

# Variabili aleatorie continue

## 1 Problemi nel discreto e nel continuo

I problemi nel discreto possono in genere essere affrontati direttamente con il calcolo combinatorio, ma si è visto che, in molti casi risulta più comoda una formulazione in termini di variabili aleatorie. Allora, ci si trova spesso a fare calcoli sulle densità: ad esempio, può essere necessario determinare la densità di una somma  $X_1 + X_2$  di variabili aleatorie indipendenti.

Anche quando esistono metodi per effettuare calcoli di questi tipi direttamente sulle densità, come ad esempio la formula della somma,

$$\begin{aligned} P\{X_1 + X_2 = x\} &= \sum_k P\{X_1 = k, X_2 = x - k\} \\ &= \sum_k P\{X_1 = k\} P\{X_2 = x - k\} \quad (\text{per l'indipendenza}) \end{aligned}$$

più essere utile lavorare invece sulle *funzioni generatrici dei momenti*,

$$m_X(z) = E(e^{zX})$$

oppure sulle *funzioni generatrici delle probabilità*,

$$\psi_X(z) = E(z^X)$$

(le due sono di fatto equivalenti:  $m_X(z) = \psi_X(e^z)$ ), che permettono di identificare più facilmente la distribuzione risultante dai calcoli. Infatti, entrambe queste funzioni identificano univocamente le densità delle variabili corrispondenti, ed esistono formule particolarmente comode per eseguire vari calcoli su tali funzioni.

Ad esempio, la funzione generatrice delle probabilità di una somma  $X_1 + X_2$  di variabili indipendenti è data semplicemente dal prodotto delle funzioni generatrici delle variabili sommate,

$$\psi_{X_1+X_2}(z) = \psi_{X_1}(z) \psi_{X_2}(z)$$

perché:

$$\begin{aligned} \psi_{X_1+X_2}(z) &= E(z^{X_1+X_2}) \\ &= E(z^{X_1} z^{X_2}) \\ &= E(z^{X_1}) E(z^{X_2}) \quad (\text{per l'indipendenza}) \\ &= \psi_{X_1}(z) \psi_{X_2}(z) \end{aligned}$$

Passando al continuo, la descrizione dei problemi diventa spesso più intuitiva. Ad esempio, nel discreto si può considerare la probabilità di fare  $k$  prove prima di ottenere l' $n$ -esimo successo; il problema corrispondente nel continuo è semplicemente il tempo di arrivo dell' $n$ -esimo evento (successo).

Tuttavia, se da un lato si ha questa maggiore intuitività, dall'altro ci possono essere notevoli difficoltà nei calcoli. Può infatti succedere, anche in casi piuttosto elementari (come ad esempio la somma di variabili aleatorie esponenziali), di ottenere formule non risolubili in modo esplicito, dalle quali si possono ottenere solo soluzioni approssimative, attraverso metodi numerici.

I problemi nel continuo tendono anche a complicarsi quando trattano densità ricavate dallo studio dei fenomeni del mondo reale, oppure dalla generalizzazione di una densità discreta a tutti i punti del continuo (con la quale si perde l'interpretazione combinatoria della densità discreta originale).

Gli strumenti utilizzati per identificare le variabili aleatorie continue che si ottengono nei problemi, e calcolarne i parametri, sono:

- la **funzioni caratteristica** per la densità continua  $f(x)$ ,

$$\phi(\theta) = \int_{\mathbb{R}} e^{i\theta x} f(x) dx$$

(che corrisponde alla trasformata di Fourier di  $f$ );

- la **funzione generatrice dei momenti**, definita in modo analogo al caso discreto,

$$m_X(z) = E(e^{zX}) = \int_{\mathbb{R}} e^{zx} f(x) dx$$

(che corrisponde invece alla trasformata di Laplace).

## 1.1 Esempio: somma di variabili geometriche

Un esempio di calcolo (nel discreto) che può essere effettuato con le funzioni generatrici delle probabilità è quello per ricavare la legge della somma  $X+Y$  di due variabili aleatorie geometriche indipendenti di parametro  $p$ .

La funzione generatrice delle probabilità di  $X$  e di  $Y$  (essa è uguale per entrambe, poiché seguono la stessa legge), cioè della distribuzione geometrica, già data in precedenza, è:

$$\psi_X(z) = \psi_Y(z) = \frac{p}{1 - z(1 - p)}$$

Allora, usando la formula per la somma dimostrata prima, la funzione generatrice delle probabilità di  $X + Y$  è:

$$\psi_{X+Y}(z) = \psi_X(z) \psi_Y(z) = \left( \frac{p}{1 - z(1 - p)} \right)^2 = \frac{p^2}{(1 - z(1 - p))^2}$$

Ci si può accorgere che la  $\psi_{X+Y}(z)$  così ottenuta è la funzione generatrice delle probabilità per un'altra distribuzione nota: la binomiale negativa di parametri  $\alpha = 2$  e  $p$ . Infatti, una variabile  $W$  binomiale negativa ha, in generale, la densità

$$p(x) = \begin{cases} \binom{\alpha + x - 1}{x} p^\alpha (1-p)^x & x = 0, 1, \dots \\ 0 & \text{altrimenti} \end{cases}$$

(con  $\alpha$  reale positivo) dalla quale si ricava la funzione generatrice delle probabilità

$$\begin{aligned} \psi_W(z) &= E(z^W) = \sum_{x=0}^{\infty} z^x p(x) \\ &= \sum_{x=0}^{\infty} z^x \binom{\alpha + x - 1}{x} p^\alpha (1-p)^x \\ &= p^\alpha \sum_{x=0}^{\infty} \binom{\alpha + x - 1}{x} (z(1-p))^x && \text{serie } \frac{1}{(1-t)^\alpha} = \sum_{x=0}^{\infty} \binom{\alpha + x - 1}{x} t^x, \\ &= p^\alpha \frac{1}{(1 - z(1-p))^\alpha} && \text{con } t = z(1-p) \\ &= \frac{p^\alpha}{(1 - z(1-p))^\alpha} \end{aligned}$$

che, per  $\alpha = 2$ , coincide appunto con  $\psi_{X+Y}(z)$ .

Più in generale, una somma di  $n$  variabili aleatorie geometriche di parametro  $p$ ,  $X_1 + \dots + X_n$ , ha la funzione generatrice delle probabilità

$$\psi_{X_1 + \dots + X_n}(z) = \frac{p^n}{(1 - z(1-p))^n}$$

ovvero segue una distribuzione binomiale negativa di parametri  $n$  e  $p$ .

Ciò corrisponde anche al significato intuitivo dato a una variabile binomiale negativa con  $\alpha = n \in \mathbb{N}$ : essa conta il numero di insuccessi prima di ottenere  $n$  successi, il quale non è altro che la somma dei numeri di insuccessi ottenuti prima di ciascun successo. Questo perché, una volta verificatosi il primo successo, dopo un numero di insuccessi contato da  $X_1$ , sapendo che le prove sono indipendenti si può “ripartire da capo” a contare gli insuccessi fino al secondo successo, usando una “nuova” variabile aleatoria geometrica  $X_2$ , e così via, fino all' $n$ -esimo successo.

## 2 Distribuzione gamma

*Definizione:* Si chiama **funzione gamma** la funzione  $\Gamma : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  definita da

$$\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1} e^{-x} dx$$

Questa funzione generalizza a  $\mathbb{R}^+$  il concetto di fattoriale, perché si può dimostrare (mediante integrazione per parti) che

$$\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$$

(e si ha che  $\Gamma(1) = 1$ ), e di conseguenza, se  $n$  è un intero positivo:

$$\Gamma(n) = (n - 1)!$$

*Definizione:* Si dice che una variabile aleatoria  $X$  segue una **legge gamma** di parametri  $\alpha > 0$  e  $\lambda > 0$ , e si scrive  $X \sim \Gamma(\alpha, \lambda)$ , se e solo se la densità di  $X$  è:

$$f(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} & \text{se } x > 0 \\ 0 & \text{altrimenti} \end{cases}$$

(come al solito, con un'integrazione, si può verificare che  $f$  è effettivamente una densità).

## 2.1 Somma di variabili aleatorie gamma

Si vuole calcolare la densità di una somma  $X_1 + \dots + X_m$  di variabili aleatorie  $X_i \sim \Gamma(\alpha_i, \lambda)$  indipendenti (di parametri  $\alpha$  diversi, ma  $\lambda$  uguale).

Innanzitutto, si osserva che, per sommare  $m$  variabili aleatorie, è sufficiente saperne sommare due alla volta, ripetendo quest'operazione fino ad averle sommate tutte. Quindi, il problema si può ridurre al caso  $m = 2$ , cioè  $X_1 + X_2$ , con  $X_1 \sim \Gamma(\alpha_1, \lambda)$  e  $X_2 \sim \Gamma(\alpha_2, \lambda)$ .

Se  $f_1(x)$  e  $f_2(x)$  sono, rispettivamente, le densità di  $X_1$  e  $X_2$ , la densità  $g(y)$  di  $X_1 + X_2$  può essere calcolata usando una formula analoga a quella del caso discreto:

$$g(y) = \int_{-\infty}^{+\infty} f_1(x) f_2(y - x) dx$$

Sostituendo, in questo integrale, le formule di  $f_1$  e  $f_2$ , e svolgendo i calcoli, si trova che  $g(y)$  è la densità di una variabile  $\Gamma(\alpha_1 + \alpha_2, \lambda)$ .

Allora, in generale, la somma  $X_1 + \dots + X_m$  segue la distribuzione  $\Gamma(\alpha_1 + \dots + \alpha_m, \lambda)$ .

## 2.2 Funzione di ripartizione

In generale, non esistono espressioni semplici per la funzione di ripartizione di  $\Gamma(\alpha, \lambda)$ . In particolare, tale espressione esiste solo nei casi  $\Gamma(m, \lambda)$ , dove  $m$  è un intero maggiore di 0:

- Per  $m = 1$ , la distribuzione  $\Gamma(1, \lambda)$  si riduce all'esponenziale di parametro  $\lambda$ , poiché la densità diventa

$$f(x) = \begin{cases} \frac{\lambda^1}{\Gamma(1)} x^{1-1} e^{-\lambda x} & \text{se } x > 0 \\ 0 & \text{altrimenti} \end{cases} = \begin{cases} \lambda e^{-\lambda x} & \text{se } x > 0 \\ 0 & \text{altrimenti} \end{cases}$$

e allora anche la funzione di ripartizione è quella dell'esponenziale:

$$F_1(x) = 1 - e^{-\lambda x}$$

- Per  $m > 1$ , si integra per parti,

$$\begin{aligned} F_m(x) &= \int_0^x \frac{\lambda^m}{\Gamma(m)} t^{m-1} e^{-\lambda t} dt \quad g'(t) = e^{-\lambda t} \implies g(t) = \frac{e^{-\lambda t}}{-\lambda} \\ &= \frac{\lambda^m}{(m-1)!} \int_0^x t^{m-1} e^{-\lambda t} dt = t^{m-1} \implies h'(t) = (m-1)t^{m-2} \\ &= \frac{\lambda^m}{(m-1)!} \left( \left[ \frac{e^{-\lambda t}}{-\lambda} t^{m-1} \right]_0^x - \int_0^x \frac{e^{-\lambda t}}{-\lambda} (m-1)t^{m-2} dt \right) \\ &= -\frac{\lambda^{m-1}}{(m-1)!} [t^{m-1} e^{-\lambda t}]_0^x + \underbrace{\frac{\lambda^{m-1}}{(m-2)!} \int_0^x t^{m-2} e^{-\lambda t} dt}_{=F_{m-1}(x)} \\ &= -\frac{\lambda^{m-1}}{(m-1)!} (x^{m-1} e^{-\lambda x} - 0) + F_{m-1}(x) \\ &= -\frac{\lambda^{m-1}}{(m-1)!} x^{m-1} e^{-\lambda x} + F_{m-1}(x) \\ &= -\frac{(\lambda x)^{m-1}}{(m-1)!} e^{-\lambda x} + F_{m-1}(x) \end{aligned}$$

e così, procedendo per induzione, si ottiene:

$$\begin{aligned} F_m(x) &= -\frac{(\lambda x)^{m-1}}{(m-1)!} e^{-\lambda x} + F_{m-1}(x) \\ &= -\frac{(\lambda x)^{m-1}}{(m-1)!} e^{-\lambda x} - \frac{(\lambda x)^{m-2}}{(m-2)!} e^{-\lambda x} + F_{m-2}(x) \\ &= -\frac{(\lambda x)^{m-1}}{(m-1)!} e^{-\lambda x} - \frac{(\lambda x)^{m-2}}{(m-2)!} e^{-\lambda x} - \dots + F_2(x) \\ &= -\frac{(\lambda x)^{m-1}}{(m-1)!} e^{-\lambda x} - \frac{(\lambda x)^{m-2}}{(m-2)!} e^{-\lambda x} - \dots - \frac{(\lambda x)^1}{1!} e^{-\lambda x} + F_1(x) \end{aligned}$$

Come appena visto,  $F_1(x)$  è la funzione di ripartizione della distribuzione esponenziale, quindi:

$$\begin{aligned}
 &= -\frac{(\lambda x)^{m-1}}{(m-1)!}e^{-\lambda x} - \frac{(\lambda x)^{m-2}}{(m-2)!}e^{-\lambda x} - \dots - \frac{(\lambda x)^1}{1!}e^{-\lambda x} + 1 - e^{-\lambda x} \\
 &= -\frac{(\lambda x)^{m-1}}{(m-1)!}e^{-\lambda x} - \frac{(\lambda x)^{m-2}}{(m-2)!}e^{-\lambda x} - \dots - \frac{(\lambda x)^1}{1!}e^{-\lambda x} - \underbrace{\frac{(\lambda x)^0}{0!}}_{=1}e^{-\lambda x} + 1 \\
 &= 1 - \sum_{k=0}^{m-1} \frac{(\lambda x)^k}{k!}e^{-\lambda x}
 \end{aligned}$$

### 3 Problema: legame tra esponenziale e Poisson

*Problema:* Si suppone che gli intervalli tra telefonate successive che giungono a un centralino siano variabili aleatorie indipendenti ed esponenziali di parametro  $\lambda$ . Qual è la probabilità che nell'intervallo di tempo  $[0, T]$ , con  $T > 0$ , giungano esattamente  $k$  telefonate?

Sia  $X_1, X_2, \dots$  una successione di variabili aleatorie esponenziali indipendenti, nella quale ogni  $X_i$  valuta il tempo trascorso tra la  $(i-1)$ -esima e l' $i$ -esima telefonata. Allora, il tempo necessario per ricevere  $n$  chiamate è dato dalla somma  $X_1 + \dots + X_n$ . Di conseguenza, il numero  $Y$  di chiamate ricevute nell'intervallo  $[0, T]$  corrisponde al più grande numero  $n$  di telefonate tale che il tempo necessario a riceverle tutte sia ancora  $\leq T$ :

$$Y = \sup\{n \mid X_1 + \dots + X_n \leq T\}$$

(per la precisione, qui si considera l'estremo superiore,  $\sup$ , e non il massimo, perché non si è al momento sicuri che il massimo numero di telefonate sia finito).

Il problema chiede di calcolare la probabilità  $P\{Y = k\}$ , cioè la densità della variabile aleatoria *discreta*  $Y$ , ma tale calcolo non è immediato. Piuttosto, conviene ricavare prima la funzione di ripartizione di  $Y$ , ovvero la probabilità  $P\{Y \leq k\}$  che giungano al centralino non più di  $k$  chiamate nell'intervallo di tempo  $[0, T]$  fissato. Infatti, si osserva che il numero di chiamate nell'intervallo è  $\leq k$  se e solo se la  $(k+1)$ -esima chiamata "non fa in tempo" ad arrivare entro lo scadere dell'intervallo:

$$\{Y \leq k\} = \{X_1 + \dots + X_{k+1} > T\}$$

Passando all'evento complementare, si vede che la funzione di ripartizione di  $Y$  può essere calcolata in termini di quella di  $X_1 + \dots + X_{k+1}$ .

$$\begin{aligned}
 P\{Y \leq k\} &= P\{X_1 + \dots + X_{k+1} > T\} \\
 &= 1 - P\{X_1 + \dots + X_{k+1} \leq T\}
 \end{aligned}$$

Ma le variabili  $X_i$  sono esponenziali di parametro  $\lambda$ , o, in altre parole, gamma di parametri  $\alpha = 1$  e  $\lambda$  ( $X_i \sim \Gamma(1, \lambda)$ ), e sono indipendenti, quindi, come visto in precedenza, la loro somma segue la legge

$$X_1 + \dots + X_{k+1} \sim \Gamma(\underbrace{1 + \dots + 1}_{k+1}, \lambda) = \Gamma(k + 1, \lambda)$$

che ha la funzione di ripartizione

$$F_{k+1}(x) = 1 - \sum_{i=0}^{(k+1)-1} \frac{(\lambda x)^i}{i!} e^{-\lambda x} = 1 - \sum_{i=0}^k \frac{(\lambda x)^i}{i!} e^{-\lambda x}$$

Allora, la funzione di ripartizione di  $Y$  è:

$$\begin{aligned} P\{Y \leq k\} &= 1 - P\{X_1 + \dots + X_{k+1} \leq T\} \\ &= 1 - F_{k+1}(T) \\ &= 1 - 1 + \sum_{i=0}^k \frac{(\lambda T)^i}{i!} e^{-\lambda T} \\ &= \sum_{i=0}^k \frac{(\lambda T)^i}{i!} e^{-\lambda T} \end{aligned}$$

Si riconosce che questa è la funzione di ripartizione della distribuzione di Poisson di parametro  $\lambda T$ . Dunque,  $Y$  è appunto una variabile aleatoria di Poisson di parametro  $\lambda T$ , e la probabilità che giungano al centralino esattamente  $k$  telefonate nell'intervallo di tempo  $[0, T]$  è data dalla densità della Poisson:

$$P\{Y = k\} = \frac{(\lambda T)^k}{k!} e^{-\lambda T}$$

Questo risultato vale in generale: se il tempo che trascorre tra ciascun evento e il successivo è dato da variabili aleatorie esponenziali indipendenti, tutte di parametro  $\lambda$  uguale, una variabile aleatoria (discreta) che conta il numero di eventi in un intervallo di tempo di ampiezza  $T$  segue la legge di Poisson di parametro  $\lambda T$ .

## 4 Problema: attesa a due sportelli

*Problema:* Un servizio viene erogato presso due sportelli: una persona che vuole usufruirne deve aspettare che si liberi *almeno uno* dei due sportelli. Se i tempi di attesa relativi ai singoli sportelli sono dati dalle variabili aleatorie indipendenti  $T_1$  e  $T_2$ , esponenziali di parametri  $\lambda_1$  e  $\lambda_2$  (rispettivamente), qual è la densità della variabile aleatoria  $T$  che indica il tempo di attesa per accedere al servizio?

Siccome è sufficiente che si liberi uno dei due sportelli, il tempo di attesa  $T$  è dato dal *minimo* tra i tempi  $T_1$  e  $T_2$  dei singoli sportelli:

$$T = \min(T_1, T_2)$$

È importante osservare che, in questo caso, *non* si tratta di una somma di variabili esponenziali, a differenza del problema precedente (si avrebbe anche qui una somma se si volesse invece accedere a entrambi gli sportelli, uno dopo l'altro). Perciò, non è corretto utilizzare gli stessi metodi di calcolo applicati prima.

Piuttosto, si osserva che l'evento  $\{T > t\}$  (il complementare di quello che definisce la funzione di ripartizione) è verificato se e solo se i tempi di attesa di entrambi gli sportelli sono maggiori di  $t$ :

$$\{T > t\} = \{\min(T_1, T_2) > t\} = \{T_1 > t, T_2 > t\}$$

se uno dei due fosse  $< t$ , allora anche  $T$ , che è il minimo di questi due tempi, sarebbe  $< t$ . Quindi, sapendo che  $T_1$  e  $T_2$  sono indipendenti ed esponenziali, la probabilità di  $\{T > t\}$  è data da

$$\begin{aligned} P\{T > t\} &= P\{\min(T_1, T_2) > t\} \\ &= P\{T_1 > t, T_2 > t\} \\ &= P\{T_1 > t\} P\{T_2 > t\} && \text{(indipendenza)} \\ &= (1 - P\{T_1 \leq t\})(1 - P\{T_2 \leq t\}) && \text{(eventi complementari)} \\ &= (1 - 1 + e^{-\lambda_1 t})(1 - 1 + e^{-\lambda_2 t}) && \text{(funzioni di ripartizione esponenziali)} \\ &= e^{-\lambda_1 t} e^{-\lambda_2 t} \\ &= e^{-(\lambda_1 + \lambda_2)t} \end{aligned}$$

e, di conseguenza, la funzione di ripartizione di  $T$  è

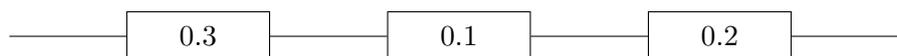
$$P\{T \leq t\} = 1 - P\{T > t\} = 1 - e^{-(\lambda_1 + \lambda_2)t}$$

ovvero  $T$  è anch'essa esponenziale, di parametro  $\lambda_1 + \lambda_2$ , e dunque la sua densità è:

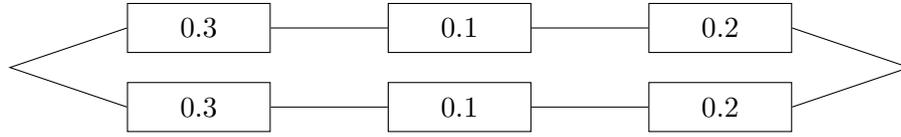
$$f(t) = (\lambda_1 + \lambda_2)e^{-(\lambda_1 + \lambda_2)t}$$

## 5 Problema: tempo di vita in serie e in parallelo

*Problema:* Un componente elettronico è formato da tre elementi in serie, ciascuno dei quali ha un tempo di vita esponenziale, rispettivamente di parametri  $\lambda = 0.3$ ,  $\mu = 0.1$  e  $\gamma = 0.2$ .



1. Sia  $T$  la variabile aleatoria che misura il tempo di vita dell'intero componente. Qual è la legge di  $T$ ?
2. Per aumentare l'affidabilità, si aggiunge un componente identico in parallelo.



Qual è la legge del tempo di vita del componente così formato?

*Soluzioni:*

1. Siano  $T_1, T_2, T_3$  le variabili aleatorie esponenziali (si suppone indipendenti) che rappresentano i tempi di vita dei tre elementi. Siccome tali elementi sono in serie, è sufficiente che si guasti uno di essi perché l'intero componente smetta di funzionare. Ciò significa che il tempo di vita del componente è dato dal minimo dei tempi di vita dei tre elementi:

$$T = \min(T_1, T_2, T_3)$$

Procedendo esattamente come nel problema precedente,

$$\begin{aligned} P\{T > t\} &= P\{\min(T_1, T_2, T_3) > t\} \\ &= P\{T_1 > t, T_2 > t, T_3 > t\} \\ &= P\{T_1 > t\} P\{T_2 > t\} P\{T_3 > t\} \\ &= e^{-\lambda t} e^{-\mu t} e^{-\gamma t} \\ &= e^{-(\lambda+\mu+\gamma)t} \end{aligned}$$

si ricava che  $T$  è una variabile aleatoria esponenziale di parametro  $\lambda + \mu + \gamma$ , la cui funzione di ripartizione è

$$F_T(t) = P\{T < t\} = 1 - e^{-(\lambda+\mu+\gamma)t}$$

(per  $t > 0$ , e uguale a 0 altrimenti).

2. Siano  $T$  e  $W$  i tempo di vita di ciascuno dei due rami paralleli. Dato che i due rami sono identici, entrambe queste variabili aleatorie sono esponenziali di parametro  $\lambda + \mu + \gamma$  (come determinato nella parte precedente del problema). Inoltre, poiché i tempi di vita dei singoli elementi che formano i rami sono indipendenti, lo sono anche  $T$  e  $W$ . Questa volta, però, il componente continua a funzionare fintanto che funziona almeno uno dei due rami paralleli, quindi il tempo di vita complessivo  $X$  è dato dal massimo di  $T$  e  $W$ :

$$X = \max(T, W)$$

Osservando che  $\max(T, W) \leq t$  se e solo se  $T \leq t$ ,  $W \leq t$ , si ricava la funzione di ripartizione di  $X$ :

$$\begin{aligned}
 F_X(t) &= P\{X \leq t\} \\
 &= P\{\max(T, W) \leq t\} \\
 &= P\{T \leq t, W \leq t\} \\
 &= P\{T \leq t\} P\{W \leq t\} \\
 &= F_T(t) F_W(t) \\
 &= (F_T(t))^2 \quad (\text{perché } T \text{ e } W \text{ seguono la stessa legge}) \\
 &= (1 - e^{-(\lambda+\mu+\gamma)t})^2
 \end{aligned}$$

(per  $t > 0$ , e uguale a 0 altrimenti).

## 6 Distribuzione uniforme

*Definizione:* Dati due numeri reali  $a$  e  $b$ , con  $a < b$ , si dice che una variabile aleatoria  $X$  ha **distribuzione uniforme**, con parametri  $a$  e  $b$ , se la sua densità di probabilità è:

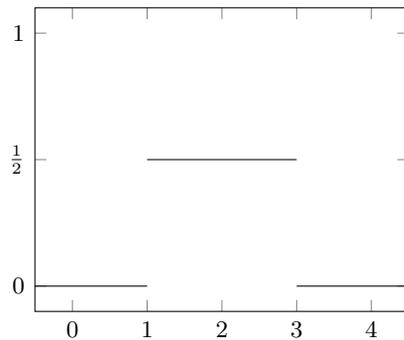
$$f(x) = \begin{cases} \frac{1}{b-a} & \text{se } a \leq x \leq b \\ 0 & \text{altrimenti} \end{cases}$$

La funzione di ripartizione corrispondente è:

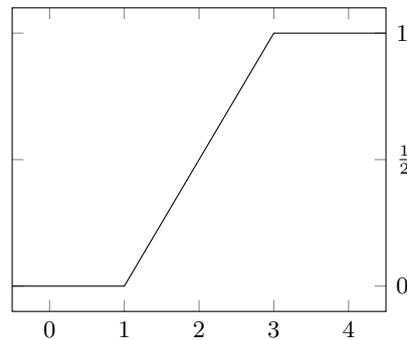
$$F(x) = \begin{cases} 0 & \text{se } x \leq a \\ \frac{x-a}{b-a} & \text{se } a < x < b \\ 1 & \text{se } x \geq b \end{cases}$$

Distribuzione uniforme con  $a = 1$  e  $b = 3$

Densità  $f(x)$



Funzione di ripartizione  $F(x)$



Per ricavare questa definizione, è sufficiente chiedere che la densità assuma un valore costante  $c > 0$  nell'intervallo  $[a, b]$ , e 0 altrimenti.

$$f(x) = \begin{cases} c & \text{se } a \leq x \leq b \\ 0 & \text{altrimenti} \end{cases}$$

Infatti, perché questa funzione sia una densità, il suo integrale su tutto  $\mathbb{R}$

$$\int_{-\infty}^{+\infty} f(x) dx = \int_a^b c dx = c \int_a^b dx = c[x]_a^b = c(b-a)$$

deve valere 1, e questo requisito fissa il valore di  $c$ :

$$c(b-a) = 1 \iff c = \frac{1}{b-a}$$

Il *valore medio* della distribuzione uniforme è

$$\begin{aligned} \mu = E(X) &= \int_{-\infty}^{+\infty} x f(x) dx = \int_a^b \frac{x}{b-a} dx = \frac{1}{b-a} \int_a^b x dx \\ &= \frac{1}{b-a} \left[ \frac{x^2}{2} \right]_a^b = \frac{1}{b-a} \cdot \frac{b^2 - a^2}{2} = \frac{(b-a)(b+a)}{2(b-a)} \\ &= \frac{a+b}{2} \end{aligned}$$

mentre la *varianza* è data da

$$\begin{aligned} \sigma^2 = \text{Var}(X) &= E(X^2) - (E(X))^2 \\ &= \int_{-\infty}^{+\infty} x^2 f(x) dx - \left( \frac{a+b}{2} \right)^2 \\ &= \frac{1}{b-a} \int_a^b x^2 dx - \frac{(a+b)^2}{4} \\ &= \frac{1}{b-a} \left[ \frac{x^3}{3} \right]_a^b - \frac{(a+b)^2}{4} \\ &= \frac{b^3 - a^3}{3(b-a)} - \frac{(a+b)^2}{4} \\ &= \frac{(b-a)(a^2 + ab + b^2)}{3(b-a)} - \frac{a^2 + 2ab + b^2}{4} \\ &= \frac{4a^2 + 4ab + 4b^2 - 3a^2 - 6ab - 3b^2}{12} \\ &= \frac{a^2 - 2ab + b^2}{12} \\ &= \frac{(b-a)^2}{12} \end{aligned}$$