Ereignisanalyse

Marcel Noack

29. April 2011

Einleitung

Der Begriff “Ereignisanalyse” bezeichnet eine Reihe statistischer Verfahren, die zur Untersuchung von Zeitintervallen zwischen aufeinander folgenden Ereignissen oder Zustandswechseln verwendet werden.

Anwendungsfelder

- Zeitdauer bis zu einem Regierungswechsel in Land x
- Wechsel der Parteipräferenz bei Person y.
- Überlebenszeiten von Patienten in medizinischen Studien, beispielsweise nach Herzoperationen oder Chemotherapie,
- Dauer von Lernprozessen in der Psychologie,
- Zeitspanne bis zu einem transregionalen Umzug in der räumlichen Mobilitätsanlyse,
- Dauer der “Herrschaft” eines Löwen über sein Rudel in der Biologie
- Dauer von Arbeitslosigkeit in ökonomischen Untersuchungen

Einleitung

Die Statistik bietet heute eine grosse Anzahl an Möglichkeiten zur Analyse von Ereignisdaten. Sie umfassen:

- Deskriptive Verfahren: Sterbetafel-Methode oder Kaplan-Meier-Schätzung
- Semiparametrisches Regressionsmodell von Cox
- Parametrische Verfahren mit und ohne Zeitabhängigkeiten: Exponentialmodell, Piecewise-Constant-Modell, Gompertz-Makeham-Modell, Weibull-Modell, Log-Logistisches-Modell

Begriffe

State: Zustand der Untersuchungseinheit. Das Set der möglichen Zustände wird Zustandsraum ("state space") genannt.

Event: Wandel von Ursprungszustand (origin state) in einen definierten Zielzustand (destination state).

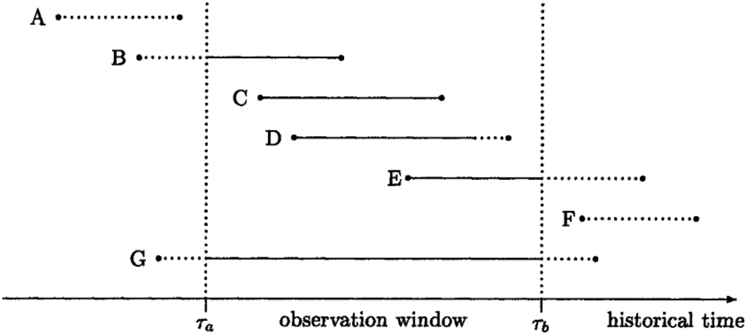
Duration: Die Verweildauer gibt an, wie lange ein Individuum in einem Zustand verharrt

Risk Period: Periode, in der ein Individuum dem Risiko ausgesetzt ist, einen Zustandswechsel zu vollziehen.

Risk Set: Anzahl der Fälle, die an einem bestimmen Zeitpunkt (stetig) oder in einem bestimmten Intervall (diskret) dem Risiko eines Zustandswechsels unterliegen. "Noch lebendig".

Zensierung

Zensierung: Dauer der Episode liegt nur inkomplett vor.



Blossfeld, H.-P. und Rohwer, G. (2002): Techniques of Event History Modeling. Mahwah, 2. Aufl. S. 40.

Dichtefunktion $f(t)$ und Verteilungsfunktion $F(t)$

- Wenn es sich bei T um eine stetige Zufallsvariable handelt, kann die Verteilung von T als Dichtefunktion $f(t)$ beschrieben werden. Sie steht mit der Verteilungsfunktion $F(t)$ in folgendem Zusammenhang:

$$f(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t)}{\Delta t} = \frac{\partial F(t)}{\partial t} \qquad (1)$$

- Es gilt:

$$F(t) = P(T \leq t) = \int_0^t f(u) du \quad \text{mit} \quad f(t) = F'(t) \qquad (2)$$

Dichtefunktion $f(t)$

$f(t)$ gibt die Wahrscheinlichkeit an, dass ein Ereignis zum Zeitpunkt t eintritt.

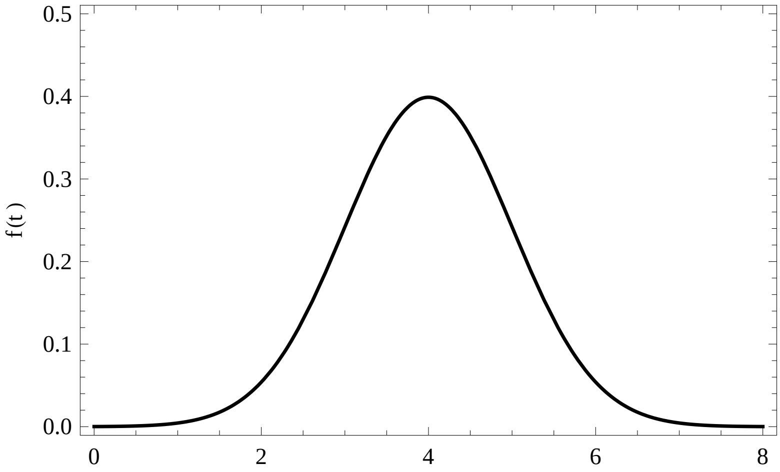


Abbildung: f(t), PDF

Verteilungsfunktion $F(t)$

$F(t)$ gibt die Wahrscheinlichkeit an, dass ein Ereignis bis zum Zeitpunkt t eintritt.

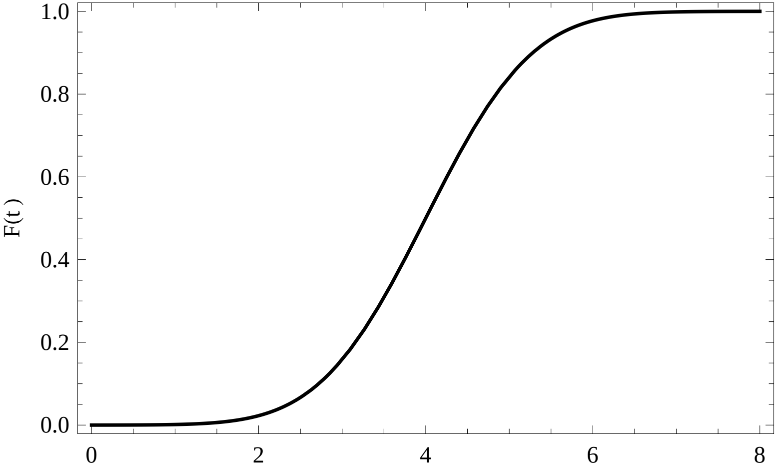


Abbildung: F(t), CDF

Survivalfunktion $S(t)$

- Die *Survivalfunktion* gibt die Wahrscheinlichkeit dafür an, das vor dem Zeitpunkt t *kein* Ereigniss eintritt.
- Individuen, denen das Ereignis noch nicht wiederfahren ist haben "überlebt" (survived).
- Bei $S(t)$ handelt es sich um eine fallende Funktion von t, mit $S(0) = 1$ und $S(t) = 0$ für $t \to \infty$.
- Die Analyse startet mit 100% "Überlebenden", bei denen sich nach unendlich langer Zeit ($t \to \infty$) bei jedem Individuum ein Zustandswechsel vom Urzustand in den Zielzustand vollzogen hat.
- Sie ist definiert als:

$$S(t) = 1 - F(t) = 1 - P(T \leq t) = P(T \geq t) = \int_t^\infty f(u)du \tag{3}$$

Dichtefunktion $f(t)$

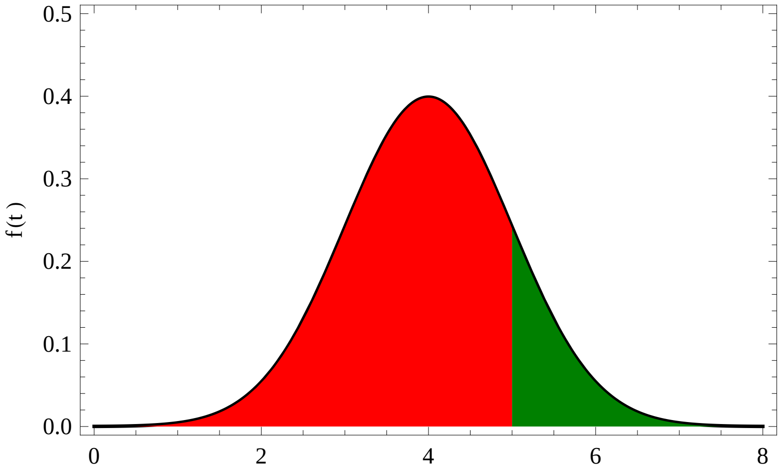


Abbildung: t=5

Verteilungsfunktion $F(t)$

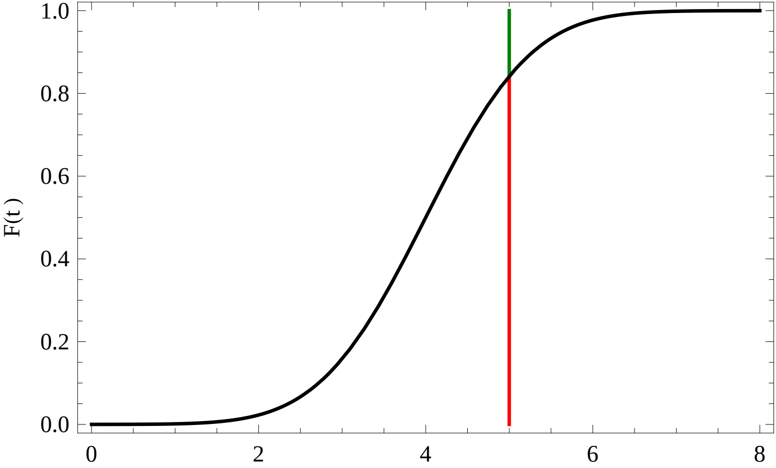


Abbildung: t=5

Survivalfunktion $S(t)$

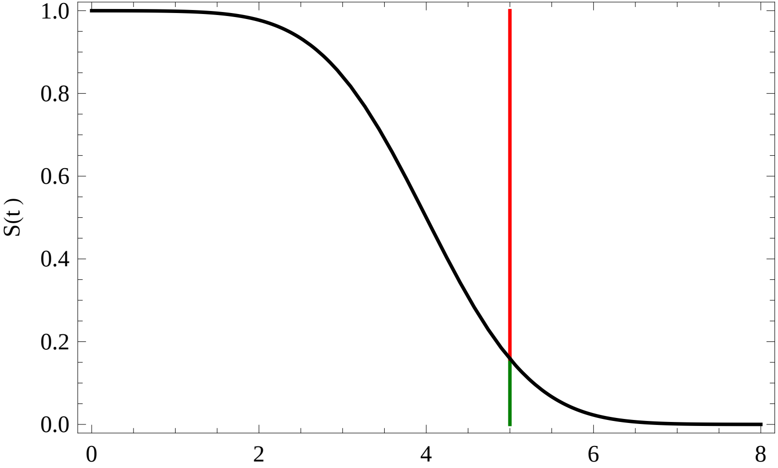


Abbildung: t=5

Verteilungsfunktion $F(t)$ und Survivalfunktion $S(t)$

- Die *Survivalfunktion* ist also das Komplement der *Verteilungsfunktion*. Sie gibt die Wahrscheinlichkeit dafür an, dass ein Ereignis vor dem Zeitpunkt t statt findet.
- Folgender Zusammenhang besteht zwischen *Survivalfunktion* und *Verteilungsfunktion*:

$$F(t) + S(t) = 1 \tag{4}$$

$$P(T \leq t) + P(T \geq t) = P(\Omega) = 1 \tag{5}$$

$$\int_0^t f(u)du + \int_t^\infty f(u)du = \int_0^\infty f(u)du = 1 \tag{6}$$

Hazardrate $h(t)$

Der “hazard”

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} = \frac{f(t)}{S(t)} \tag{7}$$

gibt die Wahrscheinlichkeit an, dass das Ereigniss in einem sehr kurzen zeitlichen Intervall - unter der Bedingung, dass das Ereignis nicht schon vorher eingetreten ist - statt findet. Aus diesem Grund ist die *hazard rate* auch als “instantaneous risk” bekannt.