

L'esponenziale e la Poisson

Sia T una variabile casuale esponenziale con parametro $\lambda = 0.25$ (quindi con valore atteso $1/\lambda = 4$). È importante calare questi valori in un contesto “reale”, non dimenticando che λ (e, di conseguenza, il valore atteso $1/\lambda$) ha un'unità di misura. In particolare, una variabile casuale esponenziale modella i “tempi di accadimento” (*failure times*) relativi ad un evento e la media ($1/\lambda$) ha quindi la stessa unità di misura di T (cioè di un *tempo*). Nel nostro esempio, 4 (cioè $1/0.25$) potrebbe essere il numero medio di secondi (o di miliardi di anni, dipende dalla sostanza) che passano prima di osservare una disintegrazione radioattiva, oppure il numero medio di minuti che si passano in coda prima di essere serviti, oppure il numero medio di anni che passano prima che un *hard disk* manifesti un difetto, e così via.

Se la media ($1/\lambda$) è un tempo, allora l'unità di misura del parametro λ di una esponenziale è il *reciproco* di un tempo ed è quindi una *velocità* (*rate*, in inglese), ovvero un *tasso*, come molto spesso viene chiamato in Italiano. Infatti, in lingua Inglese λ viene spesso indicata come *failure rate* o, anche, *hazard rate*. Nel nostro esempio potrebbe indicare che si osservano (in media) 0.25 disintegrazioni radioattive al secondo (o in un miliardo di anni) oppure che in una coda vengono servite 0.25 soggetti al minuto, oppure che gli *hard disk* manifestano un difetto in ragione di 0.25 all'anno.

Impiegando la funzione `rexp`, estraiamo un campione casuale di 10 tempi (t_1, t_2, \dots, t_{10}) da una distribuzione esponenziale con $\lambda = 0.25$ (al minuto, per esempio)

```
> set.seed(123456)
> N <- 10
> n <- c(1:N)
> x <- rexp(N,0.25)
```

e, mediante la funzione `cumsum`, associamo a ciascun tempo il corrispondente “tempo cumulato” ($T_1 = t_1$, $T_2 = t_1 + t_2$, $T_3 = t_1 + t_2 + t_3, \dots, T_{10} = t_1 + t_2 + \dots + t_{10}$). I risultati della simulazione sono i seguenti:

```
> X <- cumsum(x)
> rbind("t"=round(x,1), "T"=round(X,1))
  [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
t  2.4  2.0  5.0  4.2  4.6  7.9  0.3 10.5  0.5   2.4
T  2.4  4.4  9.4 13.7 18.2 26.1 26.4 36.9 37.4 39.8
```

Quindi, immaginando che il tempo sia espresso in minuti, abbiamo dovuto aspettare 2.4 minuti per osservare il primo evento; tra il primo evento e il secondo sono trascorsi 2 minuti, tra il secondo e il terzo 5 minuti e così via. *In media*, tra un evento e il successivo sono trascorsi circa 4 minuti:

```
> mean(x)
[1] 3.980762
```

(come ci potevamo aspettare, anche se il numero di tempi considerati è piuttosto basso).

Il tempo è naturalmente misurato in modo *continuo* (è una *variabile casuale continua*); esiste però un altro modo di vedere gli stessi risultati, facendo ricorso ad un modello *discreto* (cioè ad una *variabile casuale discreta*) che prende in esame i valori cumulati dei tempi (T_i).

Per osservare il primo evento, abbiamo dovuto aspettare 2.4 minuti ($T_1 = t_1$), per osservare un totale di 2 eventi, abbiamo dovuto aspettare 4.4 minuti ($T_2 = t_1 + t_2$), per osservarne 3, abbiamo dovuto aspettare 9.4 minuti, per osservarne 4, abbiamo dovuto aspettare 13.7 minuti, e così via; l'ultimo evento (il decimo) si è verificato dopo 39.8 minuti dall'inizio del periodo di osservazione.

Esiste un modo piuttosto efficace di rappresentare graficamente questo risultato (che è un esempio di *processo di Poisson*). Lo abbiamo fatto nella figura 1.

In ascissa sono riportati i “tempi di accadimento” cumulati (T_1, T_2, \dots, T_{10}), mentre l'ordinata corrispondente al tempo T_i è il numero totale $N(T_i)$ di eventi osservati fino al tempo T_i incluso. La lunghezza di ciascuna linea che connette un pallino pieno ad un pallino vuoto è uguale al tempo t intercorso fra i due eventi. La funzione rappresentata nella figura 1 è una *funzione a gradino* non decrescente, dove l'altezza di ogni “gradino” è uguale a 1.

Anche se i tempi t osservati nella simulazione sono (come c'era da aspettarsi) piuttosto diversi fra loro, andando da circa 17 secondi (0.28 minuti) fino a oltre 10 minuti, la media dei 10 tempi è, come abbiamo già visto, molto vicina al valore vero.

Possiamo quindi dire che *in media* occorre aspettare circa 4 minuti tra un evento e il successivo. Un modo alternativo per dire la stessa cosa è fare riferimento al parametro λ (che nel nostro caso è uguale a 0.25), la cui *dimensione* è una *velocità*. Nella nostra simulazione, quindi, $\lambda = 0.25/\text{minuto}$. Quindi ci aspettiamo di osservare (in media) 0.25 eventi in un minuto di osservazione.

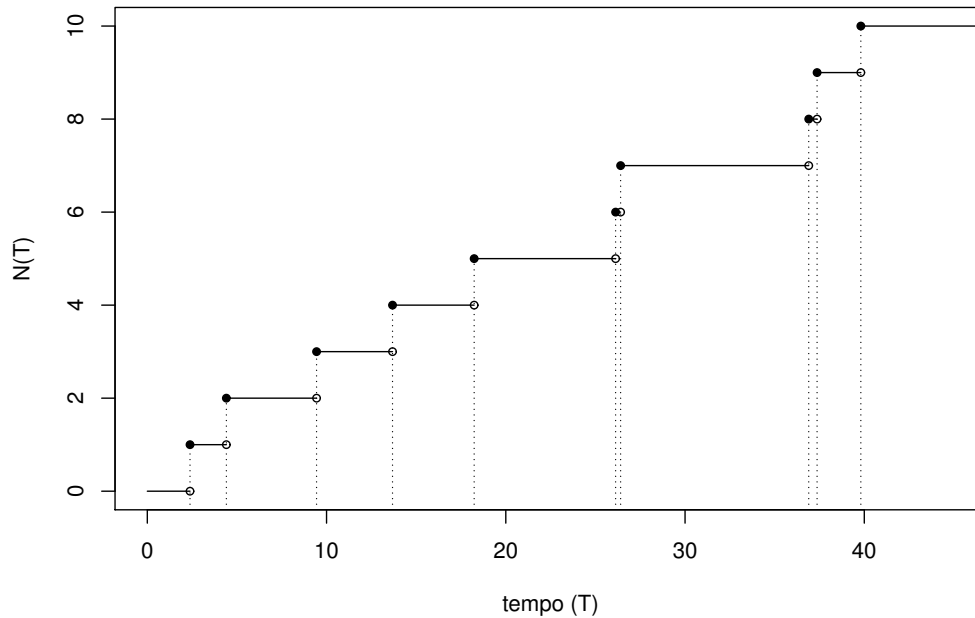


Figura 1: Esempio di un processo di Poisson.

Se la durata dell'osservazione è di circa 40 minuti (come nella nostra simulazione), ci aspettiamo di osservare complessivamente $0.25 \times 40 = 10$ eventi, che è proprio il numero di eventi che abbiamo osservato.

Possiamo allora spostare l'attenzione dal tempo intercorrente fra gli eventi al *numero di eventi* osservati in un intervallo di tempo definito (ad esempio, in un minuto di osservazione). Suddividiamo quindi i 40 minuti di osservazione riportati nella figura 1 in 40 intervalli, ciascuno della durata di un minuto (figura 2). In 31 di questi 40 intervalli non è stato osservato alcun evento; in 8 intervalli è stato osservato 1 evento e in 1 intervallo sono stati osservati 2 eventi. Il numero medio di eventi osservati in ciascun intervallo è quindi $10/40 = 0.25$. Possiamo pensare a questo valore come al valore atteso di una variabile casuale X binomiale in cui $n = 40$ e, quindi, $p = 0.25/40$ (cioè 0.00625). Le probabilità previste per $X = 0, 1, 2$ sono

```
> round(dbinom(c(0:2), 40, 0.25/40), 5)
[1] 0.77819 0.19577 0.02401
```

e la loro somma è molto vicina ad 1 (0.9980), per cui possiamo trascurare i valori da 3 a 40).

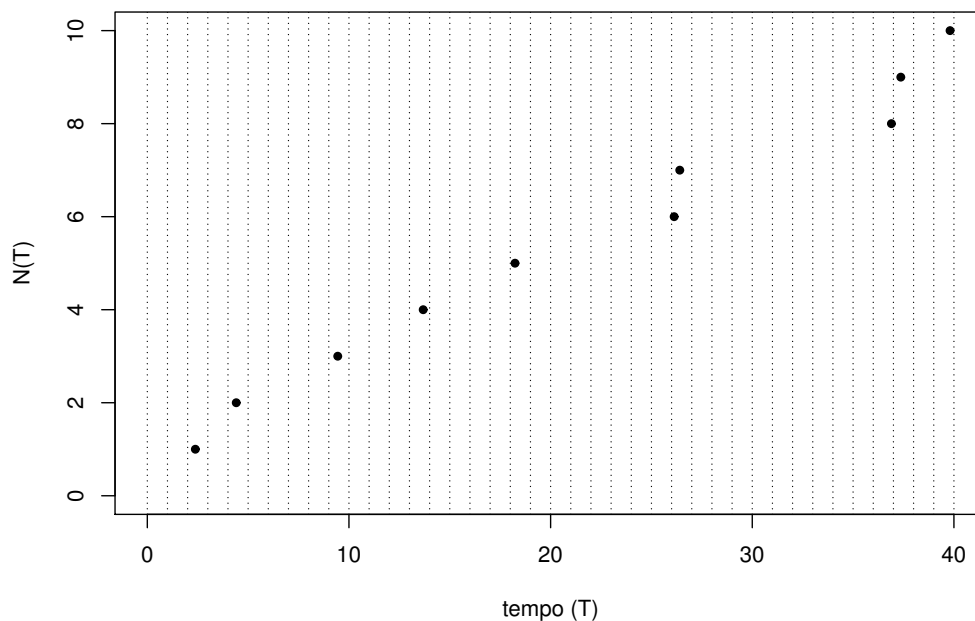


Figura 2: Esempio di un processo di Poisson.

Se moltiplichiamo le tre probabilità appena trovate per 40 (il numero di intervalli), otteniamo il numero di intervalli in cui ci aspettiamo di osservare 0, 1, 2 eventi:

```
> round(dbinom(c(0:2),40,0.25/40)*40,1)
[1] 31.1  7.8  1.0
```

che sono valori straordinariamente vicini a quelli osservati (31, 8, 1).

In effetti, potremmo pensare di dividere il periodo di osservazione (40 minuti) in un numero n molto più grande di intervalli (ognuno di durata estremamente breve, ad esempio 1/10 di secondo o anche meno) in modo tale che in ciascun intervallo può venire registrato al massimo 1 evento. Portando questo ragionamento *al limite* ($n \rightarrow \infty$), mantenendo comunque finita

(e non nulla) la probabilità p di successo in ciascuna prova (cioè la probabilità di osservare un evento in un intervallo), la distribuzione binomiale si “trasforma” nella distribuzione di Poisson¹ il cui parametro (indichiamolo qui con la lettera greca μ , dal momento che è la media) è il valore atteso np della binomiale “di partenza”.

Nella nostra simulazione, nella quale abbiamo visto che $np = 0.25$, possiamo mettere a confronto le probabilità di una binomiale con $n = 40$ e $p = 0.00625$ con quelle di una Poisson dove $\mu = np = 0.25$:

```
> a <- round(dbinom(c(0:2),40,0.25/40),5)
> b <- round(dpois(c(0:2),0.25),5)
> rbind(a,b)
      [,1]      [,2]      [,3]
a 0.77819 0.19577 0.02401
b 0.77880 0.19470 0.02434
```

osservando che i valori sono molto vicini fra loro.

Abbiamo quindi “dimostrato”² un legame estremamente interessante ed importante fra la distribuzione esponenziale e la distribuzione di Poisson: se il tempo T necessario per osservare un evento è distribuito in modo esponenziale (con parametro λ), allora il numero X di eventi osservati in un intervallo di tempo t fissato è distribuito come una Poisson (con parametro $\mu = \lambda t$).

Un altro modo per arrivare allo stesso risultato sfrutta il ragionamento seguente.³ In una distribuzione di Poisson la cui media è $\mu = \lambda t$,⁴ la probabilità di osservare 0 eventi in un intervallo di tempo t è uguale a

$$P(X = 0) = e^{-\lambda t}.$$

Questo è anche il valore della *funzione di sopravvivenza*⁵ (*survival function*) $S(t)$ di una variabile casuale T esponenziale di parametro λ :

$$P(T > t) = S(t) = e^{-\lambda t}.$$

¹Si veda, ad esempio, E. Parzen (1992). La moderna teoria delle probabilità e le sue applicazioni. Franco Angeli, Milano (pag. 139).

²Naturalmente quella che abbiamo visto non è per nulla una dimostrazione; al più è un abbozzo della linea da seguire per arrivare ad una dimostrazione sufficientemente rigorosa.

³Cooper J.C.B. (2005). The Poisson and Exponential Distributions. *Mathematical Spectrum*, 37(3), 123–125.

⁴Cioè, in una distribuzione di Poisson in cui si osservano in media λ eventi in un intervallo di tempo $t = 1$.

⁵Ricordiamo che la funzione di sopravvivenza è il complemento a 1 della funzione di ripartizione: $S(t) = 1 - F(t)$.

Abbiamo pertanto che

$$P(X = 0 | \mu = \lambda t) = P(T > t),$$

cioè la probabilità di non osservare alcun evento durante un intervallo di tempo t (per una variabile casuale di Poisson) coincide con la probabilità che il “tempo di sopravvivenza” (di una variabile casuale esponenziale) sia maggiore di t . Come già detto, si tratta di considerazioni sui legami fra queste due distribuzioni. Per una dimostrazione matematica, rimandiamo a testi specialistici.⁶

Possiamo comunque, ripetendo una simulazione già proposta nel nostro libro, far vedere come, partendo da una variabile casuale esponenziale, la distribuzione del numero di eventi osservati in un intervallo di tempo fissato, sia (molto vicina a quella di) una Poisson.

Prima di presentare questa simulazione, riportiamo qui di seguito il risultato ottenuto nella nostra prima (mini) simulazione.

t	2.4	2.0	5.0	4.2	4.6	7.9	0.3	10.5	0.5	2.4
T	2.4	4.4	9.4	13.7	18.2	26.1	26.4	36.9	37.4	39.8

Ricordiamo che l' i -esimo elemento del vettore T è il *tempo di attesa* (in minuti) che è stato necessario per osservare l' i -esimo evento. Dobbiamo allora contare quanti eventi sono stati osservati in un intervallo di tempo t *unitario*, ovvero dobbiamo contare il numero X di tempi T minori o uguali a t ; questo numero è la realizzazione della variabile casuale X di Poisson associata. Vediamo alcuni esempi:

- Se poniamo $t = 15$ minuti, nella “finestra temporale” $(0, 15]$, contiamo (osserviamo) 4 eventi, per cui $X = 4$. Infatti, $T_4 < 15$, mentre $T_5 > 15$.
- Se poniamo $t = 30$ minuti, allora $X = 7$, dal momento che nell'intervallo $(0, 30]$, contiamo 7 eventi. Infatti, $T_7 < 30$, mentre $T_8 > 30$.
- Se poniamo $t = 60$ minuti (considerando quindi un intervallo di tempo di un'ora), allora $X \geq 10$, dal momento che l'intervallo complessivo coperto dalla nostra simulazione era inferiore ad un'ora.

⁶In appendice a questo documento se ne trova una traccia.

- Se poniamo $t = 1$ minuto, allora $X = 0$, dal momento che il primo evento è stato osservato dopo 2.4 minuti (quindi, allo scadere del primo minuto non era ancora stato osservato alcun evento). Lo stesso vale se poniamo $t = 2$ minuti.

Per eseguire la simulazione (analoga a quella proposta nel libro), fissiamo un intervallo unitario $t = 15$ minuti ed estraiamo da una esponenziale con parametro $\lambda = 0.25$ un campione di n tempi (scegliendo un valore di n sufficientemente grande da “garantire” che $T_n \geq 15$), contando il numero di eventi osservato nella finestra temporale $(0, 15]$. Ripetiamo questo passo un numero molto grande di volte, in modo da avere un gran numero di realizzazioni di X . Nella simulazione che segue abbiamo ripetuto il processo 100000 volte (`nrep`), estraendo ogni volta 20 (`nexp`) tempi distribuiti come una esponenziale. Qualora $T_{20} < 15$, il risultato viene scartato (ma questo non succede nella nostra simulazione).⁷

```
> lambda <- 0.25
> nrep <- 100000
> nexp <- 20
> ris <- numeric(nrep)
> tot <- matrix(nrow=nrep,ncol=nexp)
> n <- 0
> set.seed(123456)
> for (i in 1:nrep) {
+   t <- rexp(nexp,lambda)
+   if (sum(t) >= 15) {
+     n <- n + 1
+     T <- cumsum(t)
+     tot[n,] <- t
+     x <- length(which(T <= 15))
+     ris[n] <- x
+   }
+ }
> if (n < nrep) {
+   cat("Sono stati scartati ",(nrep-n)," risultati.\n")
+ }
```

⁷Qualora dovesse succedere, saremmo in presenza di *osservazioni troncate (censored)* e i risultati della simulazione andrebbero presi con grande cautela. In questo caso è consigliabile ripetere la simulazione aumentando il valore di n .

```

+   if (n < nrep) tot <- tot[1:n,]
+   if (n < nrep) ris <- ris[1:n]
+ }
>

```

Possiamo fare un *check* sui tempi estratti, verificando i risultati empirici con quelli teorici (ricordiamo che $\lambda = 0.25$).

- La mediana di tutti i tempi estratti è molto vicina al valore teorico:

```

> c(median(tot), -log(0.5)/lambda)
[1] 2.772131 2.772589

```

- La media è quasi identica a $1/\lambda$:

```

> c(mean(tot), 1/lambda)
[1] 3.999751 4.000000

```

- La varianza è del tutto sovrapponibile a $1/\lambda^2$:

```

> c(var(as.vector(tot)), (1/lambda)^2)
[1] 15.98027 16.00000

```

Anche la media delle realizzazioni della variabile casuale X di Poisson è molto vicina al valore teorico 15λ (cioè 3.75):

```

> c(mean(ris), lambda*15)
[1] 3.74705 3.75000

```

e la media e la varianza dei conteggi ottenuti sono del tutto sovrapponibili:

```

> c(mean(ris), var(ris))
[1] 3.747050 3.767884

```

Noi però vogliamo mettere a confronto le frequenze osservate con quelle teoriche di una variabile casuale di Poisson con $\mu = \lambda t = 0.25 \times 15 = 3.75$:


```

> fo <- table(ris)
> fa <- dpois(c(0:14),lambda*15)*n
> cbind(fo,fa=round(fa,1))
      fo      fa
0  2385  2351.8
1   8818  8819.2
2  16627 16535.9
3  20790 20669.9
4  19157 19378.0
5  14412 14533.5
6   9158  9083.4
7   4886  4866.1
8   2299  2281.0
9    932   950.4
10   348   356.4
11   120   121.5
12    43    38.0
13    21    11.0
14     4     2.9

```

Le frequenze osservate sono state ottenute con la funzione `table`, mentre per ottenere quelle teoriche abbiamo prima calcolato le probabilità poissoniane corrispondenti ai valori di X da 0 a 14 e poi le abbiamo moltiplicate per il numero di repliche. Quello che abbiamo “trascurato” (ovvero un numero di eventi maggiore di 14) è davvero “trascurabile”. Infatti $P(X > 14)$ è un numero molto piccolo (circa 10^{-5}):

```

> 1 - ppois(14,lambda*15)
[1] 9.530744e-06

```

al quale è associata una frequenza attesa minore di 1:

```

> sum(fo) - sum(fa)
[1] 0.9530744

```

ragione per cui non c'è assolutamente da stupirsi se, nella nostra simulazione, non abbiamo mai osservato più di 14 eventi in 15 minuti.

Abbiamo detto che, fissato il valore di t ($t = 15$ nel nostro caso) è importante scegliere bene il valore di n (`nexp` nella nostra simulazione), cioè il

numero di tempi da estrarre, in modo che la loro somma sia “quasi certamente” maggiore di (o uguale a) t ($T_n \geq 15$), di modo che in nessuna delle `nrep` repliche $T_n < 15$. Naturalmente, disponendo di `R` possiamo andare “per tentativi”, tenendo conto che, in generale, il “costo” di ripetere una simulazione può essere considerato irrisorio. Possiamo però anche fare una scelta ragionata, basata su un solido fondamento teorico.

Sappiamo⁸ che la somma di n variabili casuali esponenziali i.i.d. con parametro λ è una variabile casuale *Gamma* con parametri n e λ . Nella parametrizzazione usata dalla funzione `pgamma` di `R`, l’argomento `shape` corrisponde a n e l’argomento `rate` corrisponde a λ . Possiamo allora chiedere a `R` di calcolare la probabilità di ottenere un valore minore di 15 in una distribuzione Gamma con parametri n e $\lambda = 0.25$. Per esempio, se $n = 10$, allora la probabilità è:

```
> pgamma(15,10,0.25)
[1] 0.005307174
```

Può sembrare un valore abbastanza basso, ma se facciamo 100000 repliche, allora dobbiamo aspettarci circa 531 campioni di 10 tempi esponenziali la cui somma è inferiore a 15 e questo non può certo andare bene. Per scegliere un opportuno valore di n calcoliamo la probabilità di estrarre da una distribuzione Gamma con $\lambda = 0.25$ un campione di n tempi (facendo variare n) la cui somma è minore di 15.

```
> n <- c(11:20)
> p <- pgamma(15,n,0.25)
> cbind(n,p,round(p*100000))
      n      p
[1,] 11 1.743111e-03 174
[2,] 12 5.280897e-04  53
[3,] 13 1.483955e-04  15
[4,] 14 3.886838e-05   4
[5,] 15 9.530744e-06   1
[6,] 16 2.196335e-06   0
[7,] 17 4.773333e-07   0
[8,] 18 9.814167e-08   0
[9,] 19 1.914342e-08   0
[10,] 20 3.551658e-09   0
```

⁸Lo abbiamo scritto a pag. 137 del libro.

L'ultima colonna riporta il numero atteso di campioni per i quali $T_n < 15$. Come vediamo, se avessimo scelto $n = 15$, ci si poteva ancora aspettare un campione per il quale $T_n < 15$, mentre già con $n = 16$ questo valore era inferiore a 1 (essendo circa 0.2). Un valore $n = 20$ garantiva “al di là di ogni ragionevole dubbio” che nessuna delle 100000 repliche avrebbe fatto registrare $T_{20} < 15$ (nel senso che avremmo dovuto fare oltre 300 *milioni* di repliche per aspettarne una per la quale $T_{20} < 15$).

Appendice

Seguiremo la linea di ragionamento illustrata da Roe Goodman,⁹ che prende in esame un “modello di rimpiazzo” (*replacement model*), nel quale un insieme di componenti elettronici hanno, ciascuno, una durata di vita modellata da una legge esponenziale con parametro λ . Indichiamo con t_1, t_2, \dots, t_n le “durate” (i *tempi di vita*) dei componenti $1, 2, \dots, n$ che possiamo considerare come variabili casuali i.i.d. tutte con lo stesso parametro λ .

Installiamo il primo componente e “accendiamo la macchina” al tempo $t = 0$. Non appena questo componente si rompe, viene immediatamente sostituito con il secondo e così via indefinitamente. Indichiamo con $N(t)$ il numero di componenti che si sono “rotti” fino al tempo t incluso. Evidentemente $N(0) = 0$. Possiamo pensare che, per $t > 0$, $N(t)$ sia una *famiglia* di variabili casuali della quale vogliamo trovare le proprietà probabilistiche.

Come abbiamo accennato in precedenza in questo documento, $N(t)$ è una variabile casuale *discreta* che dipende da un parametro continuo $t \geq 0$. D’altro canto $T_n = t_1 + t_2 + \dots + t_n$ è una variabile casuale *continua* che dipende da un parametro discreto n .

Dal momento che T_n è la somma di n variabili casuali esponenziali i.i.d., seguirà una distribuzione Gamma con parametri n e λ . Inoltre, per qualsiasi $t \geq 0$, i seguenti due eventi “coincidono”

$$\{T_n \leq t\} = \{N(t) \leq n\}$$

nel senso che se la durata totale dei primi n componenti è al massimo t , ne consegue che almeno n componenti si sono rotti entro il tempo t (e viceversa). L’uguaglianza precedente rappresenta il legame fra il parametro discreto n e il parametro continuo t . Vogliamo ora calcolare la probabilità associata a ciascuno dei due eventi:

$$P\{N(t) \leq n\} = P\{T_n \leq t\}.$$

È immediato calcolare la seconda, dal momento che è la funzione di ripartizione di una Gamma con parametri n e λ :

$$P\{N(t) \leq n\} = P\{T_n \leq t\} = \frac{\lambda}{(n-1)!} \int_0^t (\lambda x)^{n-1} e^{-\lambda x} dx.$$

⁹R. Goodman (1988). *Introduction to Stochastic Models*. The Benjamin/Cummings Publishing Company, Inc. Menlo Park, California. (pagg. 193–194)

Questo integrale può essere calcolato per parti, ponendo $u = e^{-\lambda x}$ e

$$dv = \frac{\lambda}{(n-1)!} (\lambda x)^{n-1} dx.$$

In questo modo $du = -\lambda e^{-\lambda x} dx$ e

$$v = \frac{(\lambda x)^n}{n!}.$$

Abbiamo così

$$P\{N(t) \geq n\} = \frac{(\lambda t)^n}{n!} e^{-\lambda t} + \lambda \int_0^t \frac{(\lambda x)^n}{n!} e^{-\lambda x} dx.$$

L'integrale che abbiamo appena scritto non è altro che $P\{N(t) \leq n+1\}$:

$$P\{N(t) \geq n+1\} = \frac{\lambda}{n!} \int_0^t (\lambda x)^n e^{-\lambda x} dx.$$

Pertanto:

$$P\{N(t) \geq n\} = \frac{(\lambda t)^n}{n!} e^{-\lambda t} + P\{N(t) \geq n+1\}.$$

Dal momento che

$$P\{N(t) \geq n\} - P\{N(t) \geq n+1\} = P\{N(t) = n\},$$

abbiamo che

$$P\{N(t) = n\} = \frac{(\lambda t)^n}{n!} e^{-\lambda t},$$

ovvero che $N(t)$ segue una distribuzione di Poisson con parametro $\mu = \lambda t$.