Ανάλυση Επιβίωσης

Επικ. Καθ. Σ. Ζημερας
Τμήμα Στατιστικής και Αναλογιστικών – Χρηματοοικονομικών Μαθηματικών
Πανεπιστήμιο Αιγαίου
Σάμος
2020

Κατανομές ανάλυσης επιβίωσης

- Θα παρουσιαστούν οι βασικότερες συνεχείς κατανομές (ή αλλιώς παραμετρικά μοντέλα) που εφαρμόζονται στην ανάλυση επιβίωσης. Υποθέτοντας ότι ο χρόνος αποτυχίας έχει σ.π.π. $f(t)$
- Επομένως παραμετρικές συναρτήσεις επιβίωσης ορίζονται ως εξής:

Κατανομές ανάλυσης επιβίωσης

- **Εκθετική κατανομή**

Μία τ.μ. ακολουθεί την εκθετική κατανομή (exponential distribution), $T \sim Exp(\lambda)$, $\lambda > 0$, αν

έχει σ.π.π. : $f(t) = \lambda e^{-\lambda t},\ t \geq 0, \lambda > 0$

συνάρτηση επιβίωσης : $S(t) = \int_t^{\infty} f(u;\lambda)\,du = \int_t^{\infty} \lambda e^{-\lambda u}\,du = e^{-\lambda t}$

σταθερή συνάρτηση κινδύνου : $h(t) = \dfrac{f(t;\lambda)}{S(t;\lambda)} = \lambda$, λ σταθερά

αθροιστική συνάρτηση κινδύνου : $H(t) = \int_0^t h(u)\,du = \int_0^t \lambda\,du = \lambda t$ **(2.4),**

και διάμεσο : $0.5 = S(\tau) = e^{-\lambda\tau} \Rightarrow \tau = \dfrac{-\log(0.5)}{\lambda}$.

Στην περίπτωση αυτή, έχουμε το λεγόμενο εκθετικό μοντέλο επιβίωσης.

Κατανομές ανάλυσης επιβίωσης

- **Εκθετική κατανομή**

Μπορεί να δειχθεί ότι $E(T)=\dfrac{1}{\lambda}$ και $Var(T)=\dfrac{1}{\lambda^{2}}$.

συντελεστής μεταβλητότητας ισούται με τη μονάδα.

$$C\ \ V\ \ (X)=\frac{\sigma}{\mu}100\%$$

Η εκθετική κατανομή αποτελεί ειδική περίπτωση της Gamma κατανομής αφού για $a=1$ οδηγούμαστε στην εκθετική κατανομή ($Exp(\lambda)\ ..Gamma(a=1,\lambda)$).

Ισχύουν επίσης τα εξής :

- $Exp(\lambda)\ ..Weibull(1,\lambda)$ (εκθετική ειδική περίπτωση Weibull κατανομής).
- Εάν $T\quad Exp(\lambda)$ τότε $\theta e^{T}\quad Pareto(\theta,\lambda)$.
- Εάν $T\quad Exp(\lambda)$ τότε $(T)^{1/a}\quad Weibull(a,\lambda)$.

Κατανομές ανάλυσης επιβίωσης

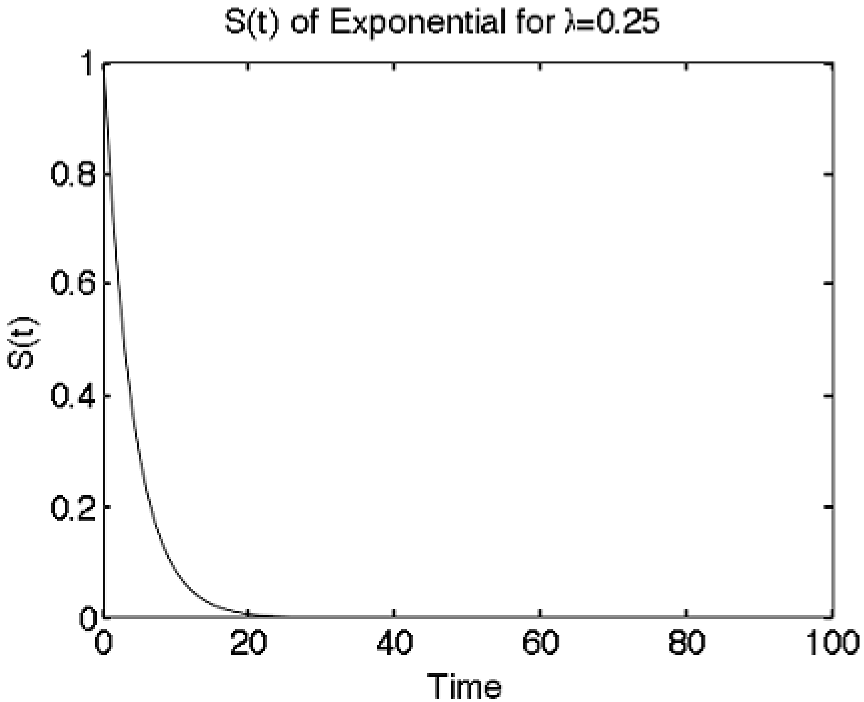
- **Εκθετική κατανομή**

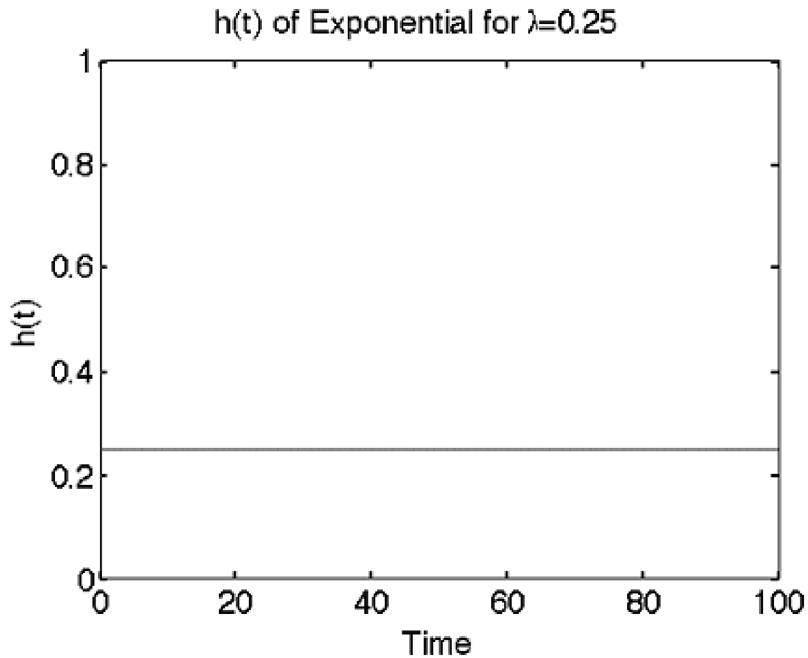
Η εκθετική κατανομή δεν είναι ιδιαίτερα ευέλικτη. Λόγω της ιδιότητάς της, **έλλειψη μνήμης (memoryless property)**, η συνάρτηση κινδύνου παραμένει σταθερή στον χρόνο.

Δηλαδή για δοσμένα $t > 0$, $s > 0$ ισχύει : $P(X > t + s | X > s) = P(X > t)$.

Από πλευράς μοντελοποίησης αυτό σημαίνει ότι η πιθανότητα αποτυχίας στην επόμενη χρονική στιγμή δεν εξαρτάται από την ηλικία του ατόμου.

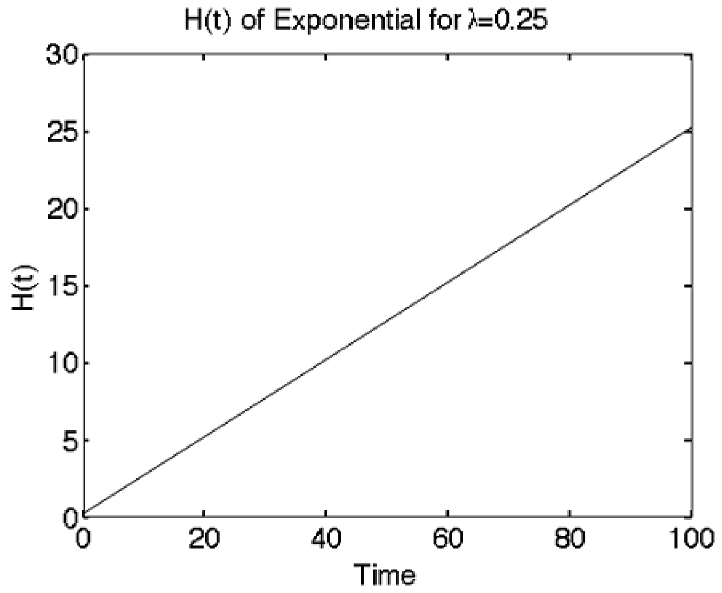
Κατά συνέπεια, η εκθετική κατανομή δεν μπορεί να μοντελοποιήσει την γήρανση.





Κατανομές ανάλυσης επιβίωσης

Εκθετική κατανομή



Παρατηρούμε ότι η συνάρτηση h(t) παραμένει σταθερή στον χρόνο ενώ η συνάρτηση H(t) αποτελεί μία γραμμή 45^o.

Κατανομές ανάλυσης επιβίωσης

- **Κατανομή Weibull**

Αποτελεί γενίκευση της εκθετικής κατανομής (για α=1 προκύπτει η εκθετική κατανομή, δηλαδή $Weibull(a=1,\lambda) \,..\, Exp(\lambda)$).

Η τ.μ. T ακολουθεί την **κατανομή Weibull (Weibull distribution)** με παραμέτρους α,λ > 0 ($T \sim Weibull(\alpha,\lambda)$, α=shape parameter, λ=scale parameter), t ≥ 0 και έχει σ.π.π. :

$$f(t) = h(t) \times S(t) = \alpha\lambda t^{a-1} e^{-\lambda t^{\alpha}}$$

συνάρτηση επιβίωσης : $S(t) = e^{-\lambda t^{a}}$, λ,t,α > 0

συνάρτηση κινδύνου : $h(t) = \alpha\lambda t^{a-1}$

αθροιστική συνάρτηση κινδύνου : $H(t) = \lambda t^{a}$

Κατανομές ανάλυσης επιβίωσης

Κατανομή Weibull

Η συνάρτηση κινδύνου είναι σταθερή εάν α = 1 (ταυτίζεται με την συνάρτηση κινδύνου του εκθετικού μοντέλου), αύξουσα ως προς t εάν α > 1 και φθίνουσα ως προς t εάν α < 1.

Είναι εμφανές ότι η Weibull είναι πιο ευέλικτη από την εκθετική.

Η σχέση $\ln(-\ln S(t))$ είναι γραμμική ως προς το $\ln(t)$. Αυτό μας επιτρέπει να αξιολογήσουμε γραφικά την καταλληλότητα του μοντέλου Weibull μέσω της γραφικής παράστασης της $\ln(-\ln \hat{S}_{KM}(t))$ ως προς $\ln(t)$ (όπου $\hat{S}_{KM}(t)$ η εκτίμηση της συνάρτησης επιβίωσης μέσω της μεθόδου Kaplan-Meier) και στην περίπτωση που προκύψει μία κατά προσέγγιση ευθεία γραμμή τότε αυτό αποτελεί ένδειξη ότι η κατανομή Weibull μπορεί να περιγράψει τα δεδομένα επαρκώς καλά.

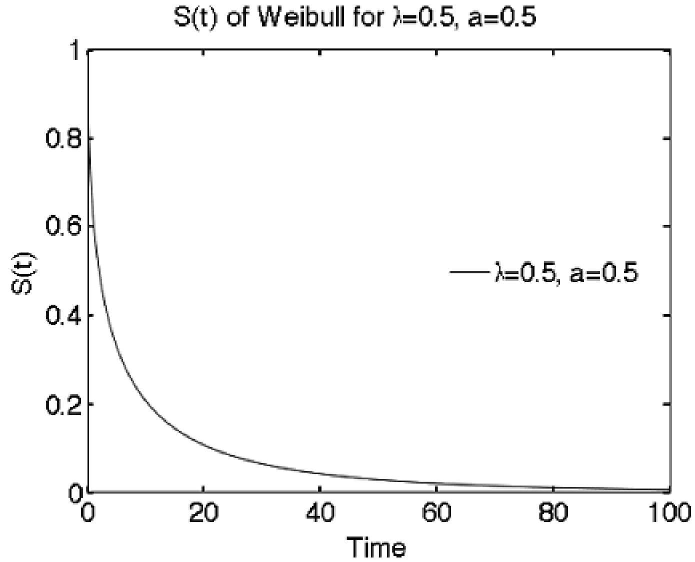
Για να επαληθεύσουμε τη γραμμική αυτή σχέση :

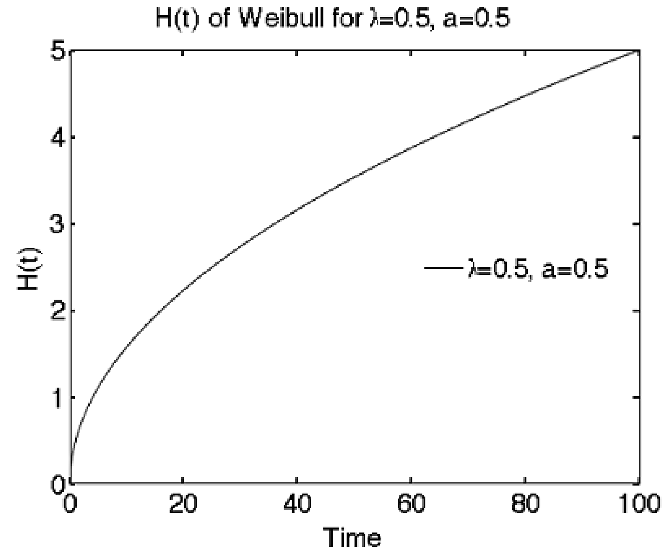
$$S(t) = e^{-\lambda t^{a}} \Rightarrow \ln\left[-\ln S(t)\right] = \ln(\lambda) + a\ln(t).$$

Στην κατανομή Weibull η $\ln[-\ln S(t)]$ είναι γραμμική συνάρτηση του $\ln(t)$ με κλίση α και σημείο τομής $\ln(\lambda)$. Εάν η κλίση είναι ίση με 1 τότε η T ακολουθεί μία εκθετική κατανομή.

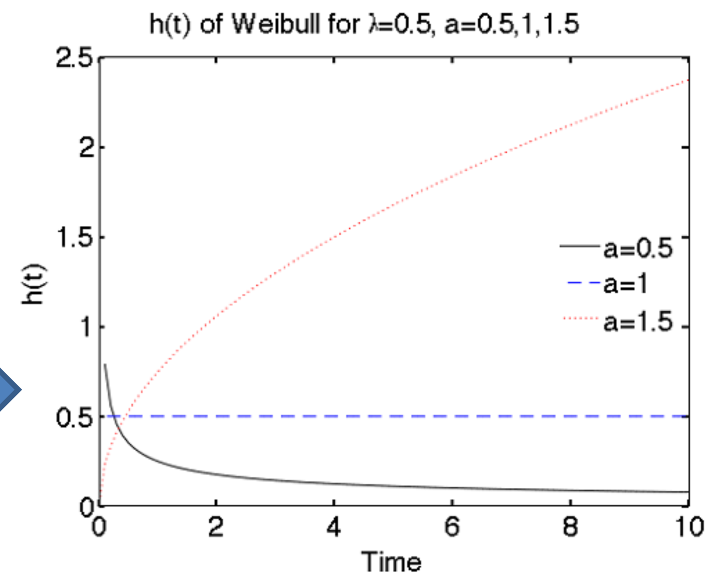
Κατανομές ανάλυσης επιβίωσης

Κατανομή Weibull





Στο τελευταίο γράφημα παρουσιάζεται η συνάρτηση h(t) για τιμές του α=0.5, 1, 1.5 από όπου μπορούμε να επιβεβαιώσουμε τα παραπάνω, δηλαδή ότι για α=1 έχουμε το εκθετικό μοντέλο, για α=0.5 η συνάρτηση είναι φθίνουσα και για α=1.5 η συνάρτηση είναι αύξουσα.



Κατανομές ανάλυσης επιβίωσης

- **Κατανομή Γάμμα**

Η τ.μ. T ακολουθεί την **κατανομή Γάμμα (Gamma distribution)** με παραμέτρους α,λ > 0

$(T \sim Gamma(\alpha,\lambda))$ και έχει σ.π.π. : $f(t) = \dfrac{\lambda^{\alpha}}{\Gamma(\alpha)} t^{\alpha-1} e^{-\lambda t}$, $t \geq 0$ **(2.9),**

όπου $\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x}\, dx$ η συνάρτηση γάμμα,

συνάρτηση επιβίωσης : $S(t) = 1 - I(\lambda t, \alpha)$

όπου $I(t,\alpha) = \dfrac{\int_0^t u^{\alpha-1} e^{-u}\, du}{\Gamma(\alpha)}$ η μη πλήρης συνάρτηση γάμμα.

Κατανομές ανάλυσης επιβίωσης

Κατανομή Γάμμα

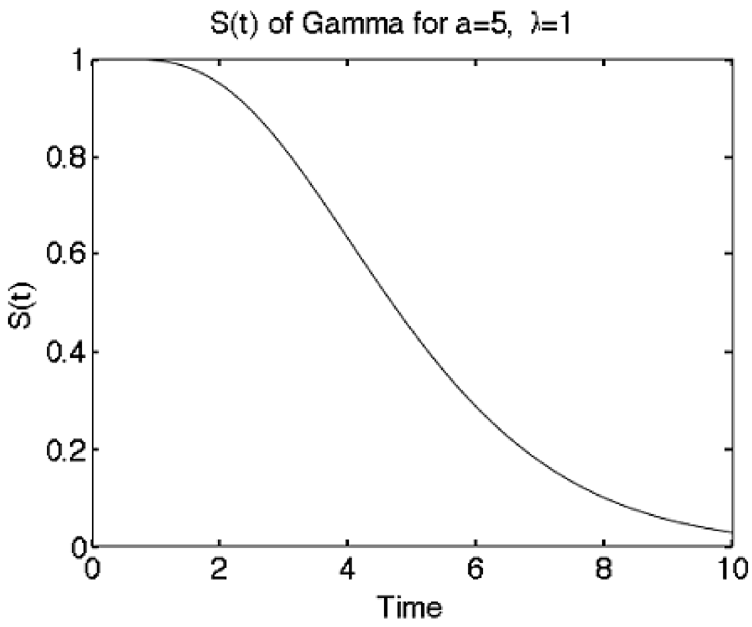
(Υπενθυμίζουμε ότι οι h(t), H(t) μπορούν σε κάθε περίπτωση να υπολογιστούν από τους τύπους : $h(t) = -\frac{d}{dt}\log S(t)$ **και** $H(t) = -\log S(t)$**).**

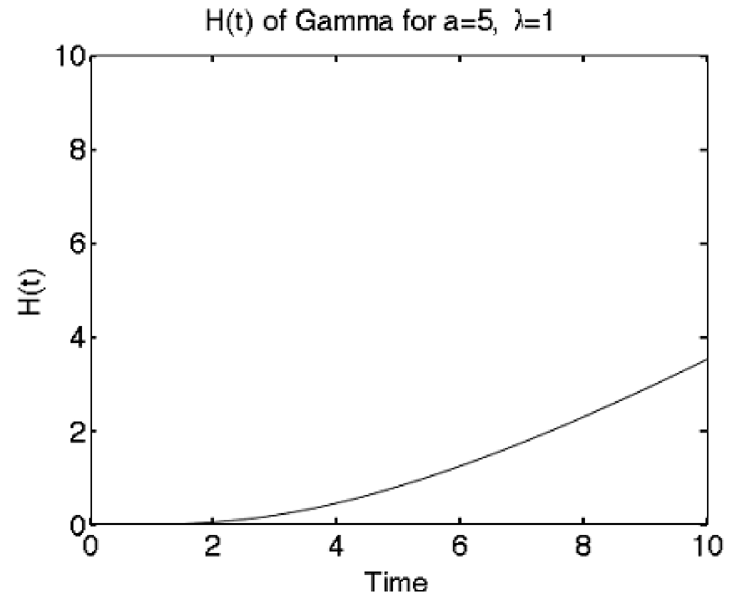
Η μέση τιμή και η διακύμανση δίνονται από τους τύπους : $E(T) = \frac{\alpha}{\lambda}, Var(T) = \frac{\alpha}{\lambda^2}$.

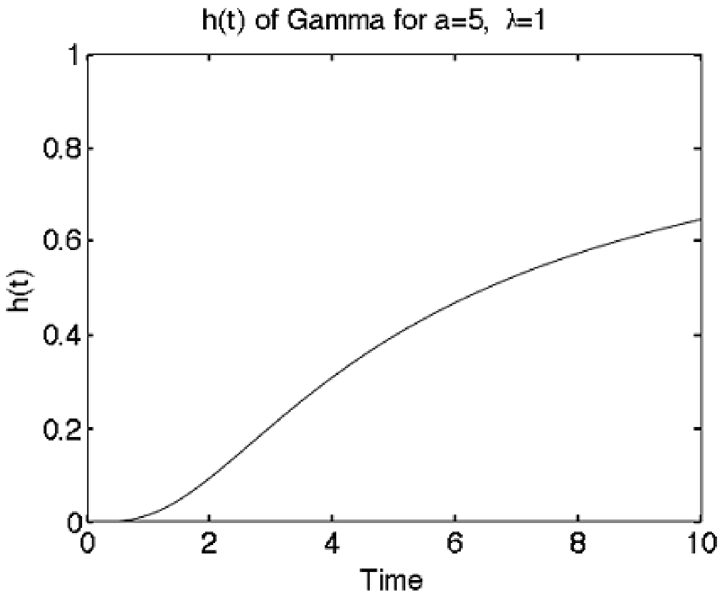
Όταν α=1 η κατανομή Γάμμα γίνεται η εκθετική κατανομή με παράμετρο λ, δηλαδή $Gamma(a = 1, \lambda) \ldots Exp(\lambda)$.

Κατανομές ανάλυσης επιβίωσης

Κατανομή Γάμμα







Η συνάρτηση κινδύνου

- είναι αύξουσα για $\alpha > 1$
- είναι φθίνουσα για $0 < \alpha < 1$

Κατανομές ανάλυσης επιβίωσης

Λογαριθμολογιστική κατανομή

Η τ.μ. Τ ακολουθεί την **λογαριθμολογιστική κατανομή (Log-logistic distribution)** με παραμέτρους α,λ > 0, t ≥ 0 ($T \quad LL(a,\lambda)$) και έχει σ.π.π. : $f(t) = \dfrac{\alpha t^{a-1}\lambda}{(1+\lambda t^{a})^{2}}$

συνάρτηση επιβίωσης : $S(t) = \dfrac{1}{1+\lambda t^{a}}$

συνάρτηση κινδύνου : $h(t) = \dfrac{a t^{\alpha-1}\lambda}{1+\lambda t^{a}}$

αθροιστική συνάρτηση κινδύνου : $H(t) = \log(1+\lambda t^{\alpha})$

Κατανομές ανάλυσης επιβίωσης

Λογαριθμολογιστική κατανομή

Η λογαριθμολογιστική κατανομή είναι ιδιαιτέρως ευέλικτη όσον αφορά την συνάρτηση κινδύνου της :

- Εάν $a<1$ η $h(t)$ είναι φθίνουσα από το $\times$ και κάτω.
- Εάν $a=1$ η $h(t)$ είναι φθίνουσα από το λ και κάτω.
- Εάν $a>1$ η $h(t)$ αυξάνει και έπειτα μειώνεται.

Ονομάζεται λογαριθμολογιστική κατανομή διότι η τ.μ. logT ακολουθεί την λογιστική κατανομή (συμμετρική κατανομή).

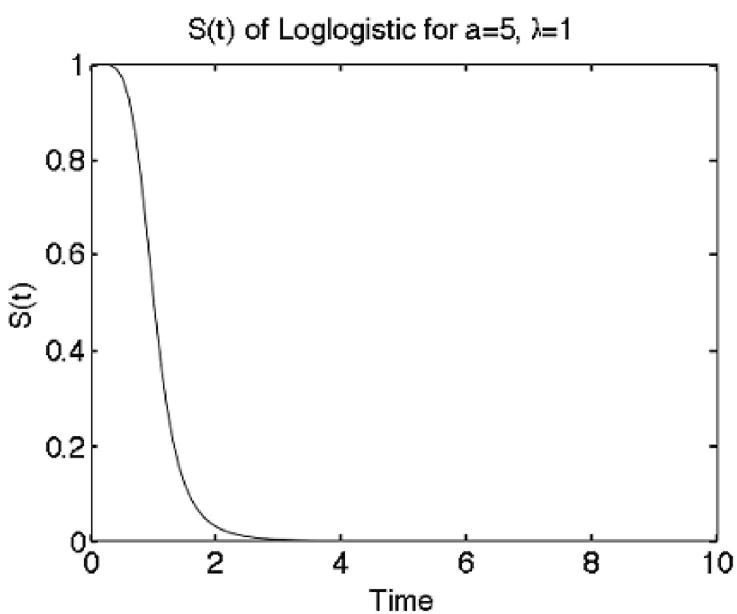
Κατανομές ανάλυσης επιβίωσης

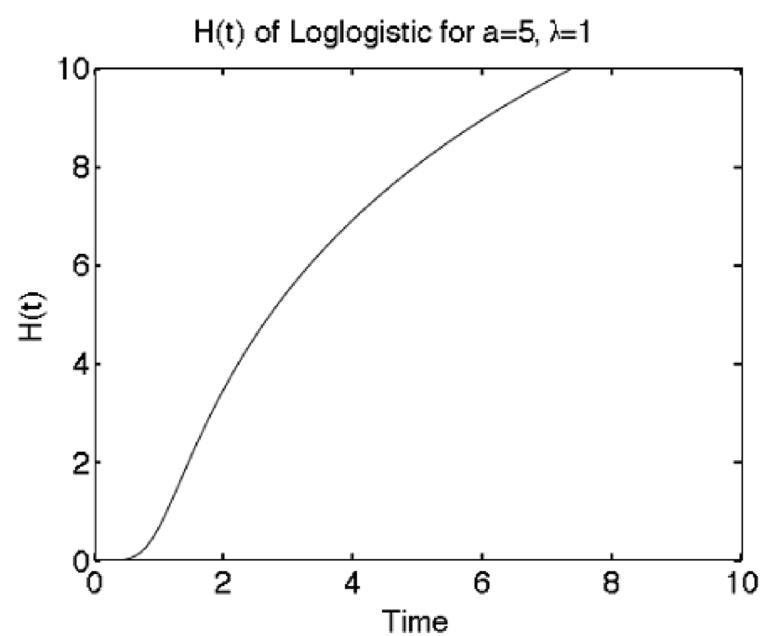
Λογαριθμολογιστική κατανομή

Σημειώνεται ότι για την $S(t)$, $odds = \dfrac{S(t)}{1 - S(t)} = \dfrac{1}{\lambda t^{a}}$

και $\mathrm{logit}(S(t)) = \log(odds) = -\log\lambda - a\log t \Leftrightarrow -\log(odds) = \log\lambda + a\log t$

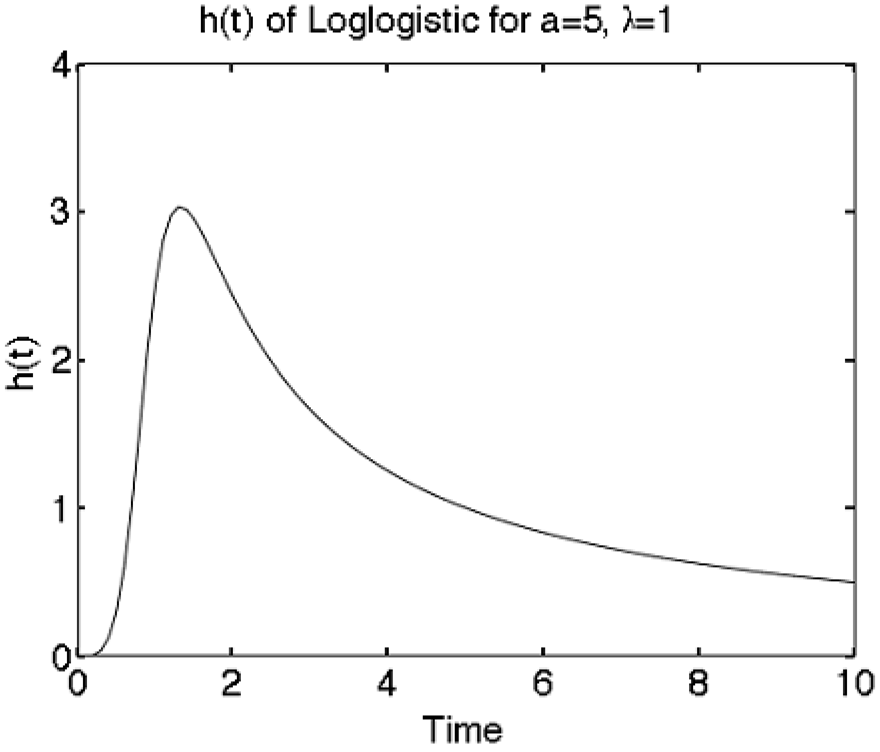
(όπου **odds (ή σχετική πιθανότητα)** ο λόγος της πιθανότητας εμφάνισης ενός ενδεχομένου έναντι της πιθανότητας μη εμφάνισής του και **logit** η συνάρτηση του λογαρίθμου του odds).





Κατανομές ανάλυσης επιβίωσης

Λογαριθμολογιστική κατανομή



Κατανομές ανάλυσης επιβίωσης

Κατανομή Pareto

Η τ.μ. Τ ακολουθεί την **κατανομή Pareto** με παραμέτρους λ>0, θ>0 ($T \sim Pareto(\theta,\lambda)$) και έχει σ.π.π. :

$$f(t) = \theta \frac{\lambda^{\theta}}{t^{\theta+1}},\ t \geq \lambda$$

συνάρτηση επιβίωσης : $S(t) = \dfrac{\lambda^{\theta}}{t^{\theta}}$

συνάρτηση κινδύνου : $h(t) = \theta \dfrac{1}{t}$

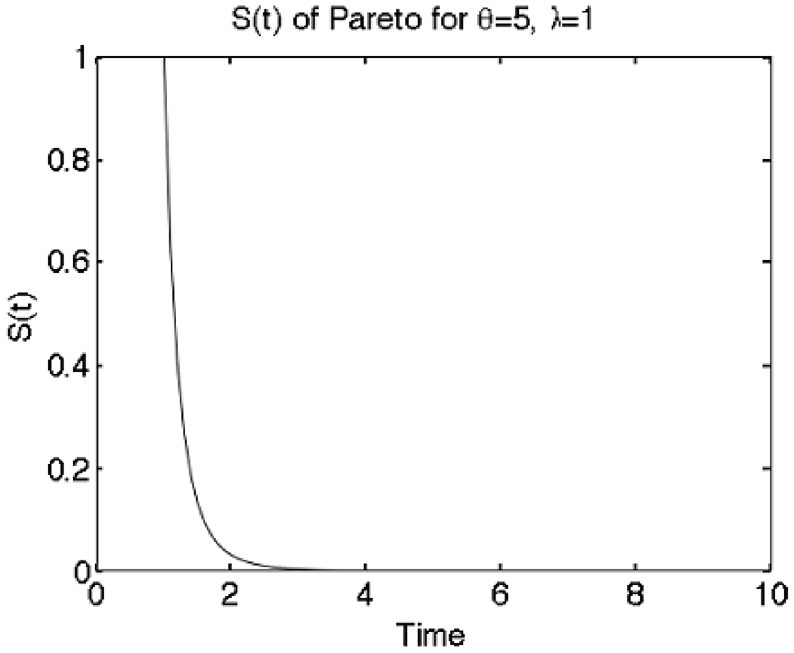
Η μέση τιμή και η διακύμανση δίνονται από τους τύπους :

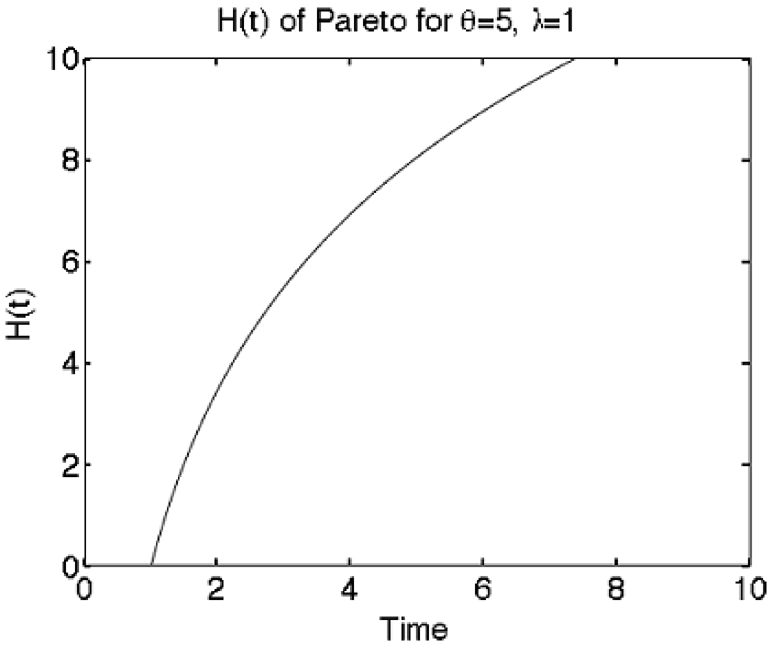
$$E(T) = \frac{\theta\lambda}{\theta-1},\ \theta > 1 \text{ και } Var(T) = \frac{\theta\lambda^2}{(\theta-1)^2(\theta-2)},\ \theta > 2.$$

Κατανομές ανάλυσης επιβίωσης

Κατανομή Pareto

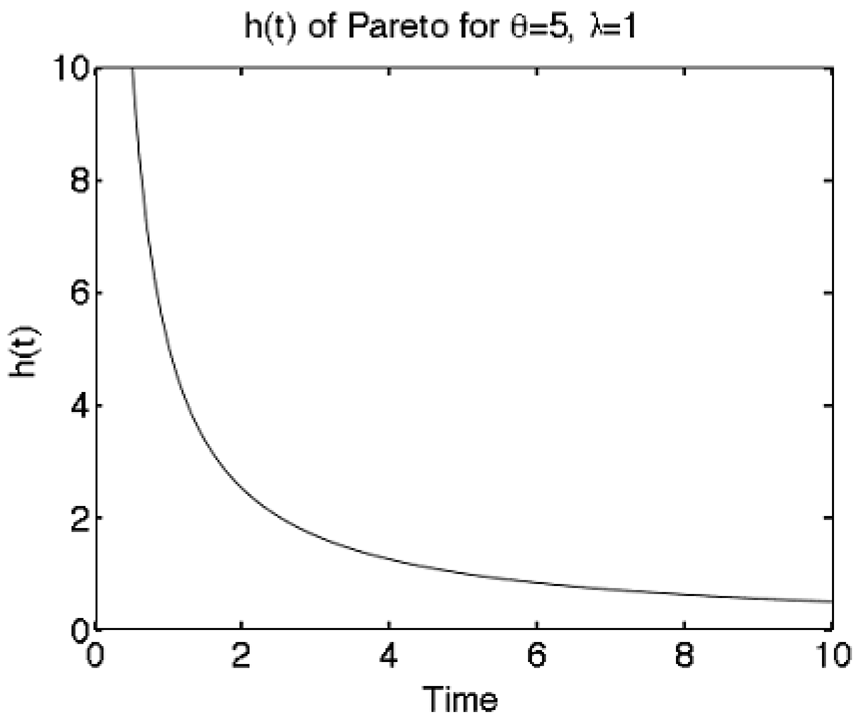
Εάν $T \quad Pareto(\theta, \lambda)$ τότε $\log\left(\frac{T}{\theta}\right) \sim Exp(\lambda)$





Κατανομές ανάλυσης επιβίωσης

Κατανομή Pareto



Κατανομές ανάλυσης επιβίωσης

Κατανομή Gompertz

Η τ.μ. Τ ακολουθεί την **κατανομή Gompertz** με παραμέτρους λ>0, α>0

$(T \sim Gompertz(\alpha, \lambda))$ και έχει σ.π.π. :

$$f(t) = \lambda e^{at} \exp\left(\frac{\lambda}{a}\left(1 - e^{at}\right)\right), t \geq 0$$

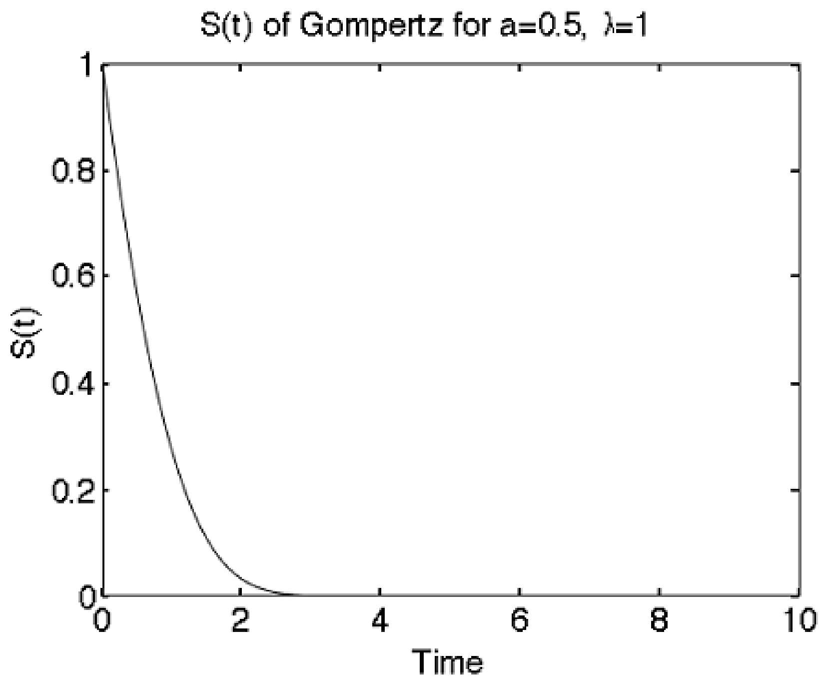
συνάρτηση επιβίωσης : $S(t) = \exp\left(\frac{\lambda}{a}\left(1 - e^{at}\right)\right), t \geq 0$

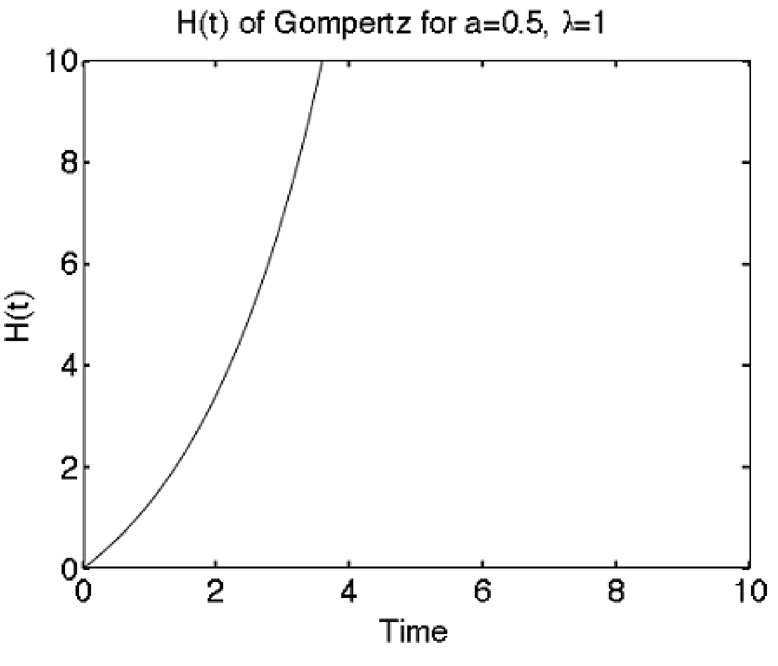
συνάρτηση κινδύνου : $h(t) = \lambda e^{at}$

Κατανομές ανάλυσης επιβίωσης

Κατανομή Gompertz

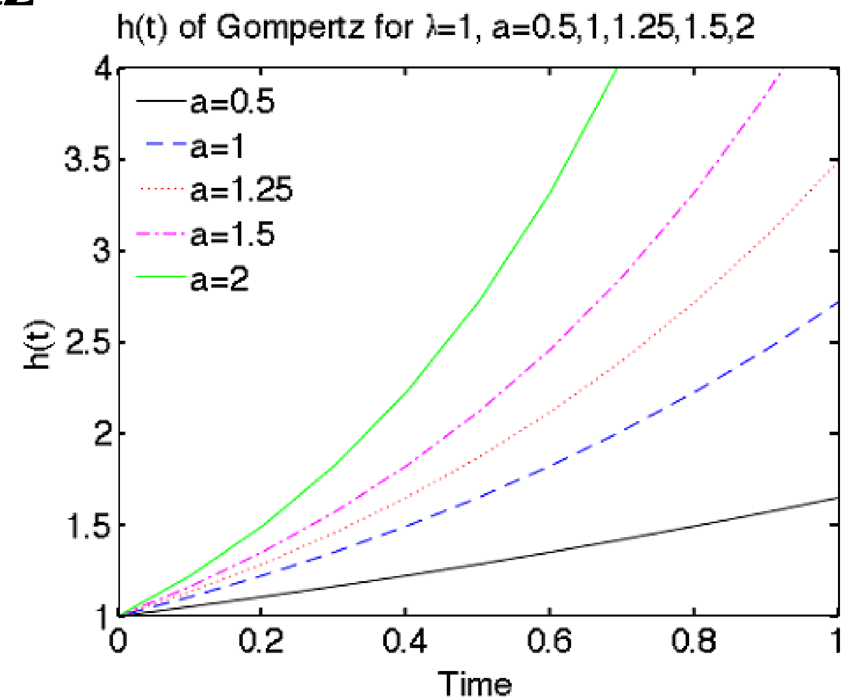
Η κατανομή Gompertz αποτελεί έναν βολικό τρόπο περιγραφής της επιβίωσης ανθρώπων και χρησιμοποιείται συχνά στην δημογραφία. Μπορεί να γενικευθεί στην Gompertz-Makeham κατανομή προσθέτοντας μία σταθερά στην συνάρτηση κινδύνου : $h(t) = c + \lambda e^{at}$.





Κατανομές ανάλυσης επιβίωσης

Κατανομή Gompertz



Παρατηρούμε ότι η συνάρτηση κινδύνου της Gompertz κατανομής είναι μία αυξανόμενη συνάρτηση όσο αυξάνουμε το α.

Κατανομές ανάλυσης επιβίωσης

Λογαριθμοκανονική κατανομή

Η **λογαριθμοκανονική κατανομή (Log-normal distribution)** στην πιο απλή μορφή της μπορεί να οριστεί ως η κατανομή μίας τ.μ. $\log T \sim Normal$.

Ο χρόνος επιβίωσης T ακολουθεί την λογαριθμοκανονική κατανομή και συμβολίζεται ως $T \quad LN(\mu,\sigma^2)$ όταν η τ.μ. $\log T$ ακολουθεί την $Normal(\mu,\sigma^2)$ και έχει σ.π.π. :

$$f_T(t) = \frac{1}{t\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}\left(\ln t - \mu\right)^2\right],\ t > 0,\ \mu \geq 0,\ \sigma > 0$$

συνάρτηση επιβίωσης $$S(t) = \frac{1}{\sigma\sqrt{2\pi}} \int_t^{\infty} \frac{1}{x} \exp\left[-\frac{1}{2\sigma^2}\left(\ln x - \mu^2\right)\right] dx$$

συνάρτηση κινδύνου $$h_T(t) = \frac{\dfrac{1}{t\sigma\sqrt{2\pi}} \exp\left[-\dfrac{(\ln at)^2}{2\sigma^2}\right]}{1 - \Phi\left(\ln \dfrac{at}{\sigma}\right)}, \quad \text{για } t > 0,$$

Κατανομές ανάλυσης επιβίωσης

Λογαριθμοκανονική κατανομή

όπου *Φ(y)* είναι η αθροιστική συνάρτηση της τυποποιημένης κανονικής κατανομής, δηλαδή $\Phi(y) = \frac{1}{\sqrt{2\pi}} \int_0^y \exp(-u^2/2)du$.

Η τιμή της παραμέρου σ^2 ελέγχει την ασυμμετρία της κατανομής: όσο μεγαλύτερη η τιμή του σ^2, τόσο μεγαλύτερη ασυμμετρία.

Η συνάρτηση κινδύνου αυξάνει αρχικά σε κάποιο μέγιστο και έπειτα μηδενίζεται καθώς ο χρόνος t τείνει στο άπειρο.

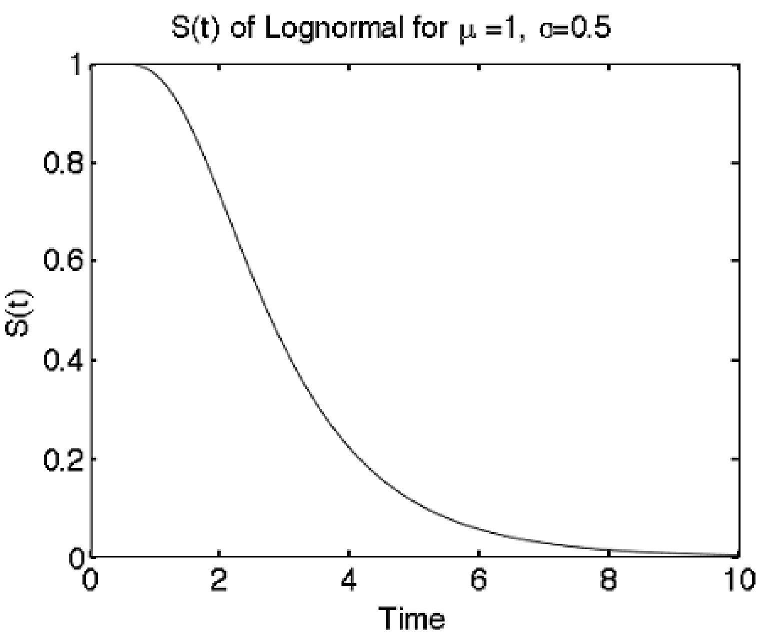
Εάν $T \sim \Lambda(\mu, \sigma^2)$ τότε η μέση τιμή της κατανομής είναι $E(T) = \exp\left(\mu + \frac{\sigma^2}{2}\right)$

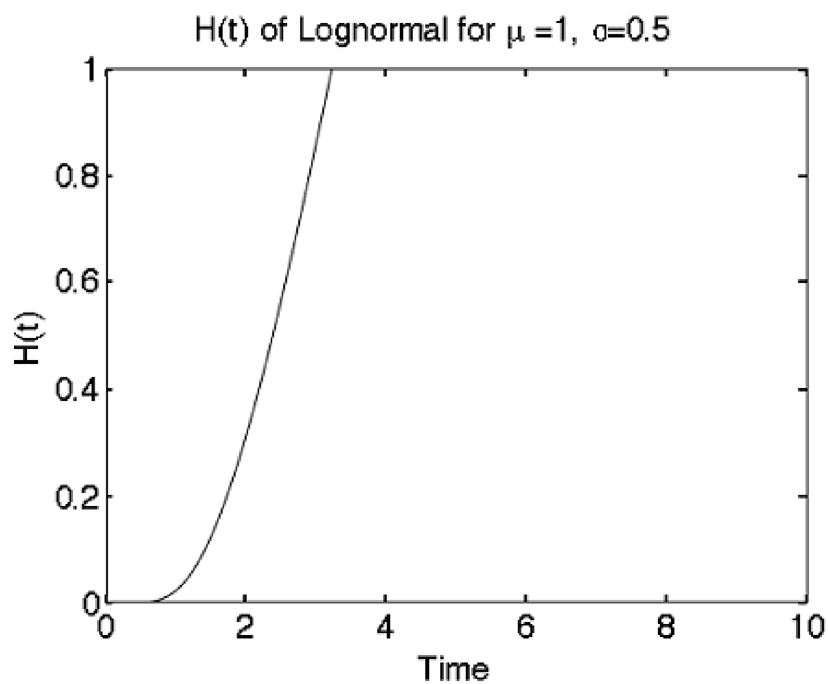
και η διακύμανση $Var(T) = \exp(\sigma^2 - 1)\exp(2\mu + \sigma^2)$.

Κατανομές ανάλυσης επιβίωσης

Λογαριθμοκανονική κατανομή

Η συνάρτηση κινδύνου αυξάνεται αρχικά σε κάποιο μέγιστο και έπειτα μηδενίζεται καθώς $t \to \infty$. Χαρακτηριστικό παράδειγμα της ιδιαίτερης συνάρτησης κινδύνου του λογαριθμοκανονικού μοντέλου είναι η περίπτωση της φυματίωσης όπου ο κίνδυνος να αποβιώσει ένας ασθενής αυξάνεται αρχικά και μειώνεται αργότερα όσο η ασθένεια προχωράει.





Κατανομές ανάλυσης επιβίωσης

Λογαριθμοκανονική κατανομή