La variabile casuale Gamma

Abbiamo già visto[1] come la somma di due variabili casuali esponenziali indipendenti con lo stesso parametro λ sia un caso particolare di una variabile casuale chiamata *Gamma* e come questa possa essere vista come la somma di n variabili casuali esponenziali i.i.d.[2] Questo ci dice che la distribuzione di probabilità di una variabile casuale Gamma ha due parametri: n e λ.
Possiamo dire che la variabile casuale Gamma è una variabile casuale piuttosto "versatile", dal momento che essa "genera" altre variabili casuali. Prima di considerarla più in dettaglio è però opportuno considerare la *funzione Gamma*.

La funzione Gamma

La *funzione Gamma* (di Eulero) è il risultato del seguente integrale:

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt \tag{1}$$

("Γ" è la lettera greca "gamma" maiuscola). Per quanto ci riguarda, l'integrale converge per tutti i valori reali di x maggiori di zero.
È facile calcolare quanto vale $\Gamma(1)$. Infatti, in questo caso l'integrale diventa

$$\int_0^\infty e^{-t} dt$$

e, quindi, il risultato è 1.
Quando $x = 2$ l'integrale da valutare è

$$\int_0^\infty t e^{-t} dt.$$

Usando l'integrazione per parti, troviamo che una primitiva è

$$-(1+t)e^{-t}$$

e, quindi, l'integrale vale (ancora) 1.

[1] A pag. 136 del libro.

[2] Questo è vero anche quando $n = 1$, per cui la variabile casuale esponenziale è un caso particolare della variabile casuale Gamma.

Sfruttando questo risultato si può calcolare

$$\int_0^{\infty} t^2 e^{-t} dt$$

una cui primitiva è

$$-(2 + 2t + t^2)e^{-t}$$

per cui l'integrale vale, in questo caso, 2.
Procedendo oltre, assegnando a x sempre valori naturali, "la" primitiva diventa sempre più "complicata". Possiamo però usare R per calcolare alcuni valori di $\Gamma(x)$:

```
> x <- c(1:9)
> rbind(x,gamma(x))
x    1    2    3    4    5    6    7    8     9
     1    1    2    6   24  120  720 5040 40320
```

osservando che i risultati corrispondono al fattoriale di $x - 1$.
Questo risultato può essere dimostrato calcolando $\Gamma(x+1)$ a partire da $\Gamma(x)$ usando l'integrazione per parti. Scriviamo innanzitutto $\Gamma(x + 1)$:

$$\Gamma(x + 1) = \int_0^{\infty} t^x e^{-t} dt.$$

Poniamo ora $u = t^x$ e $dv = e^{-t}dt$. In questo modo $du = xt^{x-1}dt$ e $v = -e^{-t}$. Abbiamo quindi

$$\Gamma(x + 1) = \Big[- t^x e^{-t} \Big]_0^{\infty} + x \int_0^{\infty} t^{x-1} e^{-t} dt.$$

Il primo addendo vale 0, mentre l'integrale che compare al secondo addendo vale $\Gamma(x)$, per cui si ottiene

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

e questo vale $\forall x > 0$.
Sapendo quindi che $\Gamma(1) = 1$, possiamo calcolare $\Gamma(2) = 1 \times \Gamma(1) = 1$, $\Gamma(3) = 2 \times \Gamma(2) = 2$, $\Gamma(4) = 3 \times \Gamma(3) = 3 \times 2$, $\Gamma(5) = 4 \times \Gamma(4) = 4 \times 3 \times 2$, $\Gamma(6) = 5 \times \Gamma(5) = 5 \times 4 \times 3 \times 2$, e così via. Abbiamo quindi

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

quando n è un intero maggiore o uguale a 1.
Possiamo quindi pensare che la funzione Gamma estenda il concetto di fattoriale a valori frazionari (maggiori di zero); per esempio

```
> gamma(c(0.5,1.5,2.5))
[1] 1.7724539 0.8862269 1.3293404
```

Il primo dei tre valori proposti,cioè $\Gamma(1/2)$, è, in effetti, $\sqrt{\pi}$:

```
> sqrt(pi)
[1] 1.772454
```

e potrebbe essere calcolato "direttamente" ricorrendo alla sostituzione $u = \sqrt{t}$ nell'integrale (1) che definisce la funzione Gamma, dal momento che questo diventa l'*integrale di Gauss* (vedi il documento nell'Appendice *on line* relativo alla Normale).

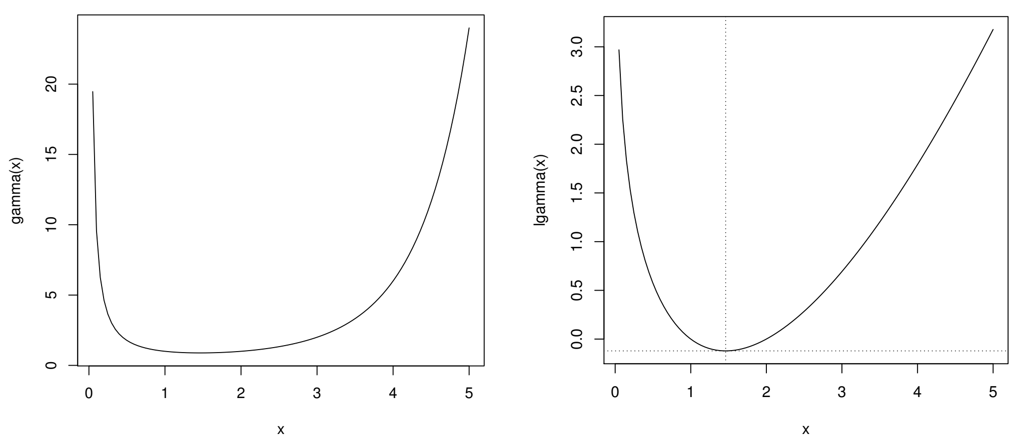

Figura 1: A sinistra: grafico della funzione Gamma. A destra: grafico del logaritmo naturale della funzione Gamma.

Nel pannello di sinistra della figura 1 abbiamo disegnato con R il grafico della funzione Gamma mediante l'istruzione `curve(gamma,0.05,5)`. La funzione ha un minimo[3]. quando x è circa 1.46163; questo minimo vale circa 0.88560

[3] Alcuni dettagli su come è stato calcolato questo minimo si trovano nell'articolo *The Minimum in the Gamma Function* di Deming e Colcord, pubblicato su *Nature* (June 1, 1935, pag. 917)

(quindi $\Gamma(x) > 0, \forall x > 0$). La posizione del minimo si vede meglio quando si riportano in ordinata i valori del logaritmo (naturale) della funzione Gamma (mediante la funzione `lgamma`), come abbiamo fatto nel pannello di destra della figura 1

```
curve(lgamma,0.05,5)
abline(v=1.46163, h=log(0.88560), lty=3)
```

oppure, dal momento che $\Gamma(1) = \Gamma(2)$, facendo uno *zoom* della funzione Gamma fra 1 e 2 con l'istruzione `curve(gamma,1,2)`.

$\Gamma(x)$ non è definito quando $x \leq 0$. Per valori positivi molto (molto) piccoli di x, $\Gamma(x) \simeq 1/x$:

```
> x <- 10^(-c(4:9))
> g <- gamma(x)
> rbind(x,g)
x    0.0001     0.00001 1.000000e-06 1.000000e-07 1e-08 1e-09
g 9999.4229 99999.42279 9.999994e+05 9.999999e+06 1e+08 1e+09
```

Una proprietà che può tornare molto utile è la seguente: per ogni $k > 0$

$$\int_0^\infty t^{x-1} e^{-kt} dt = k^{-x}\Gamma(x) \qquad (2)$$

(lo si può dimostrare mediante la sostituzione $u = kt$, per cui $t = u/k$ e $dt = 1/k\ du$).

La distribuzione Gamma

Pensando la distribuzione Gamma come la distribuzione della somma di n variabili casuali esponenziali i.i.d. con parametro λ, la funzione di densità di probabilità ha due parametri: n e λ. Il primo (n) è un *parametro di forma*, il secondo (λ) è un *parametro di velocità* (*rate parameter*); un'altra parametrizzazione prende $1/\lambda$ come secondo parametro (invece di λ); in questo caso il secondo parametro viene detto *parametro di scala*. In base alla definizione adottata, dovremmo avere $\lambda > 0$ e $n > 0$ ($n \in \mathbb{N}$). In effetti, grazie alla definizione (1) della funzione Gamma, n può assumere qualsiasi valore reale maggiore di zero ($n \in \mathbb{R}_+$).

Seguendo per il momento la "nostra" definizione della variabile casuale Gamma (somma di n variabili casuali esponenziali i.i.d.), è immediato calcolarne il valore atteso (n/λ) e la varianza (n/λ^2). Questo rimane valido anche se $n \in \mathbb{R}_+$.
La funzione di densità di probabilità della variabile casuale Gamma può essere scritta nel modo seguente:

$$f(x) = \frac{x^{n-1}\lambda^n e^{-\lambda x}}{\Gamma(n)} \tag{3}$$

e i valori $f(x)$ possono essere richiesti a R mediante la funzione `dgamma` specificando come primo argomento il valore (o i valori) di x, come secondo argomento (`shape`) il valore di n e come terzo argomento (`rate`) il valore di λ. In alternativa a `rate` si può specificare l'argomento `shape`, cioè $1/\lambda$ (la media dell'esponenziale). Quando $\lambda = 1$, la distribuzione Gamma viene spesso indicata come distribuzione Gamma *standard*.
L'integrale della (3) esteso fra 0 e infinito deve ovviamente essere uguale a 1 (altrimenti non sarebbe una distribuzione di probabilità). Dimostrarlo è facile ricorrendo alla proprietà (2). Infatti, considerando per il momento soltanto il numeratore della (3), abbiamo

$$\int_0^\infty x^{n-1}\lambda^n e^{-\lambda x} dx = \lambda^n \int_0^\infty x^{n-1} e^{-\lambda x} dx = \lambda^n \lambda^{-n}\Gamma(n) = \Gamma(n).$$

Considerando ora il denominatore della (3), vediamo come il rapporto sia uguale a 1. Pertanto

$$\int_0^\infty \frac{x^{n-1}\lambda^n e^{-\lambda x}}{\Gamma(n)} dx = 1.$$

Nella figura 2 abbiamo disegnato con R il grafico della funzione di densità di probabilità della variabile casuale Gamma per alcuni valori di n e λ. Nel pannello di sinistra della figura 2 abbiamo mantenuto costante il valore $\lambda = 2$, mentre nel pannello di destra abbiamo mantenuto costante il valore $n = 5$.

Qui di seguito elenchiamo alcune importanti proprietà della variabile casuale Gamma:

- Se $X_1, X_2, \ldots, X_k$ sono k variabili casuali indipendenti, ciascuna distribuita come una Gamma con lo stesso parametro λ, ma con parametro di forma variabile $(n_1, n_2, \ldots, n_k)$, allora la loro somma è ancora una

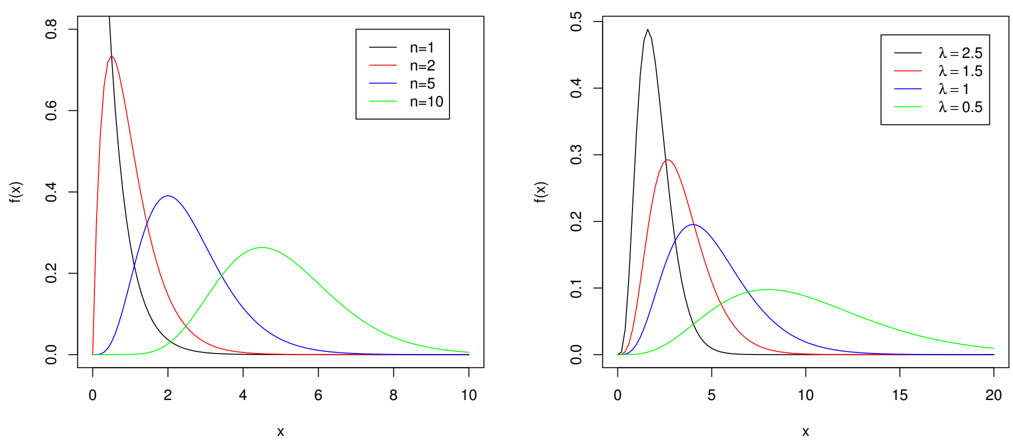

Figura 2: A sinistra: grafico della funzione di densità di probabilità della variabile casuale Gamma per diversi valori di n quando $\lambda = 2$. A destra: grafico della funzione di densità di probabilità della variabile casuale Gamma per diversi valori di λ quando $n = 5$.

Gamma con parametro di forma $n = n_1 + n_2 + \ldots + n_k$) e con parametro di velocità λ.

- Se moltiplichiamo una variabile casuale Gamma con parametri n e λ per una costante k, la variabile casuale risultante è ancora una Gamma con parametri n e $k\lambda$.

- Se $\lambda = 1/2$ e $n = \nu/2$, allora la variabile casuale Gamma è identica ad una variabile casuale chi quadrato con ν gradi di libertà.

- Se n è un intero, allora la funzione di densità di probabilità della variabile casuale Gamma descrive il *tempo di attesa* fino all'$n-$esimo "evento" in un processo di Poisson unidimensionale con tasso λ.

- Al crescere di n, la distribuzione Gamma converge alla Normale con media n/λ e varianza n/λ^2.