Introduzione ai Modelli di Durata: Alcuni Modelli Parametrici

a.a. 2009/2010 - Quarto Periodo

Prof. Filippo DOMMA

Corso di Laurea Specialistica/Magistrale in Economia Applicata

Facoltà di Economia – UniCal

Modelli Parametrici

1. Esponenziale

Le funzioni di densità, ripartizione e sopravvivenza, rispettivamente, sono:

$$f(t;\lambda) = \lambda e^{-\lambda t} \qquad t \geq 0 \qquad \lambda > 0$$

$$F(t;\lambda) = 1 - e^{-\lambda t} \qquad S(t;\lambda) = e^{-\lambda t}$$

Il valore atteso è: $E(T) = \lambda$

Proprietà: "perdita di memoria"
(tale proprietà è poco realistica)

$$P_r(T \geq t_0 + t / T \geq t_0) = P_r(T \geq t)$$

Di conseguenza, la vita media residua è: $E(T - t / T \geq t) = E(T) = \lambda$

Dalla definizione di *hazard function* si ha:

$$h(t;\lambda)=\frac{f(t;\lambda)}{S(t;\lambda)}=\frac{\lambda e^{-\lambda t}}{e^{-\lambda t}}=\lambda$$

Per la v.c. esponenziale la *hazard function* è costantemente pari a λ.

Esempio.

10, 7, 32*, 23, 22, 6, 16, 34*, 32*, 25*, 11*, 20*, 19*, 6, 17*, 35*, 6, 13, 9*, 6*, 10*
Il simbolo * indica che la durata è censurata.
Ci sono 9 durate complete e 12 censurate. Supponiamo di voler adattare un modello Esponenziale e troviamo lo stimatore di m.v. del parametro λ.

$$L(\lambda;\mathbf{t}) = \prod_{i=1}^{n} [f(t_i;\lambda)]^{\delta_i} [S(t_i;\lambda)]^{1-\delta_i} = \prod_{i=1}^{n} \lambda^{\delta_i} e^{-\lambda t_i \delta_i} e^{-\lambda t_i (1-\delta_i)} = \lambda^{\sum_{i=1}^{n}\delta_i} e^{-\lambda \sum_{i=1}^{n} t_i}$$

$$\ell(\lambda;\mathbf{t}) = \ln(\lambda)\sum_{i=1}^{n}\delta_i - \lambda\sum_{i=1}^{n} t_i \qquad \frac{\partial \ell(\lambda;\mathbf{t})}{\partial\lambda} = \frac{\sum_{i=1}^{n}\delta_i}{\lambda} - \sum_{i=1}^{n} t_i = 0 \Leftrightarrow \hat{\lambda} = \frac{\sum_{i=1}^{n}\delta_i}{\sum_{i=1}^{n} t_i}$$

$$\hat{\lambda} = \frac{9}{359} = 0.025069$$

2. Weibull

La distribuzione di Weibull è largamente utilizzata nell'analisi dei dati di durata (al punto da far pensare che molti lavori presenti in letteratura sono Weibull-dipendenti), ciò dipende da fatto che presenta una *hazard function* monotono; in particolare, può essere sempre decrescente, sempre crescente o costante a seconda del valore del parametro β. E' importante sottolineare che nelle sue diverse forme non presenta quello non-monotono. Inoltre, include come caso particolare la esponenziale ed è legata ad altre distribuzioni utilizzati nell'ambito di tali studi.
La funzione di densità e di sopravvivenza, rispettivamente, sono:

$$f(t;\lambda,\beta) = \lambda\beta(\lambda t)^{\beta-1}\exp\left(-(\lambda t)^{\beta}\right)$$

$$S(t;\lambda,\beta) = \exp\left(-(\lambda t)^{\beta}\right)$$

dove λ e β governano, rispettivamente, la scala e la forma della distribuzione.

La *hazard function* risulta essere:

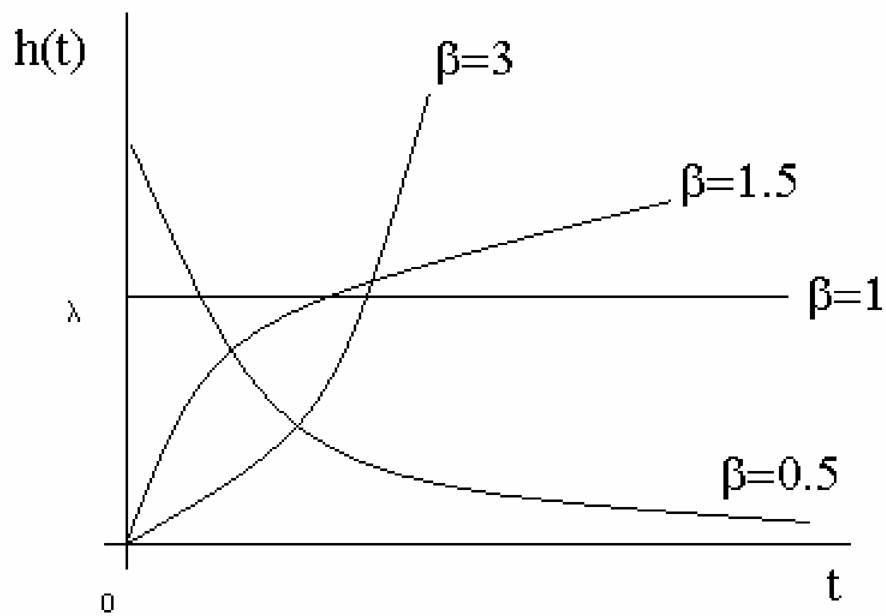
$$h(t;\beta,\lambda) = \lambda\beta\left(\lambda t\right)^{\beta-1}$$

Si osservi che per ogni t>0, si ha:

$$\frac{\partial h(t;\beta,\lambda)}{\partial t} = \lambda^{\beta}\left(\beta-1\right)t^{\beta-2}\begin{cases} >0 & \text{se} \quad \beta>1 \\ =0 & \text{se} \quad \beta=1 \\ <0 & \text{se} \quad \beta<1 \end{cases}$$

L'andamento della *hazard function* è influenzato solo ed esclusivamente dal parametro β. Per $\beta=1$ otteniamo il modello esponenziale come caso particolare.

Andamenti della *hazard function* di un modello Weibull al variare di β.



3. Log-Normale

Il modello log-Normale, analogamente a quello di Weibull, è stato intensamente utilizzato per descrivere i dati di sopravvivenza.
Si dice che T ha distribuzione log-Normale se Y=lg(T) si distribuisce secondo una Normale di parametri, diciamo, μ e σ^2.
La funzione di densità e di sopravvivenza, rispettivamente, sono:

$$f(t;\mu,\sigma)=\frac{1}{t\sigma\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{\lg(t)-\mu}{\sigma}\right)^2\right) \qquad t>0$$

$$S(t;\mu,\sigma)=1-\Phi\left(\frac{\lg(t)-\mu}{\sigma}\right)$$

dove $\Phi(.)$ è la funzione di ripartizione della normale standardizzata.

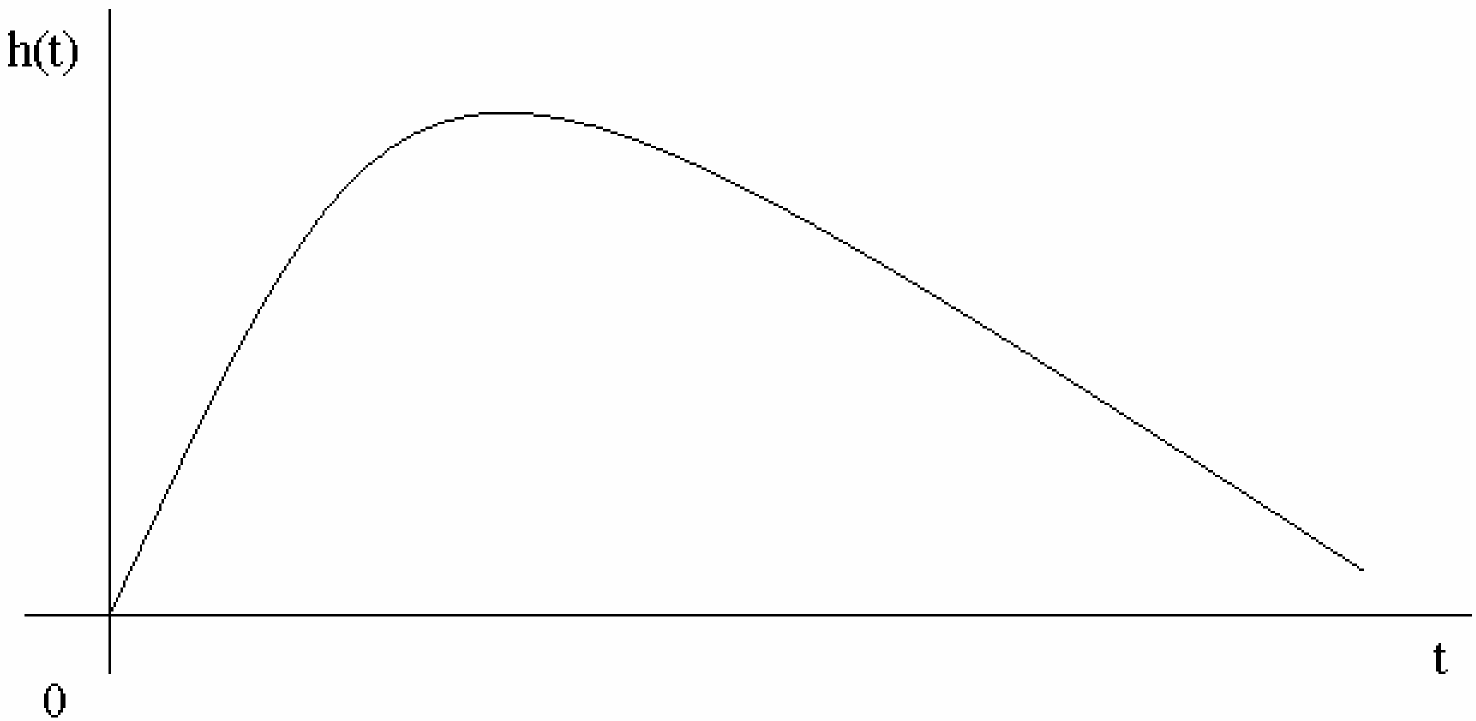
Momento primo e varianza sono:

$$E(T) = \exp\left\{\mu + \frac{\sigma^2}{2}\right\} \qquad V(T) = \left\{\exp\left(\sigma^2\right) - 1\right\} \times \exp\left\{2\mu + \sigma^2\right\}$$

La *hazard function* non ha una espressione esplicita

$$h(t;\mu,\sigma) = \frac{f(t;\mu,\sigma)}{S(t;\mu,\sigma)}$$

Si dimostra che per t=0, h(t) è zero, all'aumentare di t cresce fino a raggiungere un massimo, per poi decrescere è tendere a zero quando t tende ad infinito.



Tipico andamento della *hazard function* di una Log-Normale.

4. Log-Logistica

Una v.c. T segue una distribuzione log-logistica se Y=lg(T) segue una distribuzione Logistica. La funzione di densità e di ripartizione di T, rispettivamente, sono:

$$f(t;\alpha,\lambda)=\frac{\alpha\lambda t^{\alpha-1}}{\left(1+\lambda t^{\alpha}\right)^{2}} \qquad S(t;\alpha,\lambda)=\frac{1}{1+\lambda t^{\alpha}}$$

E’ immediato verificare che la *hazard function* è:

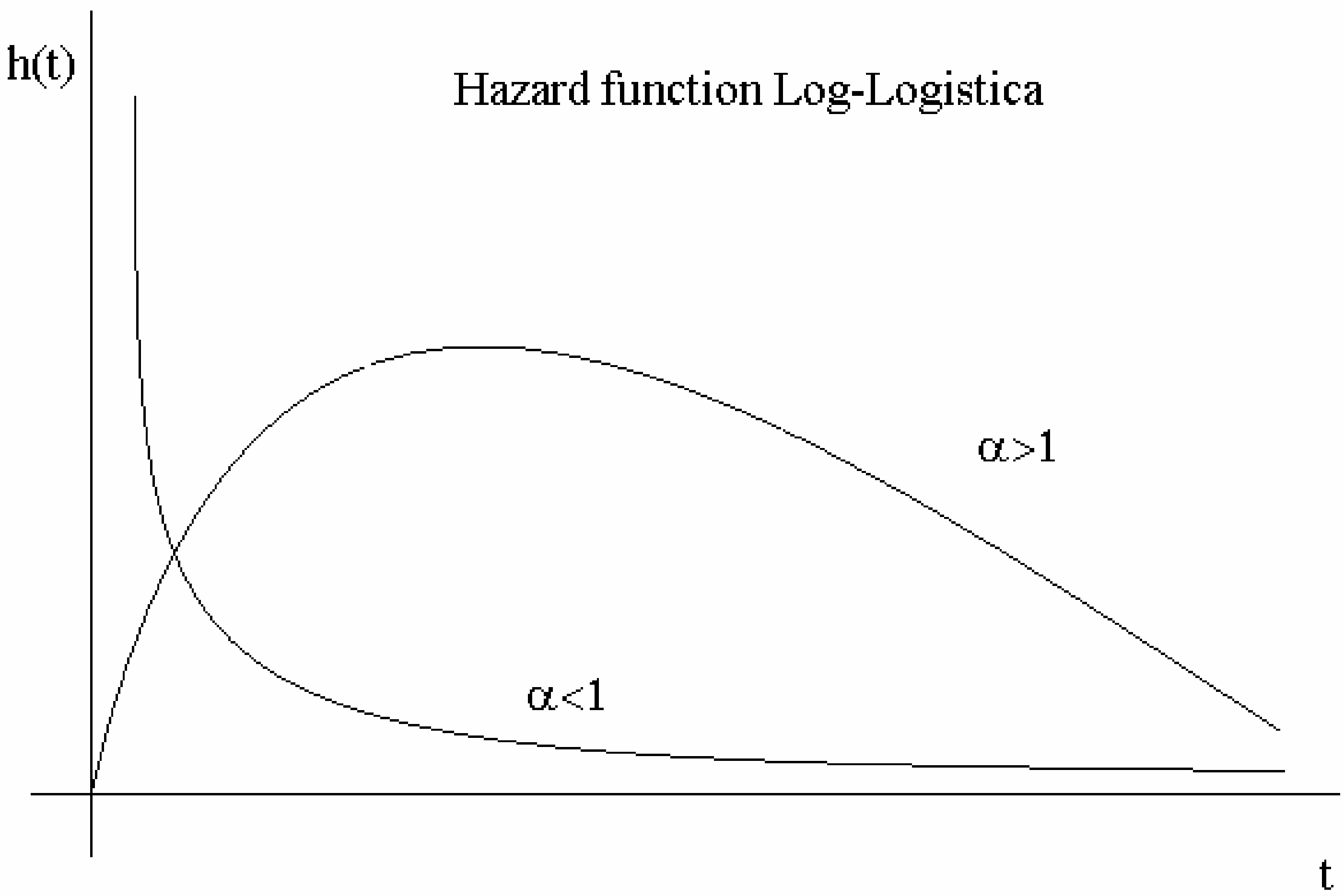
$$h(t;\alpha,\lambda)=\frac{\alpha\lambda t^{\alpha-1}}{1+\lambda t^{\alpha}}$$

$$\frac{\partial h(t;\alpha,\lambda)}{\partial t} = \frac{\alpha\lambda t^{\alpha-2}}{\left(1+\lambda t^{\alpha}\right)^{2}}\left\{(\alpha-1)-\lambda t^{\alpha}\right\}$$

Se $\alpha \leq 1$ allora $h(t;\alpha,\lambda)$ è sempre <u>decrescente</u>;

Se $\alpha > 1$ allora $h(t;\alpha,\lambda)$ presenta un <u>massimo</u> in

$$t = \left[\frac{(\alpha-1)}{\lambda}\right]^{\frac{1}{\alpha}}$$


h(t)
Hazard function Log-Logistica
α>1
α<1
t

5. Burr XII

Negli ultimi anni, tra i modelli più utilizzati, in particolar modo per l'analisi dei dati di durata in economia, vi è il modello Burr XII. Tale modello presenta le seguenti funzioni di densità e di sopravvivenza:

$$f(t; c_1, c_2, c_3) = \frac{c_1 c_3}{c_2} t^{c_3 - 1} \left[1 + \frac{t^{c_3}}{c_2} \right]^{-c_1 - 1}$$

$$S(t; c_1, c_2, c_3) = \left[1 + \frac{t^{c_3}}{c_2} \right]^{-c_1} = \left[\frac{c_2}{c_2 + t^{c_3}} \right]^{c_1}$$

con $c_i > 0$ per $i=1,2,3$.

Si noti che la v.c. Burr XII può essere vista come una generalizzazione della v.c. Log-Logistica; infatti, basta porre $c_1=1$, $c_2=1/\lambda$ e $c_3=\alpha$, ad esempio, nella funzione di sopravvivenza per ottenere la funzione di sopravvivenza della log-logistica.

Il modello Burr XII presenta la seguente *hazard function*

$$h(t;c_1,c_3) = \frac{c_1 c_3 t^{c_3-1}}{\left(c_2 + t^{c_3}\right)}$$

Si dimostra che:

se $c_3 \leq 1$ allora $h(t;c_1,c_2,c_3)$ è sempre <u>decrescente</u>;

se $c_3 > 1$ allora $h(t;c_1,c_2,c_3)$ presenta un <u>massimo</u> in

$$t^* = \left[(c_3 - 1)c_2\right]^{\frac{1}{c_3}}$$

6. Modello di Hjorth (1980) – Hazard function bathtub

Al fine di costruire un modello capace di descrivere la vita delle persone, delle cellule ect., Hjorth (1980) propone la seguente *hazard function*:

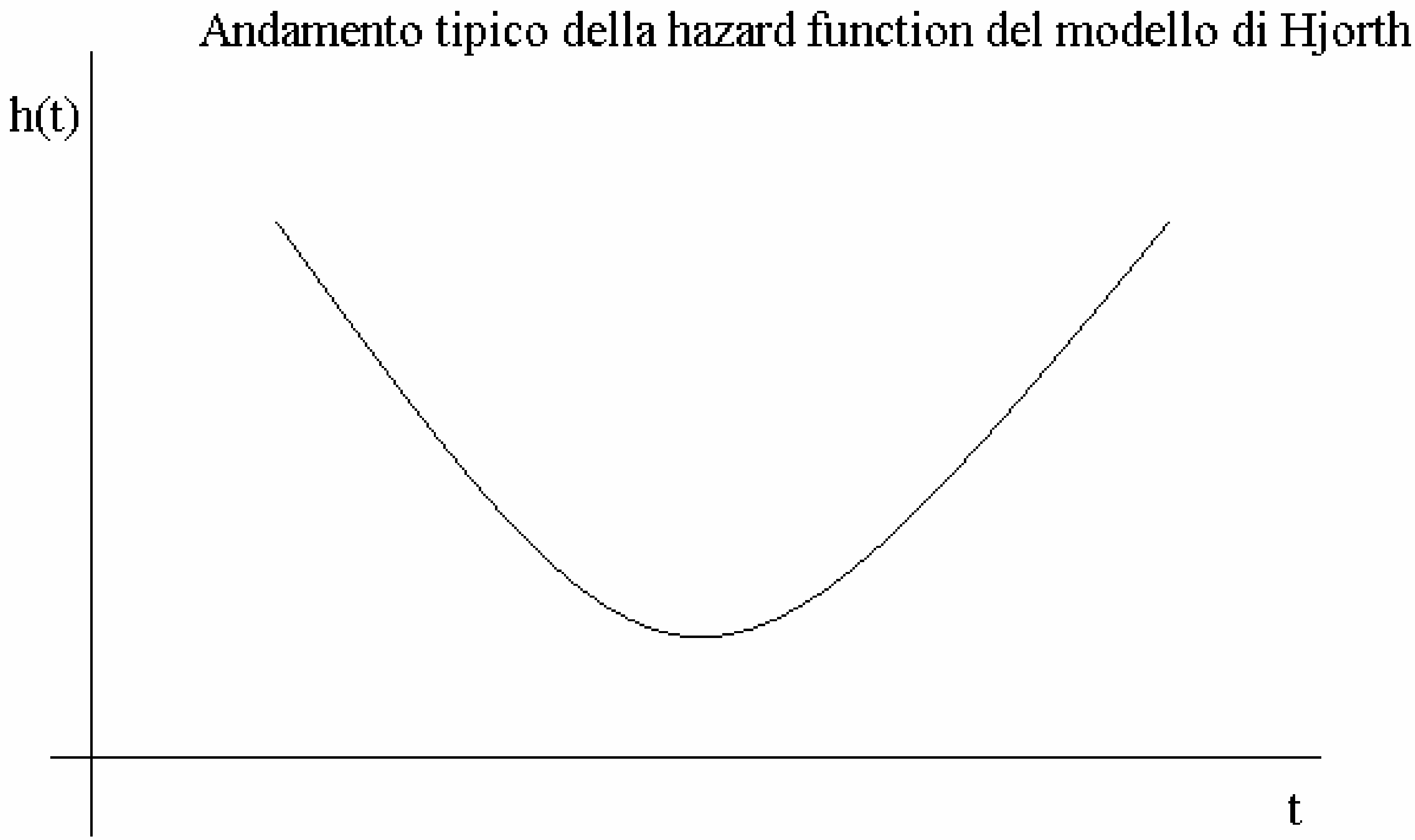
$$h(t;\beta,\lambda,\alpha) = \frac{\beta}{t+\alpha} + \delta t$$

E' semplice verificare che

$$\frac{\partial h(t;\beta,\lambda,\alpha)}{\partial t} = \frac{-\beta}{(t+\alpha)^2} + \delta = 0 \quad \Leftrightarrow \quad t^* = \sqrt{\frac{\beta}{\delta}} - \alpha$$

$$\frac{\partial^2 h(t;\beta,\lambda,\alpha)}{\partial t^2} = \frac{2\beta}{(t+\alpha)^3} > 0$$

t^* è un punto di **minimo.**


Andamento tipico della hazard function del modello di Hjorth
h(t)
t

Utilizzando la relazione esistente tra la hazard function e la funzione di sopravvivenza, abbiamo:

$$S(t;\beta,\lambda,\alpha) = \exp\left\{-\beta\int_0^t \frac{1}{u+\alpha}du - \delta\int_0^t udu\right\} = \frac{\alpha^{\beta}e^{-\frac{\delta t^2}{2}}}{(t+\alpha)^{\beta}}$$

$$f(t;\beta,\lambda,\alpha) = \left[\frac{\beta}{t+\alpha} + \delta t\right]\frac{\alpha^{\beta}e^{-\frac{\delta t^2}{2}}}{(t+\alpha)^{\beta}}$$

7. Burr III [oppure Dagum (1977, 1980)]

$$f(t;\beta,\lambda,\theta) = \beta\lambda\theta\ t^{-\theta-1}\left[1+\lambda t^{-\theta}\right]^{-\beta-1}$$

$$S(t;\beta,\lambda,\theta) = 1-\left[1+\lambda t^{-\theta}\right]^{-\beta}$$

$$h(t;\beta,\lambda,\theta) = \frac{\beta\lambda\theta\ t^{-\theta-1}\left[1+\lambda t^{-\theta}\right]^{-\beta-1}}{1-\left[1+\lambda t^{-\theta}\right]^{-\beta}}$$

Si dimostra (non è immediata) che $h(t;\beta,\lambda,\theta)$ può essere sempre decrescente, può presentare un massimo oppure un minimo ed un massimo.

Figura 1. Hazard function: caso UBT.

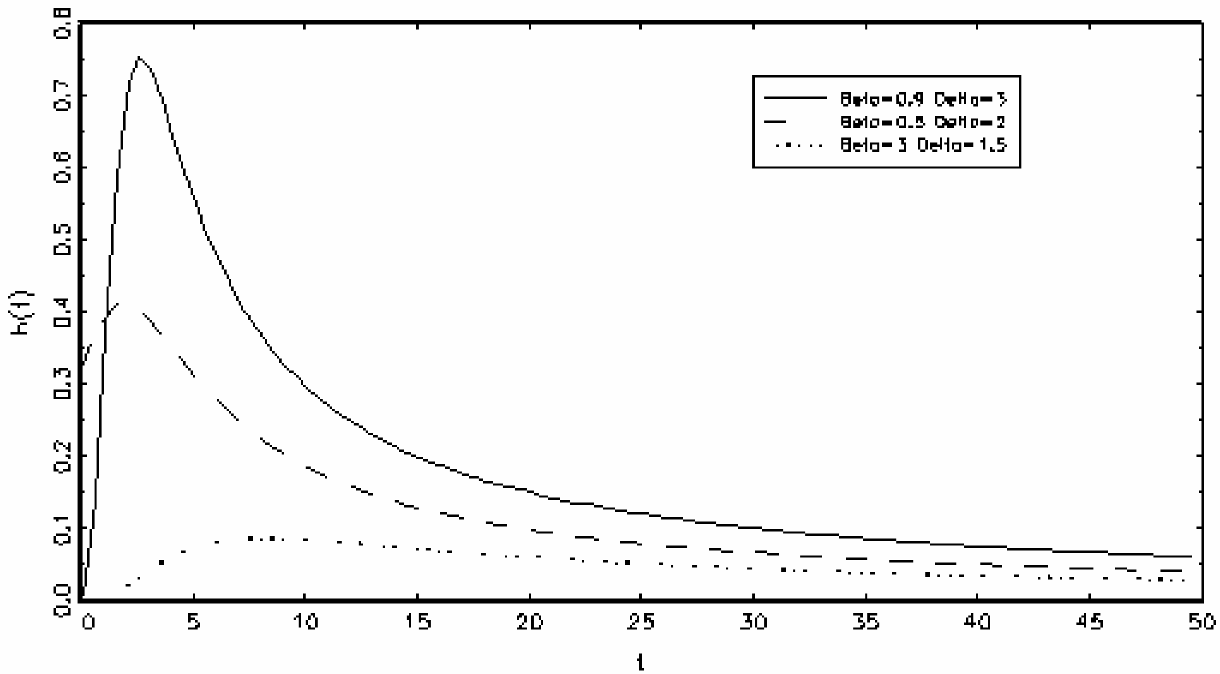


Figura 2. Hazard function: caso IFR.

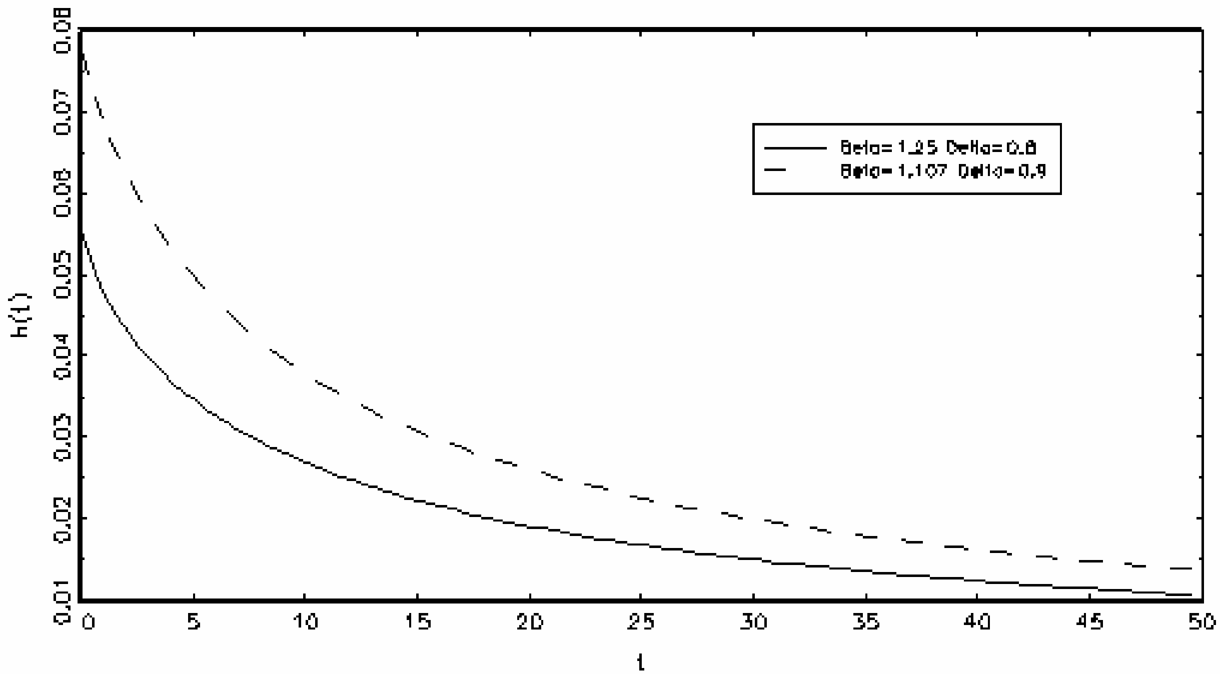


Figura 3. Hazard function: caso BT-UBT.