-\frac{x^2}{2}} \right) \frac{1}{\sqrt{2\pi}} \mathrm{d}x \\
&= -2 \int_0^{+\infty} x \left(\frac{d}{dx} e^{-\frac{x^2}{2}} \right) \frac{1}{\sqrt{2\pi}} \mathrm{d}x \\
&= -2x e^{-\frac{x^2}{2}} \frac{1}{\sqrt{2\pi}} \Big|_0^{+\infty} + 2 \int_0^{+\infty} e^{-\frac{x^2}{2}} \frac{1}{\sqrt{2\pi}} \mathrm{d}x \quad \text{(integrando per parti)} \\
&= 2 \int_0^{+\infty} e^{-\frac{x^2}{2}} \frac{1}{\sqrt{2\pi}} \mathrm{d}x = 1. \quad \text{(usando (15))}.
\end{aligned}$$

Siccome $\mathbb{E}[X|X \text{ è } \mathcal{N}(0,1)] = 0$, ne deduciamo

$$\mathrm{Var}(X|X \text{ è } \mathcal{N}(0,1)) = 1.$$

Per trattare il caso generale, usiamo il Lemma 31: data X con legge $\mathcal{N}(m, \sigma^2)$, consideriamo la variabile

$$Z = \frac{X - m}{|\sigma|}$$

che grazie al Lemma 31 ha legge $\mathcal{N}((m-m)/|\sigma|, \sigma^2/|\sigma|^2) = \mathcal{N}(0,1)$. Di conseguenza, poiché vale

$$X = |\sigma| Z + m$$

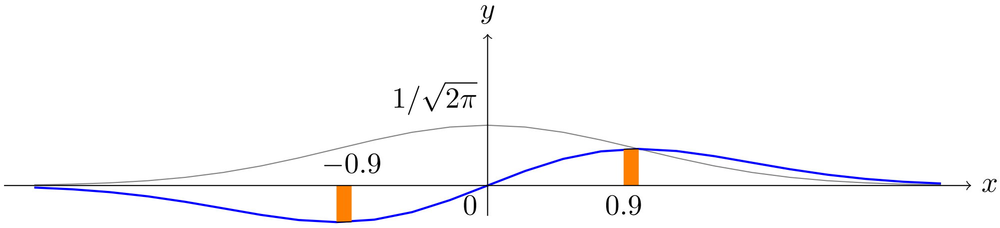

Figura 10. I contributi (in arancio) simmetrici rispetto a $x = 0$ del valore atteso di una v.a. con densità ϱ pari (linea grigia, in questo caso $\mathcal{N}(0,1)$) si cancellano, perché la funzione $x\varrho(x)$ (linea blu) è dispari.

usando quanto visto sopra per le normali standard,

$$\mathbb{E}\left[X|X \text{ è } \mathcal{N}(m,\sigma^2)\right] = \mathbb{E}\left[|\sigma|\, Z + m|Z \text{ è } \mathcal{N}(0,1)\right] = |\sigma| \cdot 0 + m = m,$$

$$\operatorname{Var}\left(X|X \text{ è } \mathcal{N}(m,\sigma^2)\right) = \operatorname{Var}\left(|\sigma|\, Z + m|Z \text{ è } \mathcal{N}(0,1)\right) = \sigma^2.$$

e quindi $\sigma(X|X \text{ è } \mathcal{N}(m,\sigma^2)) = |\sigma|$. □

Ora studiamo una proprietà *fondamentale* delle variabili gaussiane: la somma di variabili gaussiane *indipendenti* è ancora gaussiana, con l'ovvia trasformazione dei valori attesi (si sommano) e delle varianze (pure si sommano, per via dell'indipendenza). Diamo l'enunciato con dimostrazione nel caso di *due* variabili indipendenti entrambe con legge normale standard, il caso generale e pure l'estensione a n variabili è solamente tecnicamente più complicato (quindi non lo dimostriamo).

Lemma 34. *Siano X_1, X_2 due variabili aleatorie continue* indipendenti *rispettivamente con legge $\mathcal{N}(m_1,\sigma_1^2)$, $\mathcal{N}(m_2,\sigma_2^2)$ (tutto questo rispetto a un'informazione fissata I), dove m_1, $m_2 \in \mathbb{R}$, σ_1^2, $\sigma_2^2 \in (0,\infty)$. Allora, la variabile X_1+X_2 è anch'essa gaussiana (rispetto ad I), con legge $\mathcal{N}(m_1+m_2,\sigma_1^2+\sigma_2^2)$.*

Dimostrazione. Usiamo la formula di convoluzione per densità continue (Lemma 27), e sfruttiamo l'indipendenza per avere la densità condizionale:

$$\varrho(Y = y|I \cap \{X = x\}) = \varrho(Y = y|I) = \exp\left(-\frac{y^2}{2}\right)\frac{1}{\sqrt{2\pi}}.$$

Calcoliamo allora

$$\begin{aligned}\varrho(X+Y = z|I) &= \int_{\mathbb{R}} \varrho(Y = z-x|I)\varrho(X = x|I)\mathrm{d}x \\ &= \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} \exp\left(-\frac{(z-x)^2}{2}\right)\exp\left(-\frac{x^2}{2}\right)\frac{1}{\sqrt{2\pi}}\mathrm{d}x \\ &= \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} \exp\left[-\frac{1}{2}\left(x^2+(z-x)^2\right)\right]\frac{1}{\sqrt{2\pi}}\mathrm{d}x.\end{aligned}$$

Concentriamoci sull'esponente che appare nell'integrale (rispetto alla variabile x)

$$\begin{aligned} x^2 + (z-x)^2 &= x^2 + z^2 - 2xz + x^2 \\ &= z^2 + 2x^2 - 2xz \\ &= \left(\sqrt{2}x\right)^2 - 2\left(\sqrt{2}x\right)\left(\frac{z}{\sqrt{2}}\right) + \left(\frac{z}{\sqrt{2}}\right)^2 + \frac{z^2}{2} \\ &= \left(\sqrt{2}x - \frac{z}{\sqrt{2}}\right)^2 + \frac{z^2}{2} \\ &= \left(\frac{x-\frac{z}{2}}{1/\sqrt{2}}\right)^2 + \left(\frac{z}{\sqrt{2}}\right)^2. \end{aligned}$$

Questa identità algebrica ci permette di ricondurre l'integrale a quello di una densità gaussiana (precisamente una densità $\mathcal{N}(\frac{z}{2}, \frac{1}{2})$), infatti troviamo le identità

$$\begin{aligned} &\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \exp\left[-\frac{1}{2}\left(x^2 + (z-x)^2\right)\right] \frac{1}{\sqrt{2\pi}} \mathrm{d}x \\ &\quad= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \exp\left[-\frac{1}{2}\left(\frac{x-\frac{z}{2}}{1/\sqrt{2}}\right)^2 - \frac{1}{2}\left(\frac{z}{\sqrt{2}}\right)^2\right] \frac{1}{\sqrt{2\pi}} \mathrm{d}x \\ &\quad= \frac{1}{\sqrt{2\pi \cdot 2}} \exp\left[-\frac{1}{2}\left(\frac{z}{\sqrt{2}}\right)^2\right] \int_{\mathbb{R}} \exp\left[-\frac{1}{2}\left(\frac{x-\frac{z}{2}}{1/\sqrt{2}}\right)^2\right] \frac{1}{\sqrt{2\pi \cdot \frac{1}{2}}} \mathrm{d}x \\ &\quad= \frac{1}{\sqrt{2\pi \cdot 2}} \exp\left[-\frac{1}{2}\left(\frac{z}{\sqrt{2}}\right)^2\right]. \end{aligned}$$

e riconosciamo quindi la densità di una gaussiana $\mathcal{N}(0, 2)$. □

Combinando il Lemma 34 e il Lemma 31, otteniamo una proprietà molto importante delle variabili gaussiane: se $X_1, X_2, \ldots X_n$ sono n variabili indipendenti *tutte* gaussiane dei medesimi parametri m, σ^2, allora la media aritmetica

$$\frac{1}{n}\sum_{i=1}^n X_i$$

è una variabile gaussiana, di media $\frac{1}{n}\sum_{i=1}^n m = m$ e varianza $\frac{1}{n^2}\sum_{i=1}^n \sigma^2 = \frac{\sigma^2}{n}$. Il fatto che la varianza della media aritmetica tende a zero è un'altra espressione della legge dei grandi numeri, la quale assicura che con grande probabilità, al tendere di n all'infinito, la media aritmetica è molto vicina al valore atteso m. Se vogliamo essere più precisi – in un certo senso vogliamo capire di quale ordine di infinitesimo si tratta – possiamo dividere la differenza tra la media aritmetica e il valore atteso m per la quantità $\sqrt{\sigma^2/n}$ (ossia la deviazione standard). Usando un'altra volta il Lemma 31, deduciamo che

$$\left(\frac{1}{n}\sum_{i=1}^n X_i - m\right)\sqrt{\frac{n}{\sigma^2}}$$

ha legge $\mathcal{N}(0,1)$, ossia la normale standard. Potremmo anche esprimere più semplicemente il risultato scrivendo l'uguaglianza

$$\frac{1}{n}\sum_{i=1}^{n} X_i = m + \sqrt{\frac{\sigma^2}{n}} Z$$

dove Z ha legge $\mathcal{N}(0,1)$.

Questi calcoli (molto semplici) che valgono per n variabili gaussiane indipendenti in realtà valgono in un senso approssimato anche per n variabili indipendenti tutte con la stessa legge, non necessariamente gaussiane. Il teorema preciso è noto come "teorema del limite centrale" (*central limit theorem* in inglese), dove l'aggettivo centrale significa "importante". Infatti questo teorema giustifica l'apparizione di variabili gaussiane in molti ambiti, come approssimazione di somme di variabili indipendenti.

Teorema 35 (Teorema del limite centrale)**.** *Fissati $m \in \mathbb{R}$, $\sigma^2 \in (0,\infty)$, per ogni $n \geq 1$, siano $X_1, \ldots, X_n \in \mathbb{R}$ variabili aleatorie indipendenti tutte con la stessa legge (discreta o continua), aventi valore atteso e varianza*

$$\mathbb{E}[X_1|I] = \ldots = \mathbb{E}[X_n|I] = m \quad \mathrm{Var}(X_1|I) = \ldots = \mathrm{Var}(X_n|I) = \sigma^2.$$

Allora, per ogni intervallo $A \subseteq \mathbb{R}$, vale

$$\lim_{n\to\infty} P\left(\left(\frac{1}{n}\sum_{i=1}^{n} X_i - m\right)\sqrt{\frac{n}{\sigma^2}} \in A|I\right) = \int_A e^{-\frac{1}{2}x^2}\frac{1}{\sqrt{2\pi}}\mathrm{d}x.$$

In modo più informale, possiamo dire che, per n grande, vale l'approssimazione

$$\frac{1}{n}\sum_{i=1}^{n} X_i \approx m + \sqrt{\frac{\sigma^2}{n}} Z,$$

dove Z ha legge $\mathcal{N}(0,1)$. Oppure equivalentemente (ma sempre in modo formale),

$$\frac{1}{n}\sum_{i=1}^{n} X_i \approx W,$$

dove W ha legge $\mathcal{N}(m, \sigma^2/n)$.

7. Problemi

Problema 36. La direzione di un supermercato deve acquistare un nuovo freezer per l'esposizione dei surgelati in vendita. La durata di funzionamento in mesi di un modello di freezer per supermercato è descritta da una variabile aleatoria T con distribuzione esponenziale di parametro $1/120$.

(1) Si calcoli la funzione di sopravvivenza del freezer (cioè la funzione $t \mapsto P(T > t|\Omega)$), valore atteso e varianza di T.
(2) Dopo aver installato il freezer, sapendo che sono passati 60 mesi di funzionamento corretto, si calcoli la densità della variabile "tempo di funzionamento rimanente" (ossia $T - 60$).
(3) Mentre i primi 5 anni di funzionamento sono coperti da garanzia, il costo della riparazione del freezer è $\exp(T/10)$ per una rottura che avviene in un tempo T (in mesi) tra i 5 e i 10 anni di funzionamento

(dopo i 10 anni, o a una seconda rottura, l'azienda decide di rottamare il freezer). Calcolare il valore atteso del costo di riparazione, prima di installare il freezer. Calcolarne il valore atteso sapendo che non si è rotto nei primi 5 anni di funzionamento.

Una soluzione. (1) Si tratta di usare le formule generali per le variabili esponenziali, nel caso in cui $\lambda = 1/120$. Troviamo

$$P(T > t|\Omega) = e^{-t/120}, \quad \text{per } t \geq 0,$$

$$\mathbb{E}\left[T|\Omega\right] = 120, \quad \operatorname{Var}(T|\Omega) = 120^2, \quad \sigma(T|\Omega) = 120.$$

(2) Usiamo l'assenza di memoria delle variabili esponenziali: la legge di $T - 60$, rispetto a $\Omega \cap \{T \geq 60\}$ è ancora $\mathcal{E}(1/120)$, ossia

$$\varrho((T-60) = t|\Omega \cap \{T \geq 60\}) = \frac{1}{120}e^{-t/120}, \quad \text{per } t \geq 0.$$

(3) Il valore atteso del costo di riparazione C prima di installare il freezer (ossia rispetto alla informazione Ω) è

$$\begin{aligned}
\mathbb{E}\left[C|\Omega\right] &= \mathbb{E}\left[C|\Omega \cap T \leq 60\right] P(T \leq 60|\Omega) \\
&\quad + \mathbb{E}\left[C|\Omega \cap 60 < T \leq 120\right] P(60 < T \leq 120|\Omega) \\
&\quad + \mathbb{E}\left[C|\Omega \cap T > 120\right] P(T > 120|\Omega) \\
&= \mathbb{E}\left[\exp(T/10)|\Omega \cap 60 < T \leq 120\right] P(60 < T \leq 120|\Omega) \\
&= \int_{60}^{120} \exp(t/10)\exp(-t/120)\frac{1}{120}\mathrm{d}t \\
&= \int_{60}^{120} \exp\left(t(1/10 - 1/120)\right)\frac{1}{120}\mathrm{d}t \\
&= \frac{\exp\left(11t/120\right)}{11/120}\Big|_{60}^{120}\frac{1}{120} = \frac{\exp\left(11\right) - \exp\left(11/2\right)}{11}.
\end{aligned}$$

Sapendo invece che nei primi 5 anni di funzionamento non si è rotto, calcoliamo similmente

$$\begin{aligned}
\mathbb{E}\left[C|\Omega \cap \{T > 60\}\right] &= \mathbb{E}\left[C|\Omega \cap T \leq 60 \cap \{T > 60\}\right] P(T \leq 60|\Omega \cap \{T > 60\}) \\
&\quad + \mathbb{E}\left[C|\Omega \cap 60 < T \leq 120 \cap \{T > 60\}\right] P(60 < T \leq 120|\Omega \cap \{T > 60\}) \\
&\quad + \mathbb{E}\left[C|\Omega \cap T > 120 \cap \{T > 60\}\right] P(T > 120|\Omega \cap \{T > 60\}) \\
&= \mathbb{E}\left[\exp(T/10)|\Omega \cap 60 < T \leq 120\right] P(60 < T \leq 120| \cap \{T > 60\}) \\
&= \int_{60}^{120} \exp\left(t(1/10 - 1/120)\right)\frac{1}{120}\mathrm{d}t\frac{P(60 < T \leq 120|\Omega \cap \{T > 60\})}{P(60 < T \leq 120|\Omega)} \\
&= \frac{\exp\left(11\right) - \exp\left(11/2\right)}{11}\frac{P(60 < T \leq 120|\Omega \cap \{T > 60\})}{P(60 < T \leq 120|\Omega)}.
\end{aligned}$$

Calcoliamo infine, usando la formula di Bayes,

$$\begin{aligned}
\frac{P(60 < T \leq 120|\Omega \cap \{T > 60\})}{P(60 < T \leq 120|\Omega)} &= \frac{P(T > 60|\Omega \cap \{60 < T < 120\} P(60 < T \leq 120|\Omega)}{P(60 < T \leq 120|\Omega)P(T > 60|\Omega)} \\
&= \frac{P(T > 60|\Omega \cap \{60 < T < 120\}}{P(T > 60|\Omega)} = \frac{1}{P(T > 60|\Omega)} = e^{1/2}.
\end{aligned}$$

Quindi vale

$$\mathbb{E}\left[C|\Omega\cap\{T>60\}\right]=\frac{\exp(11)-\exp(11/2)}{11}e^{1/2}.$$

Problema 37. Il tempo di vita, in anni, di un componente elettronico è un una variabile aleatoria continua $T\in[1,10]$ di densità:

$$\varrho(T=t|\Omega)=\begin{cases}\frac{k}{t^4} & \text{se } t\in[1,10]\\ 0 & \text{altrimenti}\end{cases}$$

(1) Trovare la costante k. Calcolare valore atteso e varianza di T.
(2) Calcolare la funzione di ripartizione e la funzione di sopravvivenza di T (rispetto a $P(\cdot|\Omega)$)
(3) Calcolare la probabilità per il tempo di vita di superare 2 anni.
(4) Sapendo che il componente ha vissuto per più di 2 anni, calcolare la densità di T, il valore atteso di T e la varianza.

Una soluzione. (1) La costante k deve essere tale per cui

$$1=P(T\in[1,10]|\Omega)=\int_1^{10}kt^{-4}\mathrm{d}t=k\frac{t^{-3}}{-3}\Big|_1^{10}=k\frac{1-10^{-3}}{3}.$$

Perciò $k=3/(1-10^{-3})\approx 3$. Calcoliamo valore atteso e varianza di T riconducendoci ad integrali:

$$\mathbb{E}\left[T|\Omega\right]=\int_1^{10}tkt^{-4}\mathrm{d}t=k\frac{t^{-2}}{-2}\Big|_1^{10}=k\frac{1-10^{-2}}{2}\approx\frac{k}{2},$$

mentre per la varianza, troviamo prima

$$\mathbb{E}\left[T^2|\Omega\right]=\int_1^{10}t^2kt^{-4}\mathrm{d}t=k\frac{t^{-1}}{-1}\Big|_1^{10}=k\frac{1-10^{-1}}{1}=k0,9$$

e poi calcoliamo

$$\mathrm{Var}\left(T|\Omega\right)=\mathbb{E}\left[T^2|\Omega\right]-\left(\mathbb{E}\left[T|\Omega\right]\right)^2=k0,9-\frac{k^2}{4}\approx 0,45,$$

da cui $\sigma(T|\Omega)\approx\sqrt{0,45}=0.67$.

(2) Calcoliamo, per $t\in[1,10]$,

$$P(T\le t|\Omega)=\int_1^t ks^{-4}\mathrm{d}s=k\frac{s^{-3}}{-3}\Big|_1^t=\frac{k}{3}(1-t^{-3})\approx 1-t^{-3},$$

mentre la funzione di sopravvivenza di T è

$$P(T>t|\Omega)=1-P(T\le t|\Omega)=1-\frac{k}{3}(1-t^{-3})\approx t^{-3}.$$

(3) Usando il punto precedente, troviamo $P(T>2|\Omega)=1-\frac{k}{3}(1-s^{-3})\approx 1/8$.

(4) Dobbiamo rifare i calcoli del punto (1) rispetto però all'informazione $\Omega\cap\{T>2\}$. Usiamo quindi il Lemma 10, per trovare la densità, per $t\in[2,10]$ (altrove la densità vale 0)

$$\varrho(T=t|\Omega\cap\{T>2\})=kt^{-4}/P(T>2|\Omega),$$

dove $P(T>2|\Omega)$ si calcola usando il punto sopra.

Problema 38. Un albergo ha due prenotazioni per il pomeriggio, ciascuna indicante un tempo di arrivo tra le 14:00 e le 20:00 (*supponiamo quindi che i tempi di arrivo T_1 e T_2 siano variabili indipendenti con legge uniforme*). Calcolare

(1) la densità del tempo di arrivo del primo cliente (in ordine di arrivo), il valore atteso e la varianza;
(2) la densità del tempo di arrivo del secondo cliente (in ordine di arrivo), il valore atteso e la varianza;
(3) la probabilità che il secondo cliente (in ordine di arrivo) arrivi dopo le 15:00, sapendo che il primo cliente (in ordine di arrivo) è arrivato prima delle 14:30.
(4) (*difficile*) la probabilità che il secondo cliente (in ordine di arrivo) arrivi dopo le 15:00, sapendo che il primo cliente (in ordine di arrivo) è arrivato esattamente alle 14:30.

Una soluzione. Misuriamo il tempo in ore (e frazioni di ora) a partire dalle 14:00, così che T_1 e T_2 sono variabili continue uniformi a valori in $[0,6]$ (entrambe).

(1) Per calcolare la densità del tempo di arrivo del primo cliente, ossia di $F := \min\{T_1, T_2\}$, calcoliamo prima la funzione di sopravvivenza di F, per $t \in [0,6]$:

$$\begin{aligned} P(F > t|\Omega) &= P\left(\min\{T_1, T_2\} > t|\Omega\right) \\ &= P\left(T_1 > t \quad \text{e} \quad T_2 > t|\Omega\right) \\ &= P\left(T_1 > t|\Omega\right) P\left(T_2 > t|\Omega\right) \\ &= \left(\int_t^6 \frac{ds}{6}\right)^2 = \left(1 - \frac{t}{6}\right)^2. \end{aligned}$$

A questo punto otteniamo la densità calcolando

$$-\frac{d}{dt} P(F > t|\Omega) = \frac{1}{3}\left(1 - \frac{t}{6}\right).$$

Calcoliamo il valore atteso di F:

$$\mathbb{E}\left[F|\Omega\right] = \int_0^6 t \cdot \frac{1}{3}\left(1 - \frac{t}{6}\right) \mathrm{d}t = \frac{1}{3} \cdot \frac{6^2}{2} - \frac{1}{18}\frac{6^3}{3} = 6 - 4 = 2.$$

Calcoliamo la varianza di F:

$$\mathbb{E}\left[F^2|\Omega\right] = \int_0^6 t^2 \cdot \frac{1}{3}\left(1 - \frac{t}{6}\right) \mathrm{d}t = \frac{1}{3} \cdot \frac{6^3}{3} - \frac{1}{18}\frac{6^4}{4} = 6,$$

da cui

$$\mathrm{Var}\left(F|\Omega\right) = 6 - 4 = 2.$$

(2) Nel caso di $L := \max\{T_1, T_2\}$ lavoriamo con la funzione di ripartizione, per $t \in [0,6]$,

$$\begin{aligned} P(L \leq t|\Omega) &= P\left(\max\{T_1, T_2\} \leq t|\Omega\right) \\ &= P\left(T_1 \leq t \quad \text{e} \quad T_2 \leq t|\Omega\right) \\ &= P\left(T_1 \leq t|\Omega\right) P\left(T_2 \leq t|\Omega\right) \\ &= \left(\int_0^t \frac{ds}{6}\right)^2 = \left(\frac{t}{6}\right)^2. \end{aligned}$$

A questo punto otteniamo la densità calcolando

$$\frac{d}{dt}P(L \leq t|\Omega) = \frac{1}{3}\frac{t}{6}.$$

Calcoliamo il valore atteso di L:

$$\mathbb{E}\left[L|\Omega\right] = \int_0^6 t \cdot \frac{1}{3}\frac{t}{6}\mathrm{d}t = \frac{1}{18}\frac{6^3}{3} = 4.$$

Calcoliamo la varianza di L:

$$\mathbb{E}\left[L^2|\Omega\right] = \int_0^6 t^2 \cdot \frac{1}{3}\frac{t}{6}\mathrm{d}t = \frac{1}{18}6^4 4 = 18,$$

da cui

$$\mathrm{Var}\left(L|\Omega\right) = 18 - 4^2 = 2.$$

(3) Dobbiamo calcolare $P(L > 1|F < 1/2)$. Qui si può procedere in tanti modi diversi. Una possibilità (forse la via più breve) è di calcolare

$$\begin{aligned}
P(L > 1|F < 1/2) &= 1 - P(L \leq 1|F < 1/2) = 1 - P(F < 1/2|L \leq 1)\frac{P(L \leq 1|\Omega)}{P(F < 1/2|\Omega)} \\
&= 1 - \left(1 - P(F \geq 1/2|L \leq 1)\right)\frac{P(L \leq 1|\Omega)}{P(F < 1/2|\Omega)} \\
&= 1 - \left(1 - \frac{P(F \geq 1/2 \quad \text{e} \quad L \leq 1|\Omega)}{P(L \leq 1|\Omega)}\right)\frac{P(L \leq 1|\Omega)}{P(F < 1/2|\Omega)}.
\end{aligned}$$

Questa identità ci richiede solamente di calcolare

$$P(L \leq 1|\Omega) = \int_0^1 \frac{1}{3}\frac{t}{6}\mathrm{d}t = \frac{1}{18}\cdot\frac{1}{2} = \frac{1}{36},$$

$$P(F < 1/2|\Omega) = \int_0^{1/2} \frac{1}{3}\left(1 - \frac{t}{6}\right)\mathrm{d}t = \frac{1}{6} - \frac{1}{18}\cdot\frac{1}{8},$$

e infine

$$\begin{aligned}
P(F \geq 1/2 \quad \text{e} \quad L \leq 1|\Omega) &= P(1/2 \leq X_1 \leq 1 \quad \text{e} \quad 1/2 \leq X_2 \leq 1|\Omega) \\
&= \left(\int_{1/2}^1 \frac{\mathrm{d}t}{6}\right)^2 = \frac{1}{12^2}.
\end{aligned}$$

Troviamo allora

$$P(L > 1|F < 1/2) = 1 - \left(1 - \frac{3 \cdot 12}{12 \cdot 12}\right)\frac{1}{36\left(1/6 - 1/(18 * 8)\right)} = \frac{20}{23}.$$

(4) Dobbiamo calcolare $P(L > 1|F = 1/2)$. Anche in questo caso possiamo procedere in diversi modi. Una possibilità è di ripetere parzialmente quanto fatto sopra (usando la formula di Bayes, caso discreto/continuo) per trovare

$$P(L > 1|F = 1/2) = 1 - P(L \leq 1|F = 1/2) = 1 - \varrho(F = 1/2|L \leq 1)\frac{P(L \leq 1|\Omega)}{\varrho\left(F = 1/2|\Omega\right)}.$$

Abbiamo già visto che

$$P(L \leq 1|\Omega) = \frac{1}{36}, \quad \varrho(F = 1/2|\Omega) = \frac{1}{3}\left(1 - \frac{1/2}{6}\right) = \frac{11}{36}.$$

Ci rimane da calcolare solamente $\varrho(F = 1/2|L \leq 1)$, per farlo calcoliamo prima la funzione di sopravvivenza di F rispetto all'informazione $\Omega \cap \{L \leq 1\}$. Notiamo che, rispetto a questa informazione, siccome $F \leq L$, abbiamo con probabilità 1, $F \in [0,1]$, quindi calcoliamo solamente la funzione di sopravvivenza per $t \in [0,1]$:

$$\begin{aligned} P(F > t|L \leq 1) &= \frac{P(F > t \quad \text{e} \quad L \leq 1|\Omega)}{P(L \leq 1|\Omega)} \\ &= \frac{P(t < X_1 \leq 1 \quad \text{e } t < X_2 \leq 1|\Omega)}{P(L \leq 1|\Omega)} \\ &= 36 \left(\int_t^1 \frac{\mathrm{d}s}{6} \right)^2 = (1-t)^2. \end{aligned}$$

Derivando rispetto a t otteniamo la densità:

$$\varrho(F = t|L \leq 1) = -\frac{d}{dt} P(F > t|L \leq 1) = 2(1-t).$$

Ponendo $t = 1/2$, concludiamo che

$$\varrho(F = 1/2|L \leq 1) = 1,$$

e quindi

$$P(L > 1|F = 1/2) = 1 - \frac{1/36}{11/36} = 1 - \frac{1}{11} = \frac{10}{11}.$$

Problema 39. Un'urna (non vuota) contiene palline blu o rosse, e poniamo $X \in [0,1]$ la frazione di palline rosse sul numero totale (neppure il numero totale è noto, quindi supponiamo che X sia uniforme, basandoci solamente su questa informazione iniziale).

(1) Calcolare la densità di X, il valore atteso e la varianza.
(2) Si estrae una pallina e si vede che è rossa. Calcolare la densità di X, il valore atteso e la varianza.
(3) Dopo aver estratto una pallina e visto che è rossa, la rimettiamo nell'urna e mescoliamo. Qual è la probabilità di estrarre nuovamente una pallina rossa?
(4) Dopo aver estratto una pallina e visto che è rossa, averla rimessa nell'urna e mescolato, estraiamo una seconda pallina e vediamo che è blu. Calcolare la densità di X, il valore atteso e la varianza.
(5) Supponiamo di aver fatto in sequenza $n \geq 1$ estrazioni (con reimmissione) e di aver visto solamente palline rosse. Qual è la densità di X, il valore atteso e la sua varianza? Qual è la probabilità di estrarre nuovamente una pallina rossa alla $(n+1)$-esima estrazione?

Una soluzione. (1) È dato nel testo che la densità di X rispetto all'informazione iniziale Ω è uniforme su $[0,1]$, quindi

$$\varrho(X = x|\Omega) = 1 \quad \text{per } x \in [0,1],$$

$$\mathbb{E}\left[X|\Omega\right] = \frac{1}{2} \quad \mathrm{Var}\left(X|\Omega\right) = \frac{1}{12}.$$

(2) Poniamo A l'evento "prima pallina estratta è rossa". Dobbiamo calcolare, per $x \in [0,1]$,

$$\varrho(X = x|\Omega \cap A) = P(A|\Omega \cap \{X = x\})\frac{\varrho(X = x|\Omega)}{P(A|\Omega)},$$

avendo usato Bayes continuo/discreto. Calcoliamo separatamente

$$P(A|\Omega \cap \{X = x\}) = x,$$

perché si tratta di estrarre da un'urna di cui sappiamo che la frazione di palline rosse è x,

$$\varrho(X = x|\Omega) = 1$$

perché X è uniforme (rispetto a Ω) e infine (come è anche scritto nell'enunciato della formula di Bayes, che stiamo usando)

$$P(A|\Omega) = \int_0^1 P(A|\Omega \cap \{X = x\})\varrho(X = x|\Omega)\mathrm{d}x = \int_0^1 x\mathrm{d}x = \frac{1}{2}.$$

Troviamo

$$\varrho(X = x|\Omega \cap A) = 2x.$$

Il valore atteso vale

$$\mathbb{E}\left[X|\Omega \cap A\right] = \int_0^1 x2x\mathrm{d}x = \frac{2}{3},$$

mentre

$$[\mathbb{E}\left[X^2|\Omega \cap A\right] = \int_0^1 x^2 2x\mathrm{d}x = \frac{2}{4} = \frac{1}{2},$$

e quindi la varianza vale

$$\mathrm{Var}\left(X|\Omega \cap A\right) = \frac{1}{2} - \left(\frac{2}{3}\right)^2 = \frac{1}{18}.$$

(3) Poniamo B l'evento "seconda pallina estratta è rossa". Dobbiamo calcolare, usando l'evento A definito prima,

$$P(B|\Omega \cap A) = \int_0^1 P(B|\Omega \cap A \cap \{X = x\})\varrho(X = x|\Omega \cap A)\mathrm{d}x,$$

dove abbiamo "decomposto" rispetto ai possibili contenuti dell'urna (usando la probabilità condizionale). D'altra parte, se conosciamo la frazione di palline rosse $X = x$, allora A e B sono indipendenti (solito schema di estrazioni con reimmissione, prima parte del corso) quindi

$$P(B|\Omega \cap A \cap \{X = x\}) = P(B|\Omega \cap \{X = x\}) = x.$$

Però abbiamo anche $\varrho(X = x|\Omega \cap A) = 2x$, quindi

$$P(B|\Omega \cap A) = \int_0^1 x2x\mathrm{d}x = \frac{2}{3},$$

e non $1/2$ come uno potrebbe aspettarsi: il fatto è che la prima estrazione ci permette di "imparare" qualcosa sul contenuto dell'urna (su cui siamo incerti). (4) Dobbiamo calcolare

$$
\begin{aligned}
\varrho(X=x|\Omega\cap A\cap B^c) &= P\left(B^c|\Omega\cap A\cap\{X=x\}\right)\frac{\varrho(X=x|\Omega\cap A)}{P(B^c|\Omega\cap A)}\\
&= P\left(B^c|\Omega\cap\{X=x\}\right)\frac{\varrho(X=x|\Omega\cap A)}{1-P(B|\Omega\cap A)}\\
&= (1-x)\frac{2x}{1/3} = 6x(1-x).
\end{aligned}
$$

Calcoliamo valore atteso e varianza:

$$
\mathbb{E}\left[X|\Omega\cap A\cap B^c\right] = \int_0^1 x6x(1-x)\mathrm{d}x = 6\left(\frac{x^3}{3}-\frac{x^4}{4}\right)\Big|_0^1 = 6\left(\frac{1}{3}-\frac{1}{4}\right) = \frac{1}{2},
$$

$$
\mathbb{E}\left[X^2|\Omega\cap A\cap B^c\right] = \int_0^1 x^2 6x(1-x)\mathrm{d}x = 6\left(\frac{x^4}{4}-\frac{x^5}{5}\right)\Big|_0^1 = 6\left(\frac{1}{4}-\frac{1}{5}\right) = \frac{3}{10},
$$

$$
\mathrm{Var}\left(X|\Omega\cap A\cap B^c\right) = \frac{3}{10}-\frac{1}{4} = \frac{1}{20}.
$$

(5) Sia C l'evento "n palline rosse estratte in n estrazioni", e calcoliamo la densità di X sapendo C:

$$
\varrho(X=x|\Omega\cap C) = P(C|\Omega\cap\{X=x\})\frac{\varrho(X=x|\Omega)}{P(C|\Omega)}.
$$

Abbiamo

$$
P(C|\Omega\cap\{X=x\}) = x^n,
$$

(estrazioni con reimmersione da un'urna il cui contenuto è noto), mentre

$$
P(C|\Omega) = \int_0^1 P(C|\Omega\cap\{X=x\})\varrho(X=x|\Omega)\mathrm{d}x = \int_0^1 x^n\mathrm{d}x = \frac{1}{n+1}.
$$

Di conseguenza, per $x\in[0,1]$,

$$
\varrho(X=x|\Omega\cap C) = (n+1)x^n.
$$

Calcoliamo valore atteso e varianza

$$
\mathbb{E}\left[X|\Omega\cap C\right] = \int_0^1 x(n+1)x^n\mathrm{d}x = \frac{n+1}{n+2},
$$

$$
\mathbb{E}\left[X^2|\Omega\cap C\right] = \int_0^1 x^2(n+1)x^n\mathrm{d}x = \frac{n+1}{n+3},
$$

$$
\mathrm{Var}\left(X|\Omega\cap C\right) = \frac{n+1}{n+3}-\left(\frac{n+1}{n+2}\right)^2 = \frac{n+1}{(n+2)^2(n+3)}.
$$

Calcoliamo infine la probabilità di estrarre una pallina rossa alla $n+1$-esima estrazione, sapendo C: posto D l'evento "$n+1$-esima pallina estratta è rossa, abbiamo

$$
P(D|\Omega\cap C) = \int_0^1 P(D|\Omega\cap C\cap\{X=x\})\varrho(X=x|\Omega\cap C)\mathrm{d}x = \int_0^1 x(n+1)x^n\mathrm{d}x = \frac{n+1}{n+2}.
$$

Problema 40. È noto che tempo di vita (in anni) di un componente elettronico senza difetti è modellizzato come una variabile aleatoria T con legge esponenziale di parametro $1/2$. Tuttavia, è noto che vi sono anche componenti difettosi, il cui tempo di vita T ha legge uniforme sull'intervallo $[0,2]$. Sappiamo anche che il 75% dei componenti sono senza difetti, ma il rimanente 25% è difettoso. Ci viene recapitato uno di questi componenti.

(1) Calcolare valore atteso e varianza del tempo di vita del componente.
(2) Calcolare la probabilità che il suo tempo di vita sia maggiore di 1 anno.
(3) Dopo un anno, il componente continua a funzionare: qual è la probabilità che sia un componente senza difetti? Calcolare il valore atteso e la varianza del tempo di vita rimanente.
(4) Sapendo che esattamente dopo un anno il componente si è rotto, qual è la probabilità che fosse un componente difettoso?

Una soluzione. Poniamo D l'evento "il pezzo ricevuto è difettoso", (e D^c significa che il pezzo è sano). (1) Calcoliamo

$$\begin{aligned}\mathbb{E}\left[T|\Omega\right] &= \mathbb{E}\left[T|D\right]P(D|\Omega) + \mathbb{E}\left[T|D^c\right]P(D^c|\Omega)\\ &= \mathbb{E}\left[T|D\right]\frac{1}{4} + \mathbb{E}\left[T|D^c\right]\frac{3}{4}\\ &= 1\cdot\frac{1}{4} + 2\frac{3}{4} = \frac{7}{4}\end{aligned}$$

avendo usato il fatto che la densità di T rispetto a D è uniforme su $[0,1]$, e rispetto a D^c è esponenziale di parametro $1/2$. Calcoliamo in modo simile,

$$\begin{aligned}\mathbb{E}\left[T^2|\Omega\right] &= \mathbb{E}\left[T^2|D\right]P(D|\Omega) + \mathbb{E}\left[T^2|D^c\right]P(D^c|\Omega)\\ &= \mathbb{E}\left[T^2|D\right]\frac{1}{4} + \mathbb{E}\left[T^2|D^c\right]\frac{3}{4}\\ &= \left(\operatorname{Var}(T|D) + \mathbb{E}\left[T|D\right]^2\right)\frac{1}{4} + \left(\operatorname{Var}(T|D^c) + \mathbb{E}\left[T|D^c\right]^2\right)\frac{3}{4}\\ &= \left(\frac{1}{3}+1\right)\frac{1}{4} + (4+4)\frac{3}{4} = \frac{19}{3}.\end{aligned}$$

e infine

$$\operatorname{Var}(T|\Omega) = \frac{19}{3} - \left(\frac{7}{4}\right)^2 \approx 3,27.$$

(2) Calcoliamo

$$\begin{aligned}P(T>1|\Omega) &= P(T>1|D)P(D|\Omega) + P(T>1|D^c)P(D^c|\Omega)\\ &= \int_1^2 \frac{\mathrm{d}x}{2}\cdot\frac{1}{4} + e^{-1/2}\frac{3}{4} = \frac{1}{2}\frac{1}{4} + e^{-1/2}\frac{3}{4} \approx 0,28.\end{aligned}$$

(3) Dobbiamo calcolare

$$P(D^c|T>1) = P(T>1|D^c)\frac{P(D^c|\Omega)}{P(T>1|\Omega)} = \frac{e^{-1/2}\frac{3}{4}}{\frac{1}{2}\frac{1}{4} + e^{-1/2}\frac{3}{4}} \approx 0,55.$$

Il tempo di vita rimanente è $T-1$. Calcoliamone il valore atteso:

$$\begin{aligned}\mathbb{E}\left[(T-1)|T>1\right] &= \mathbb{E}\left[(T-1)|\left\{T>1\right\}\cap D\right]P(D|T>1)\\ &\quad+\mathbb{E}\left[(T-1)|\left\{T>1\right\}\cap D^c\right]P(D^c|T>1)\\ &= \mathbb{E}\left[(T-1)|\left\{T>1\right\}\cap\left\{T \text{ è unif su } [0,2]\right\}\right](1-P(D^c|T>1))\\ &\quad+\mathbb{E}\left[(T-1)|\left\{T>1\right\}\cap\left\{T \text{ è } \mathcal{E}(1/2)\right\}\right]P(D^c|T>1)\\ &= \int_1^2 (t-1)\mathrm{d}t(1-P(D^c|T>1))+2P(D^c|T>1)\\ &= \frac{1}{2}(1-P(D^c|T>1))+2P(D^c|T>1)\\ &\approx= \frac{1}{2}\cdot 0,45+2\cdot 0,55 = 1,33.\end{aligned}$$

(4) Calcoliamo

$$P(D|T=1) = \varrho\left(T=1|D\right)\frac{P(D|\Omega)}{\varrho(T=1|\Omega)}.$$

Sappiamo che

$$\varrho\left(T=1|D\right)=\frac{1}{2},$$

$$P(D|\Omega)=\frac{1}{4},$$

e infine troviamo

$$\begin{aligned}\varrho\left(T=1|\Omega\right) &= \varrho\left(T=1|D\right)P(D|\Omega)+\varrho\left(T=1|D^c\right)P(D^c|\Omega)\\ &= \frac{1}{2}\cdot\frac{1}{4}+\frac{1}{2}e^{-1/2}\frac{3}{4},\end{aligned}$$

perciò

$$P(D|T=1)=\frac{\frac{1}{2}\cdot\frac{1}{4}}{\frac{1}{2}\cdot\frac{1}{4}+\frac{1}{2}e^{-1/2}\frac{3}{4}}.$$

Problema 41. La posizione di un possibile pixel rotto su modello di schermo LCD è modellizata tramite due variabili aleatorie X, $Y\in[0,1]$ (ascissa e ordinata rispetto a un vertice dello schermo, per semplicità variabili continue). Ci sono due modelli che prevedono la posizione dei pixel rotti: uno (modello U) prevede che X e Y siano variabili aleatorie indipendenti, ciascuna con legge uniforme di parametri $[0,1]$. Un altro (modello F) prevede invece sempre che X e Y siano indipendenti, con densità

$$\varrho(X=z|\Omega\cap F)=\varrho(Y=z|\Omega\cap F)=c\left(\frac{1}{z^{1/2}}+\frac{1}{(1-z)^{1/2}}\right).$$

Supponiamo di dare inizialmente probabilità uniforme alla validità dei due modelli (quindi ciascuno ha probabilità $1/2$, rispetto alla informazione iniziale Ω).

(1) Calcolare c affinché la funzione sopra sia una densità di probabilità. Calcolare valore atteso e varianza delle coordinate X, Y previste dal modello F.
(2) Calcolare valore atteso e varianza delle coordinate X, Y rispetto alla informazione iniziale.

(3) Supponiamo di osservare un pixel rotto in uno schermo, in posizione $X = 1/2$, $Y = 1/2$. Come cambia la probabilità della validità dei due modelli?
(4) Supponiamo di osservare un pixel rotto in uno schermo, in posizione $X = 1/100$, $Y = 1/100$. Come cambia la probabilità della validità dei due modelli?

Una soluzione. (1) Dobbiamo imporre che

$$\int_0^1 c\left(\frac{1}{z^{1/2}} + \frac{1}{(1-z)^{1/2}}\right) \mathrm{d}z = 1.$$

Notiamo che

$$\int_0^1 \frac{1}{z^{1/2}}\mathrm{d}z = \left.\frac{z^{1/2}}{1/2}\right|_0^1 = 2,$$

e pure, con il cambio di variabile $y = 1 - z$,

$$\int_0^1 \frac{1}{(1-z)^{1/2}}\mathrm{d}z = \int_0^1 \frac{1}{y^{1/2}}\mathrm{d}y = 2,$$

quindi deve essere $c = 1/4$. Calcoliamo valore atteso di X se vale il modello F:

$$\begin{aligned}
\mathbb{E}\left[X|\Omega \cap F\right] &= \int_0^1 z\frac{1}{4}\left(\frac{1}{z^{1/2}} + \frac{1}{(1-z)^{1/2}}\right)\mathrm{d}z \\
&= \frac{1}{4}\left(\int_0^1 z^{1/2}\mathrm{d}z + \int_0^1 \frac{1-y}{y^{1/2}}\mathrm{d}y\right) \quad \text{cambio di variabile } y = 1-z \\
&= \frac{1}{4}\left(\frac{1}{3/2} + \frac{1}{1/2} - \frac{1}{3/2}\right) = \frac{1}{2}.
\end{aligned}$$

A questo risultato si poteva arrivare facilmente notando che la densità $c\left(\frac{1}{z^{1/2}} + \frac{1}{(1-z)^{1/2}}\right)$ è *simmetrica* rispetto all'asse $x = 1/2$. Calcoliamo

$$\begin{aligned}
\mathbb{E}\left[X^2|\Omega \cap F\right] &= \int_0^1 z^2\frac{1}{4}\left(\frac{1}{z^{1/2}} + \frac{1}{(1-z)^{1/2}}\right)\mathrm{d}z \\
&= \frac{1}{4}\left(\int_0^1 z^{3/2}\mathrm{d}z + \int_0^1 \frac{(1-y)^2}{y^{1/2}}\mathrm{d}y\right) \quad \text{cambio di variabile } y = 1-z \\
&= \frac{1}{4}\left(\int_0^1 z^{3/2}\mathrm{d}z + \int_0^1 \left(\frac{1}{y^{1/2}} - 2y^{1/2} + y^{3/2}\right)\mathrm{d}y\right) \\
&= \frac{1}{4}\left(\frac{1}{5/2} + \frac{1}{1/2} - 2\frac{1}{3/2} + \frac{1}{5/2}\right) \approx 0,37.
\end{aligned}$$

e quindi

$$\mathrm{Var}\left(X|\Omega \cap F\right) \approx 0,37 - \frac{1}{4} = 0,12.$$

I calcoli per la coordinata Y sono identici. (2) Rispetto alla informazione iniziale, abbiamo

$$\mathbb{E}\left[X|\Omega\right] = \mathbb{E}\left[X|\Omega \cap F\right]\frac{1}{2} + \mathbb{E}\left[X|\Omega \cap U\right]\frac{1}{2} = \frac{1}{2},$$

mentre

$$\mathbb{E}\left[X^2|\Omega\right] = \mathbb{E}\left[X^2|\Omega\cap F\right]\frac{1}{2} + \mathbb{E}\left[X^2|\Omega\cap U\right]\frac{1}{2} \approx 0,37\frac{1}{2} + \frac{1}{3}\frac{1}{2} = 0,35.$$

e quindi

$$\operatorname{Var}(X|\Omega) = \mathbb{E}\left[X^2|\Omega\right] - \mathbb{E}\left[X|\Omega\right]^2 \approx 0,35 - \frac{1}{4} = 0,10.$$

(3) Dobbiamo applicare la formula di Bayes, aggiornando l'informazione Ω a $\Omega\cap\{X=1/2\}\cap\{Y=1/2\}$. Proviamo a farlo "aggiornando" un passo alla volta, ossia calcoliamo prima

$$\begin{aligned} P\left(U|\Omega\cap\{X=1/2\}\right) &= \varrho(X=1/2|\Omega\cap U)\frac{P(U|\Omega)}{\varrho(X=1/2|\Omega)} \\ &= 1\frac{1}{1+\frac{1}{4}\left(\left(\frac{1}{2}\right)^{-1/2}+\left(\frac{1}{2}\right)^{-1/2}\right)} \\ &= \frac{1}{1+\frac{\sqrt{2}}{2}} \approx 0,58. \end{aligned}$$

Di conseguenza,

$$P\left(F|\Omega\cap\{X=1/2\}\right) = 1 - P\left(U|\Omega\cap\{X=1/2\}\right) \approx 0,42.$$

Ora calcoliamo (ancora usando Bayes discreto/continuo)

$$P\left(U|\{X=1/2\}\cap\{Y=1/2\}\right) = \varrho\left(Y=1/2|\{X=1/2\}\cap U\right)\frac{P(U|\{X=1/2\})}{\varrho(Y=1/2|\{X=1/2\})},$$

dove

$$\begin{aligned} \varrho(Y=1/2|\Omega\cap\{X=1/2\}) &= \varrho(Y=1/2|\Omega\cap\{X=1/2\}\cap U)P(U|\Omega\cap\{X=1/2\}) \\ &\quad + \varrho(Y=1/2|\Omega\cap\{X=1/2\}\cap F)P(F|\Omega\cap\{X=1/2\}). \end{aligned}$$

Se vale il modello U oppure il modello F, allora X e Y sono indipendenti, quindi

$$\varrho\left(Y=1/2|\Omega\cap\{X=1/2\}\cap U\right) = \varrho\left(Y=1/2|\Omega\cap U\right) = 1,$$

$$\varrho(Y=1/2|\Omega\cap\{X=1/2\}\cap F) = \varrho\left(Y=1/2|\Omega\cap F\right) = \frac{\sqrt{2}}{2}.$$

Di conseguenza abbiamo

$$\varrho(Y=1/2|\Omega\cap\{X=1/2\}) = 1\cdot 0,58 + \frac{\sqrt{2}}{2}\cdot 0,42 \approx 0,88$$

e

$$P\left(U|\Omega\cap\{X=1/2\}\cap\{Y=1/2\}\right) \approx \frac{0,58}{0,88} \approx 0,66.$$

(4) Si tratta di ripetere i calcoli di prima, stavolta aggiornando l'informazione Ω a $\Omega\cap\{X=1/100\}\cap\{Y=1/100\}$. Proviamo a farlo "aggiornando"

un passo alla volta, ossia calcoliamo prima

$$
\begin{aligned}
P\left(U|\Omega \cap \{X = 1/100\}\right) &= \varrho(X = 1/100|\Omega \cap U)\frac{P(U|\Omega)}{\varrho(X = 1/100|\Omega)} \\
&= 1\frac{1}{1 + \frac{1}{4}\left(\left(\frac{1}{100}\right)^{-1/2} + \left(\frac{99}{100}\right)^{-1/2}\right)} \\
&\approx \frac{1}{1 + \frac{1}{4}\left(10 + 1\right)} = 0,27.
\end{aligned}
$$

Di conseguenza,

$$
P\left(F|\Omega \cap \{X = 1/100\}\right) = 1 - P\left(U|\Omega \cap \{X = 1/100\}\right) \approx 0,73.
$$

Ora calcoliamo (ancora usando Bayes discreto/continuo)

$$
P\left(U|\left\{X = 1/2\right\} \cap \{Y = 1/100\}\right) = \varrho\left(Y = 1/100|\left\{X = 1/2\right\} \cap U\right)\frac{P(U|\left\{X = 1/100\right\})}{\varrho(Y = 1/100|\left\{X = 1/100\right\})},
$$

dove

$$
\begin{aligned}
\varrho(Y = 1/100|\Omega \cap \{X = 1/100\}) = {}& \varrho(Y = 1/100|\Omega \cap \{X = 1/100\} \cap U)P(U|\Omega \cap \{X = 1/100\}) \\
&+ \varrho(Y = 1/100|\Omega \cap \{X = 1/100\} \cap F)P(F|\Omega \cap \{X = 1/100\}).
\end{aligned}
$$

Se vale il modello U oppure il modello F, allora X e Y sono indipendenti, quindi

$$
\begin{aligned}
\varrho\left(Y = 1/100|\Omega \cap \{X = 1/100\} \cap U\right) &= \varrho\left(Y = 1/100|\Omega \cap U\right) = 1, \\
\varrho(Y = 1/100|\Omega \cap \{X = 1/100\} \cap F) &= \varrho\left(Y = 1/100|\Omega \cap F\right) \approx 2,75.
\end{aligned}
$$

Di conseguenza abbiamo

$$
\varrho(Y = 1/2|\Omega \cap \{X = 1/2\}) \approx 1 \cdot 0,27 + 2,75 \cdot 0,42 \approx 2,28
$$

e

$$
P\left(U|\Omega \cap \{X = 1/2\} \cap \{Y = 1/2\}\right) \approx \frac{0,27}{2,28} \approx 0,12.
$$

Problema 42. Il prezzo di un'azione di una certa azienda (in un istante futuro t_1) può essere ben rappresentato (al tempo $t_0 < t_1$) secondo alcuni studiosi dei mercati da una variabile della forma $C = \exp(X)$, dove $X \in \mathbb{R}$ è una variabile gaussiana di parametri $\mathcal{N}(m, \sigma^2)$.

(1) Calcolare il valore atteso e la varianza di C.
(2) Calcolare la densità di C.
(3) Gli studiosi sanno che $\sigma^2 = 1$ è un parametro realistico, mentre sono totalmente indecisi sul fatto che il modello sia meglio adattato alla realtà ponendo $m = 1$ oppure $m = 2$. Supponendo che, arrivato il momento t_1 il prezzo osservato risulti 3, quale scelta del parametro è (era) più opportuna? con quale probabilità?

Una soluzione. (1) Calcoliamo

$$
\begin{aligned}
\mathbb{E}\left[C|\Omega\right] = \mathbb{E}\left[\exp(X)|X \text{ è } \mathcal{N}(m, \sigma^2)\right] &= \int_{\mathbb{R}} \exp(x)\exp\left(-\frac{1}{2}\left(\frac{x - m}{\sigma}\right)^2\right)\frac{\mathrm{d}x}{\sqrt{2\pi\sigma^2}} \\
&= \int_{\mathbb{R}} \exp\left(x - \frac{1}{2}\left(\frac{x - m}{\sigma}\right)^2\right)\frac{\mathrm{d}x}{\sqrt{2\pi\sigma^2}}.
\end{aligned}
$$

Concentriamoci sull'esponente

$$\begin{aligned}
x - \frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2 &= -\frac{1}{2\sigma^2}\left(-2\sigma^2 x + x^2 - 2xm + m^2\right)\\
&= -\frac{1}{2\sigma^2}\left[\left(x-(m+\sigma^2)\right)^2 + m^2 - (m+\sigma^2)^2\right]\\
&= -\frac{1}{2\sigma^2}\left[\left(x-(m+\sigma^2)\right)^2 - 2\sigma^2 m - \sigma^4\right]\\
&= -\frac{1}{2\sigma^2}\left(x-(m+\sigma^2)\right)^2 + m + \frac{\sigma^2}{2}.
\end{aligned}$$

Questa identità algebrica ci permette di scrivere

$$\begin{aligned}
&\int_{\mathbb{R}} \exp\left(x - \frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2\right)\frac{\mathrm{d}x}{\sqrt{2\pi\sigma^2}} =\\
&\quad = e^{m+\frac{\sigma^2}{2}} \int_{\mathbb{R}} \exp\left(-\frac{1}{2\sigma^2}\left(x-(m+\sigma^2)\right)^2\right)\frac{\mathrm{d}x}{\sqrt{2\pi\sigma^2}}\\
&\quad = e^{m+\frac{\sigma^2}{2}}.
\end{aligned}$$

Abbiamo quindi trovato che

$$\mathbb{E}\left[\exp(X)|X \text{ è } \mathcal{N}(m,\sigma^2)\right] = e^{m+\frac{\sigma^2}{2}}.$$

Questo fatto può essere usato direttamente per calcolare la varianza di C: infatti $(\exp(X))^2 = \exp(2X)$, e $2X$ è $\mathcal{N}(2m, 4\sigma^2)$, quindi

$$\mathbb{E}\left[\exp(2X)|X \text{ è } \mathcal{N}(m,\sigma^2)\right] = e^{2m+2\sigma^2}.$$

e allora

$$\operatorname{Var}\left(\exp(X)|X \text{ è } \mathcal{N}(m,\sigma^2)\right) = e^{2m+2\sigma^2} - e^{2m+\sigma^2} = e^{2m+\sigma^2}(e^{\sigma^2}-1).$$

(2) Per calcolare la densità di C, ragioniamo come nella dimostrazione del Lemma 11: dato un intervallo $[a,b] \subseteq (0,\infty)$ (perché $C = \exp(X) \in (0,\infty)$), osserviamo che

$$\begin{aligned}
P(a \le C \le b|\Omega) &= P(a \le \exp(X) \le b|\Omega)\\
&= P(\log(a) \le X \le \log(b)|\Omega)\\
&= \int_{\log(a)}^{\log(b)} \varrho(X = x|\Omega)\mathrm{d}x\\
&= \int_a^b \varrho(X = \log(y)|\Omega)\frac{1}{y}\mathrm{d}y \quad \text{cambio di variabile } y = \exp(x).
\end{aligned}$$

Ricordando la densità di una gaussiana $\mathcal{N}(m,\sigma^2)$, otteniamo, per $y \in (0,\infty)$,

$$\varrho(C = y) = \varrho(X = \log(y)|\Omega)\frac{1}{y} = \exp\left(-\frac{1}{2}\left(\frac{\log(y)-m}{\sigma}\right)^2\right)\frac{1}{\sqrt{2\pi\sigma^2}y}.$$

(3) Si tratta di applicare la formula di Bayes,

$$P(\text{m=1}|C=3) = \varrho(C=3|m=1)\frac{P(m=1|\Omega)}{\varrho(C=3|\Omega)},$$

e calcolare

$$\varrho(C=3|m=1) = \exp\left(-\frac{1}{2}\left(\log(3)-1\right)^2\right)\frac{1}{\sqrt{2\pi}3}$$

$$\varrho(C=3|m=2) = \exp\left(-\frac{1}{2}\left(\log(3)-2\right)^2\right)\frac{1}{\sqrt{2\pi}3},$$

$$\varrho(C=3|\Omega) = \varrho(C=3|m=1)\frac{1}{2} + \varrho(C=3|m=2)\frac{1}{2},$$

da cui

$$P(\text{m=1}|C=3) = \frac{\exp\left(-\frac{1}{2}\left(\log(3)-1\right)^2\right)}{\exp\left(-\frac{1}{2}\left(\log(3)-1\right)^2\right)+\exp\left(-\frac{1}{2}\left(\log(3)-2\right)^2\right)} \approx 0,6.$$

Problema 43. Il tempo di vita di una resistenza elettrica è rappresentato da una variabile esponenziale di parametro $1/3$ (cioè finché funziona è un cavo che fa passare corrente dissipandone un po', quando è rotto invece non passa più nulla).

(1) Supponiamo di prendere n resistenze e collegarle in serie (cioè di collegarle una dietro l'altra). Qual è il tempo di vita della resistenza così costruita? calcolarne la legge, valore atteso e varianza.
(2) Supponiamo di prendere n resistenze e collegarle in parallelo (cioè di collegarne tutti gli estremi sinistri ad un punto del circuito e gli estremi destri ad un altro punto). Qual è il tempo di vita della resistenza così costruita? calcolarne la legge, valore atteso e varianza.
(3) Ripetere gli stessi calcoli con tempi di vita aventi legge uniforme sull'intervallo $[0,1]$.

Una soluzione. (1) Il tempo di vita della resistenza costruita collegando in serie è il minimo dei tempi di vita di tutte le resistenze, ossia

$$T_{ser} = \min\left\{T_1, T_2, \dots, T_n\right\},$$

dove T_1, T_2, $\dots T_n$ sono variabili esponenziali di parametro $1/3$. Non avendo altra informazione, assumiamo che le variabili T_1, $\dots T_n$ siano *indipendenti*. Calcoliamo allora la funzione di sopravvivenza

$$\begin{aligned} P(T_{ser} > t|\Omega) &= P(\min\left\{T_1, T_2, \dots, T_n\right\} > t|\Omega) \\ &= P(T_1 > t \text{ e} T_2 > t \text{ e} \dots \text{e} T_n > t|\Omega) = \\ &= P(T_1 > t|\Omega)\cdot P(T_2 > t|\Omega)\cdot \dots \cdot P(T_n > t|\Omega) \\ &= \left(e^{-t/3}\right)^n = e^{-t(n/3)} \end{aligned}$$

avendo usato il fatto che la funzione di sopravvivenza di una esponenziale di parametro λ è $e^{-t\lambda}$ (per $t \geq 0$). La densità si calcola derivando e cambiando di segno: per $t > 0$, abbiamo

$$\varrho(T_{ser} = t|\Omega) = -\frac{d}{dt}e^{-t(n/3)} = \frac{n}{3}e^{-t(n/3)},$$

che è una densità esponenziale di parametro $n \cdot \frac{1}{3}$. (In generale, il minimo di n variabili esponenziali indipendenti di parametro λ è una variabile esponenziale di parametro $n\lambda$).

(2) Se collegate in parallelo, il tempo di vita è il massimo dei tempi di vita delle resistenze, ossia

$$T_{par} = \max\{T_1, T_2, \dots, T_n\},$$

dove T_1, T_2, $\dots T_n$ sono come sopra (in particolare, indipendenti). Calcoliamo la funzione di ripartizione

$$\begin{aligned} P(T_{par} \leq |\Omega) &= P(\max\{T_1, T_2, \dots, T_n\} \leq t|\Omega) \\ &= P(T_1 \leq t \text{ e} T_2 \leq t \text{ e} \dots \text{e} T_n \leq t|\Omega) = \\ &= P(T_1 \leq t|\Omega) \cdot P(T_2 \leq t|\Omega) \cdot \dots \cdot P(T_n \leq t|\Omega) \\ &= \left(1 - e^{-t/3}\right)^n. \end{aligned}$$

avendo usato il fatto che la funzione di ripartizione di una esponenziale di parametro λ è $1 - e^{-t\lambda}$ (per $t \geq 0$). La densità si calcola derivando: per $t > 0$, abbiamo

$$\varrho(T_{par} = t|\Omega) = \frac{d}{dt}\left(1 - e^{-t/3}\right)^n = \frac{n}{3}\left(1 - e^{-t/3}\right)^{n-1} e^{-t/3}.$$

(3) Gli stessi calcoli però con leggi uniformi sull'intervallo $[0, 1]$ ci portano ai risultati, per $t \in [0, 1]$:

$$\begin{aligned} P(T_{ser} > t|\Omega) = (1-t)^n, \quad &\Rightarrow \quad \varrho(T_{ser} = t|\Omega) = n(1-t)^{n-1}, \\ P(T_{par} \leq t|\Omega) = t^n, \quad &\Rightarrow \quad \varrho(T_{par} = t|\Omega) = nt^{n-1}. \end{aligned}$$

Problema 44. Il tempo di attesa per un processo prima che sia svolto da un server è rappresentato da una variabile aleatoria $T \in [0, \infty)$ con legge esponenziale di parametro $\Lambda \in (0, \infty)$. Un progettista è incerto su quale sia un buon valore di Λ, sa più o meno che si trova vicino ad 1, e quindi scrive che Λ è pure una variabile aleatoria esponenziale di parametro 1: quindi nel suo modello dato $\lambda \in \mathbb{R}$, la variabile T è esponenziale di parametro λ, condizionatamente all'evento $\{\Lambda = \lambda\}$.

(1) Scrivere la densità di T rispetto all'informazione iniziale, e calcolarne valore atteso e varianza.
(2) Supponiamo che il progettista osservi che il tempo impiegato per un processo è 1. Come cambia la densità di Λ?
(3) Supponiamo che il progettista osservi che i tempi impiegati per due processi (che suppone indipendenti dato $\Lambda = \lambda$) sono rispettivamente 1 e 2. Come cambia la densità di Λ?

Una soluzione. (1)Scriviamo, per $t \geq 0$,

$$\begin{aligned} \varrho(T = t|\Omega) &= \int_0^\infty \varrho(T = t|\Omega \cap \{\Lambda = \lambda\})\varrho(\Lambda = \lambda|\Omega)\mathrm{d}\lambda \\ &= \int_0^\infty \lambda e^{-\lambda t} e^{-\lambda}\mathrm{d}\lambda \\ &= \int_0^\infty \lambda e^{-\lambda(t+1)}\mathrm{d}\lambda = \frac{1}{(t+1)^2}\int_0^\infty (t+1)\lambda e^{-\lambda(t+1)}(t+1)\mathrm{d}\lambda \\ &= \frac{1}{(t+1)^2}\int_0^\infty y e^{-y}\mathrm{d}y = \frac{1}{(t+1)^2} \end{aligned}$$

Il valore atteso di T rispetto alla informazione iniziale è $+\infty$, ossia l'integrale improprio diverge: dato $M > 0$,

$$\begin{aligned}\int_0^M \frac{t}{(t+1)^2}\mathrm{d}t = \int_0^M \frac{t}{(t+1)^2}\mathrm{d}t \\ = \left(\frac{1}{t+1} + \log(t+1)\right)\Big|_0^M = \frac{1}{M+1} + \log(M+1) - 1 \to +\infty \quad \text{per } M \to +\infty.\end{aligned}$$

Quindi neppure la varianza si può definire.

(2) Usiamo la formula di Bayes (caso continuo/continuo)

$$\begin{aligned}\varrho(\Lambda = \lambda|\Omega \cap \{T = 1\}) &= \varrho(\Lambda = \lambda|\Omega)\frac{\varrho(T = 1|\Omega \cap \{\Lambda = \lambda\})}{\varrho(T = 1|\Omega)} \\ &= e^{-\lambda}\frac{\lambda e^{-1\cdot\lambda}}{1/(1+1)} \\ &= 2\lambda e^{-2\lambda}.\end{aligned}$$

Notiamo che Λ non è più una variabile esponenziale.

(3) Possiamo sfruttare il risultato del punto precedente per aggiornare la densità di Λ usando la formula di Bayes (caso continuo/continuo), rispetto alla informazione $\Omega \cap \{T = 1\}$ e incorporando la nuova informazione $\{T' = 2\}$, dove T' è una variabile esponenziale di parametro λ, indipendente da T, dato $\Omega \cap \{\Lambda = \lambda\}$. Abbiamo

$$\varrho(\Lambda = \lambda|\Omega\cap\{T = 1\}\cap\{T' = 2\}) = \varrho(\Lambda = \lambda|\Omega\cap\{T = 1\})\frac{\varrho(T' = 2|\Omega \cap \{T = 1\} \cap \{\Lambda = \lambda\})}{\varrho(T' = 2|\Omega \cap \{T = 1\})}.$$

Notiamo che, per indipendenza di T da T' dato $\Lambda = \lambda$, vale, per ogni $t \geq 0$ (in particolare per $t = 2$,

$$\varrho(T' = t|\Omega \cap \{T = 1\} \cap \{\Lambda = \lambda\}) = \varrho(T' = 2|\Omega \cap \{\Lambda = \lambda\}) = \lambda e^{-\lambda t}.$$

Il denominatore della formula di Bayes si può calcolare integrando sui i possibili valori di $\Lambda = \lambda$, ossia

$$\begin{aligned}\varrho(T' = 2|\Omega \cap \{T = 1\}) &= \int_0^\infty \varrho(T' = 2|\Omega \cap \{T = 1\} \cap \{\Lambda = \lambda\})\varrho(\Lambda = \lambda|\Omega \cap \{T = 1\})\mathrm{d}\lambda \\ &= \int_0^\infty \varrho(T' = 2|\Omega \cap \{\Lambda = \lambda\})\varrho(\Lambda = \lambda|\Omega \cap \{T = 1\})\mathrm{d}\lambda \\ &= \int_0^\infty \lambda e^{-2\lambda} \cdot 2\lambda e^{-2\lambda}\mathrm{d}\lambda \\ &= 2\int_0^\infty \lambda^2 e^{-4\lambda}\mathrm{d}\lambda \\ &= \frac{2}{4^3}\int_0^\infty (4\lambda)^2 e^{-4\lambda}4\mathrm{d}\lambda \\ &= \frac{2}{4^3}\int_0^\infty y^2 e^{-y}\mathrm{d}y = \frac{1}{4^2}.\end{aligned}$$

In conclusione, abbiamo

$$\varrho(\Lambda = \lambda|\Omega \cap \{T = 1\} \cap \{T' = 2\}) = 2\lambda e^{-2\lambda} \cdot \frac{\lambda e^{-2\lambda}}{1/4^2} = 32\lambda^2 e^{-4\lambda}.$$

Appendice A. Regole di calcolo (v.a. continue)

Sia $E \subseteq \mathbb{R}$ intervallo e I un evento (eventualmente condizionale).
Densità. $X \in E$ è variabile aleatoria con densità $\varrho : E \to [0, \infty)$ (rispetto all'informazione I) se

$$P(X \in A|I) = \int_A \varrho(x)\mathrm{d}x \quad \text{per ogni } A \subseteq E \text{ intervallo.}$$

Notazione: $\varrho(x) = \varrho(X = x|I)$.

(i) Si ha $\int_E \varrho(X = x|I)\mathrm{d}x = 1$ (utile per determinare costanti).
(ii) Si può estendere la densità fuori da E ponendola uguale a zero.
(iii) Densità e condizionamento (B intervallo):

$$\varrho(X = x|I \cap \{X \in B\}) = \begin{cases} \varrho(X = x|I)/P(X \in B|I) & \text{se } x \in B, \\ 0 & \text{altrimenti.} \end{cases}$$

(iv) Densità e trasformazioni affini (se $\lambda \neq 0$)

$$\varrho((\lambda X + c) = y|I) = \varrho\left(X = \frac{y - c}{\lambda}|I\right)\frac{1}{|\lambda|}.$$

Probabilità condizionale. Data $X \in E$ v.a. con densità $\varrho(X = \cdot|I)$, per ogni B evento e $A \subseteq E$ intervallo vale (per definizione)

$$P(B \cap \{X \in A\}|I) = \int_A P(B|\{X = x\} \cap I)\varrho(X = x|I)\mathrm{d}x.$$

Posto $A = E$, si ha (il caso più frequente in cui si applica)

$$P(B|I) = \int_E P(B|\{X = x\} \cap I)\varrho(X = x|I)\mathrm{d}x.$$

Vale $P(B|\{X = x\} \cap I) \in [0, 1]$.

Formula di Bayes: caso continuo/discreto Date $X \in E$, v.a. continua e $Y \in F$ v.a. discreta, si ha

$$\varrho(X = x|\{Y = y\} \cap I) = \frac{P(Y = y|I \cap \{X = x\})\varrho(X = x|I)}{P(Y = y|I)}$$

e pure (scambiando)

$$P(Y = y|I \cap \{X = x\}) = \frac{\varrho(X = x|\{Y = y\} \cap I)P(Y = y|I)}{\varrho(X = x|I)}.$$

Per calcolare i "denominatori":

$$P(Y = y|I) = \int_E P(Y = y|I \cap \{X = x\})\varrho(X = x|I)\mathrm{d}x,$$

$$\varrho(X = x|I) = \sum_{y \in F} \varrho(X = x|I \cap \{Y = y\})P(Y = y|I).$$

Formula di Bayes: caso continuo/continuo Date $X \in E$, $Y \in F$, v.a. continue, si ha

$$\varrho(X = x| \{Y = y\} \cap I) = \frac{\varrho(Y = y|I \cap \{X = x\})\varrho(X = x|I)}{\varrho(Y = y|I)}.$$

Dove

$$\varrho(Y = y|I) = \int_E \varrho(Y = y|I \cap \{X = x\})\varrho(X = x|I)\mathrm{d}x.$$

Indipendenza. Date $X \in E$, $Y \in F \subseteq \mathbb{R}$ v.a. si dicono indipendenti (rispetto ad I) se

$P(X \in A \,\mathrm{e}\, Y \in B|I) = P(X \in A|I)P(Y \in B|I)$ per ogni $A \subseteq E$, $B \subseteq F$ intervalli.

Formulazioni equivalenti

$P(X \in A| \{Y \in B\} \cap I) = P(X \in A|I)$ per ogni $A \subseteq E$, $B \subseteq F$ intervalli,

$P(X \in A| \{Y = y\} \cap I) = P(X \in A|I)$ per ogni $A \subseteq E$ intervallo, $y \in F$,

(se X v.a. continua) $\varrho(X = x| \{Y = y\} \cap I) = \varrho(X = x|I)$ per ogni $x \in E$, $y \in F$.

Funzione ripartizione/sopravvivenza. Per ogni $t \in \mathbb{R}$,

$$P(X \leq t|I) = \int_{(-\infty,t]\cap E} \varrho(X = x|I)\mathrm{d}x, \quad \text{(ripartizione)}$$

$$P(X > t|I) = \int_{(t,+\infty)\cap E} \varrho(X = x|I)\mathrm{d}x, \quad \text{(sopravvivenza)}$$

$$P(X \leq t|I) + P(X > t|I) = 1.$$

Dalla funzione di ripartizione/sopravvivenza alla densità:

$$\frac{d}{dt}P(X \leq t|I) = \varrho(X = t|I) \qquad -\frac{d}{dt}P(X > t|I) = \varrho(X = t|I).$$

Valore atteso. Data $X \in E$ variabile aleatoria con densità $\varrho(X = \cdot|I)$, $f : E \to \mathbb{R}$ si pone

$$\mathbb{E}\left[f(X)|I\right] = \int_E f(x)\varrho(X = x|I)\mathrm{d}x$$

(se l'integrale di Riemann eventualmente improprio è ben definito).

(i) (linearità) $\mathbb{E}\left[cf(X) + g(Y)|I\right] = c\mathbb{E}\left[f(X)|I\right] + \mathbb{E}\left[g(Y)|I\right]$.

(ii) (decomposizione) Se $A_1, \ldots, A_n$ sono un sistema di alternative,

$$\mathbb{E}\left[f(X)|I\right] = \sum_{i=1}^{n} \mathbb{E}\left[f(X)|A_i \cap I\right] P(A_i|I).$$

(iii) (monotonia) Se $f(X) \geq 0$, allora $\mathbb{E}\left[f(X)|I\right] \geq 0$.

(iv) (varianza)

$$\mathrm{Var}\left(f(X)|I\right) = \mathbb{E}\left[(f(X) - \mathbb{E}\left[f(X)|I\right])^2|I\right] = \mathbb{E}\left[f(X)^2|I\right] - (\mathbb{E}\left[f(X)|I\right])^2 \geq 0$$

$$\mathrm{Var}\left(\lambda f(X) + c|I\right) = \lambda^2\, \mathrm{Var}\left(f(X)|I\right).$$

$\mathrm{Var}\left(f(X) + g(Y)|I\right) = \mathrm{Var}\left(f(X)|I\right) + \mathrm{Var}\left(g(Y)|I\right)$ se X, Y v.a. indipendenti.

Appendice B. Esempi di densità

Densità uniforme. Sia $E = (a, b) \subseteq \mathbb{R}$ intervallo (aperto o chiuso non cambia alcunché). Si dice che $X \in E$ è una v.a. uniforme (continua) se

$$\varrho(X = x | X \text{ unif. su } (a,b)) = \frac{1}{b-a} \quad \text{per ogni } x \in (a,b),$$

ossia è *costante* su E. Eventualmente si può porre $\varrho(X = x|X \text{ unif. su } (a,b)) = 0$ se $x \notin (a,b)$.

$$\begin{aligned}
\mathbb{E}\left[X | X \text{ unif. su } (a,b)\right] &= \frac{a+b}{2}, \\
\operatorname{Var}\left(X | X \text{ unif. su } (a,b)\right) &= \frac{(b-a)^2}{12}. \\
P(X \le t | X \text{ unif. su } (a,b)) &= \begin{cases} 0 & \text{se } a < t < b, \\ (t-a)/(b-a) & \text{se } t \in (a,b), \\ 1 & \text{se } t \ge b. \end{cases}
\end{aligned}$$

Densità esponenziale. Sia $\lambda > 0$ un parametro (fissato). Si dice che $X \in [0, \infty)$ è una v.a. esponenziale di parametro λ (brevemente X è $\mathcal{E}(\lambda)$) se

$$\varrho(X = x | X \text{ è } \mathcal{E}(\lambda)) = \lambda e^{-\lambda x} \quad \text{per ogni } x \in [0, \infty),$$

(Eventualmente si può porre $\varrho(X = x|) = 0$ se $x < 0$).

$$\begin{aligned}
\mathbb{E}\left[X | X \text{ è } \mathcal{E}(\lambda)\right] &= \frac{1}{\lambda}, \\
\operatorname{Var}\left(X | X \text{ è } \mathcal{E}(\lambda)\right) &= \frac{1}{\lambda^2}. \\
P(X > t | X \text{ è } \mathcal{E}(\lambda)) &= \begin{cases} 1 & \text{se } t < 0, \\ e^{-\lambda t} & \text{se } t \ge 0. \end{cases}
\end{aligned}$$

Le variabili aleatorie esponenziali indicano per lo più "tempi" (di durata di dispositivi, attese ecc.). Altre proprietà:

(i) (assenza di memoria) Se X è $\mathcal{E}(\lambda)$ rispetto ad una informazione I, allora per ogni $x_0 \ge 0$ (fissato) la variabile $X - x_0$, ossia il "tempo rimanente dopo x_0" è $\mathcal{E}(\lambda)$ rispetto all'informazione $I \cap \{X \ge x_0\}$ (oppure anche $I \cap \{X > x_0\}$)
$$\varrho((X - x_0) = x | I \cap \{X \ge x_0\}) = \varrho(X = x|I).$$

(ii) (minimo di esponenziali indipendenti) Se X_1 è $\mathcal{E}(\lambda_1)$, X_2 è $\mathcal{E}(\lambda_2)$ e le due v.a. sono *indipendenti*, allora $\min\{X_1, X_2\}$ è $\mathcal{E}(\lambda_1 + \lambda_2)$ (si generalizza anche al minimo di n esponenziali tutte indipendenti tra loro).

Densità gaussiana.

Siano $m \in \mathbb{R}$, $\sigma > 0$ due parametri (fissato). Si dice che $X \in \mathbb{R}$ è una v.a. gaussiana di parametri (m, σ^2) (brevemente $\mathcal{N}(m, \sigma^2)$) se

$$\varrho(X = x | X \text{ è } \mathcal{N}(m, \sigma^2)) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2} \cdot \left(\frac{x-m}{\sigma} \right)^2 \right)$$

Notiamo che ϱ è sempre *strettamente* positiva (ma è "praticamente" nulla se $|x - m| > 3\sigma$)

$$\mathbb{E}\left[X | X \text{ è } \mathcal{N}(m, \sigma^2) \right] = m,$$
$$\mathrm{Var}\left(X | X \text{ è } \mathcal{N}(m, \sigma^2) \right) = \sigma^2.$$

Le variabili aleatorie gaussiane sono anche dette "normali" o "centrali" (per la loro rilevanza nel teorema limite centrale).

(i) (trasformazioni affini) Se X è $\mathcal{N}(m, \sigma^2)$ e $a, b \in \mathbb{R}$ sono parametri con $a \neq 0$, la variabile

$$aX + b \quad \text{è } \mathcal{N}(am + b, a^2\sigma^2).$$

(ii) (somma di gaussiane indipendenti) Se X_1 è $\mathcal{N}(m_1, \sigma_1^2)$ e X_2 è $\mathcal{N}(m_2, \sigma_1^2)$ e sono v.a. *indipendenti* allora

$$X_1 + X_2 \quad \text{è } \mathcal{N}(m_1 + m_2, \sigma_1^2 + \sigma_2^2).$$

(questo si generalizza a n variabili gaussiane indipendenti tra loro).

E-mail address, D. Trevisan: `dario.trevisan@unipi.it`