
Zeit-bis-zum-Ereignis Daten III

**(Maximum Likelihood Schätzung und ihre
Anwendung auf Time to event Daten)**

Dr. Dirk Hasenclever

IMISE, Universität Leipzig

Dirk.Hasenclever@IMISE.uni-Leipzig.de

Maximum Likelihood Methode zur Herleitung von Schätzverfahren I

Seien $X_1, \dots, X_N$ eine mathematische Stichprobe unabhängige und identisch verteilter diskreter Zufallsvariable mit einer gemeinsamen Verteilung $F_X(x, \theta_X)$, welche in der parametrischen Familie von Verteilungsfunktionen $F_X(x, \theta)$ mit $\theta \in \Omega$ und Wahrscheinlichkeitsfunktion $\text{pr}_X(X = x, \theta)$ liegt.

Für eine konkrete (realisierte) Stichprobe $x_1, \dots, x_N$ ist die Likelihoodfunktion

$$L(\theta \mid x_1, \dots, x_N) := \prod_{i=1}^N \text{pr}_X(X = x_i, \theta)$$

eine zufällige (da von der Stichprobe abhängig) Funktion von θ .

Ihr Wert ist die Wahrscheinlichkeit der Daten gegeben den Parameter θ .

Maximum Likelihood Methode zur Herleitung von Schätzverfahren II

Der - gegeben die beobachteten Daten - plausibelste Wert von θ ist

$$\theta_{ML} = \operatorname{argmax}_{\theta} (L(\theta | x_1, \dots, x_N))$$

d.h. derjenige Wert θ_{ML} , der die Wahrscheinlichkeit der Beobachtung $L(\theta | x_1, \dots, x_N)$ maximiert
also die Daten "am wahrscheinlichsten" erscheinen lässt.

Zur Bestimmung des Maximums der Funktion $L(\theta | x_1, \dots, x_N)$ kann man die üblichen Methoden der Analysis bzw. der numerischen Mathematik einsetzen.

Das Verfahren, aus einer Stichprobe θ_{ML} zu berechnen, heißt

Maximum Likelihood Schätzverfahren.

Beispiel: Maximum Likelihood Schätzer für Wahrscheinlichkeit in einem Binomialen Versuch

Binomialer Versuch : k Erfolge aus N :

$$L(p) = \binom{N}{k} p^k (1-p)^{(N-k)}$$

Leichter ist es mit dem Logarithmus der Likelihood zu arbeiten :

$$\begin{aligned} l(p) = \ln(L(p)) &= \ln\left(\binom{N}{k}\right) + k \cdot \ln p + (N-k) \cdot \ln(1-p) \\ &\propto k \cdot \ln p + (N-k) \cdot \ln(1-p) \quad \text{[nur Terme mit p relevant]} \end{aligned}$$

Setze erste Ableitung Null :

$$\begin{aligned} \frac{dl}{dp} &= \frac{k}{p} - \frac{N-k}{1-p} = 0 \Leftrightarrow \\ k - kp &= Np - kp \Leftrightarrow \\ p &= \frac{k}{N} \end{aligned}$$

Der Häufigkeitsschätzer ist auch der Maximum Likelihood Schätzer!

Maximum Likelihood Methode zur Herleitung von Schätzverfahren III

Für stetige Zufallsvariable ist $pr_X(X = x) = 0$ für alle x des Trägers von X .

Um eine Likelihoodfunktion auch für stetige Zufallsvariable zu definieren, betrachten wir stattdessen die Wahrscheinlichkeit, das X in ein Intervall der Breite Δx um x fällt:

$$pr_X(X \in [x - \Delta x / 2 ; x + \Delta x / 2]) \approx f_X(x) \cdot \Delta x$$

Da Δx^N nicht vom Parameter θ abhängt, können wir analog zum diskreten Fall definieren:

Für eine konkrete (realisierte) Stichprobe $x_1, \dots, x_N$ einer stetigen Variable ist die Likelihoodfunktion

$$L(\theta | x_1, \dots, x_N) := \prod_{i=1}^N f_X(x_i, \theta)$$

ML-Schätzer für die Parameter der Normalverteilung

$$L(\mu, \sigma^2 \mid x_1, \dots, x_N) = \prod_{i=1}^N \frac{1}{\sqrt{2\pi \cdot \sigma^2}} \exp\left(-\frac{(x_i - \mu)^2}{2 \cdot \sigma^2}\right)$$

Loglikelihood:

$$l(\mu, \sigma^2 \mid x_1, \dots, x_N) = \sum_{i=1}^N \left(-\ln(\sqrt{2\pi}) - \ln(\sigma) - \frac{(x_i - \mu)^2}{2 \cdot \sigma^2} \right)$$

$$\frac{\partial l}{\partial \mu} = \sum_{i=1}^N \left(\frac{(x_i - \mu)}{\sigma^2} \right) = 0 \Leftrightarrow \sum_{i=1}^N x_i = N\mu \Rightarrow \hat{\mu}_{ML} = \frac{1}{N} \sum_{i=1}^N x_i$$

$$\frac{\partial l}{\partial \sigma^2} = -\frac{N}{\sigma} + \sum_{i=1}^N \left(\frac{(x_i - \mu)^2}{\sigma^3} \right) = 0 \Leftrightarrow N = \sum_{i=1}^N \left(\frac{(x_i - \mu)^2}{\sigma^2} \right)$$

$$\Rightarrow \hat{\sigma}_{ML}^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \hat{\mu}_{ML})^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2$$

Allgemeiner Satz über Maximum Likelihood Schätzer

Maximum Likelihood Schätzer (MLE) sind i.A. asymptotisch multivariat normalverteilt:

$$\hat{\beta} \sim \text{Normal}[\beta, I^{-1}(\beta)]$$

$$I(\beta) = E(J(\beta)) \quad \text{mit} \quad J(\beta) = -\left(\frac{\partial^2 l(\beta)}{\partial \beta_i \partial \beta_j}\right)_{i,j} \quad (\text{I ist die Loglikelihood})$$

$I(\beta)$ kann man schätzen durch:

$$I(\beta) \approx I_{\text{obs}} = -\left(\frac{\partial^2 l(\beta)}{\partial \beta_i \partial \beta_j}\right)_{i,j} \Big|_{\hat{\beta}} \quad (\text{beobachtete Informationsmatrix})$$

Dieser Satz gilt nur unter gewissen Regularitätsbedingungen, die meistens erfüllt sind.
Vorsicht, wenn der Träger der Verteilungsfamilie von deren Parametern abhängt!

Heuristik dazu

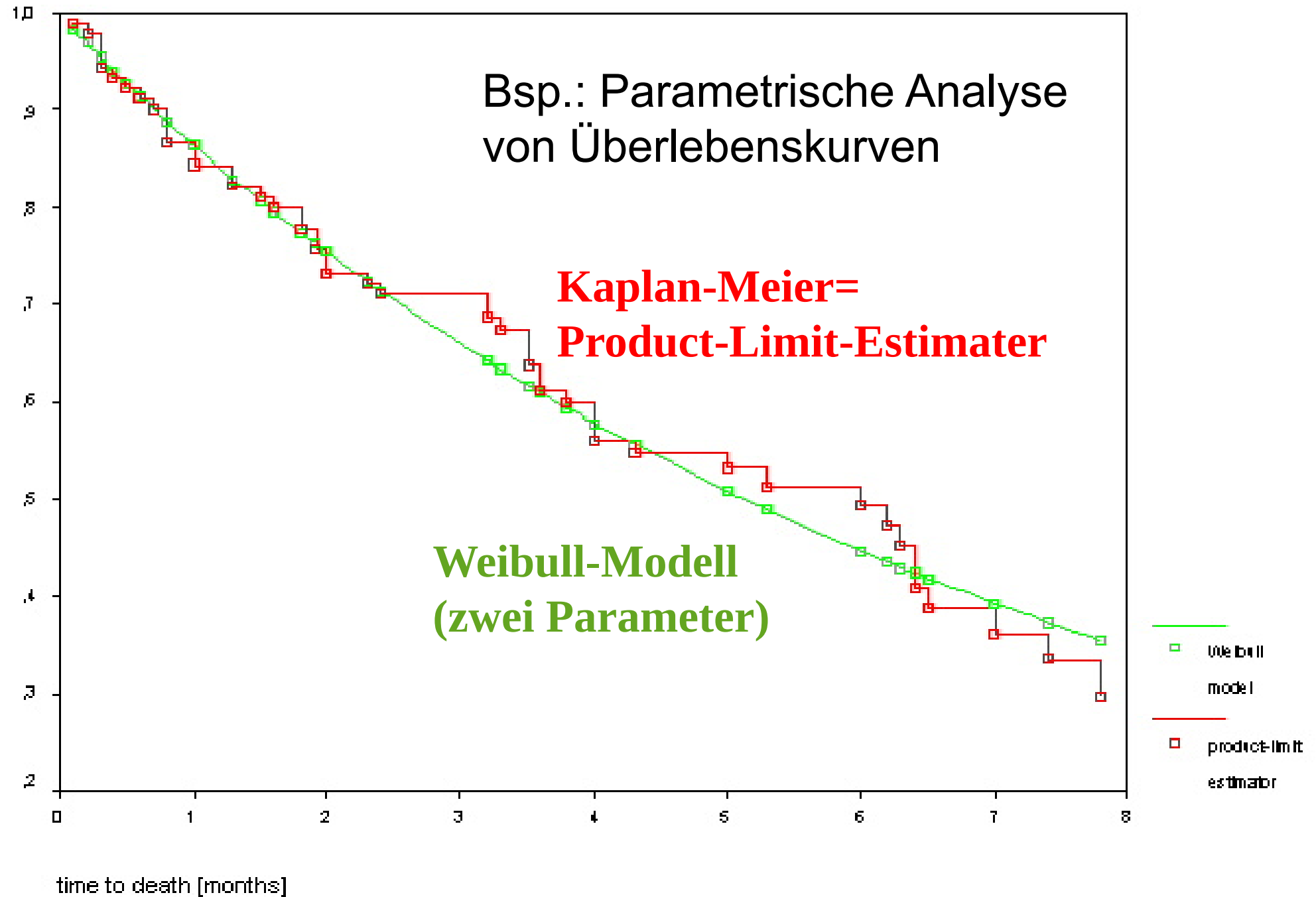
- Taylor-Entwicklung im Maximum β :
- $\text{Loglikelihood}(\beta) \approx$
 $\text{Matrix_zweite_Ableitungen}(\beta) * (\hat{\beta} - \beta)^2$
(in kleiner Umgebung)
- Je gekrümmter die Loglikelihood desto genauer ist $\hat{\beta}$ durch die Daten bestimmt.
- Also Standardfehler² $\sim$
 $1 / \text{Matrix_zweite_Ableitungen}(\beta)$

Parametrische Modelle für Überlebenskurven

Parametrische Modelle mit wenigen Parametern wie z.B. eine Weibull-Verteilung kann man an Überlebensdaten durch Maximum Likelihood Schätzung der Parameter anpassen.

Dazu muss man zensierte Daten geeignet in der Likelihood Funktion berücksichtigen.

Larynx data



Likelihoodfunktion für zensierte Daten I

Daten : (t_i, δ_i) $i = 1, \dots, n$ $\delta_i = \begin{cases} 1 & \text{Ereignis beobachtet} \\ 0 & \text{Beobachtung zensiert} \end{cases}$

IDEE : Modelliere den Zensurprozess mit
durch eine bivariate, teilweise latente Verteilung :

$(T, C) \sim \text{Dichte } h(t, c) dt dc$ (zweidimensional)

$T := \text{'wahrer'}$, eventuell unbeobachtbarer Ereigniszeitpunkt

$C := \text{'wahrer'}$, eventuell unbeobachtbarer Zensurzeitpunkt

Also : $\delta_i = \begin{cases} 1 & t_i \leq c_i \quad (c_i \text{ unbeobachtet}) \\ 0 & t_i > c_i \quad (t_i \text{ unbeobachtet}) \end{cases}$

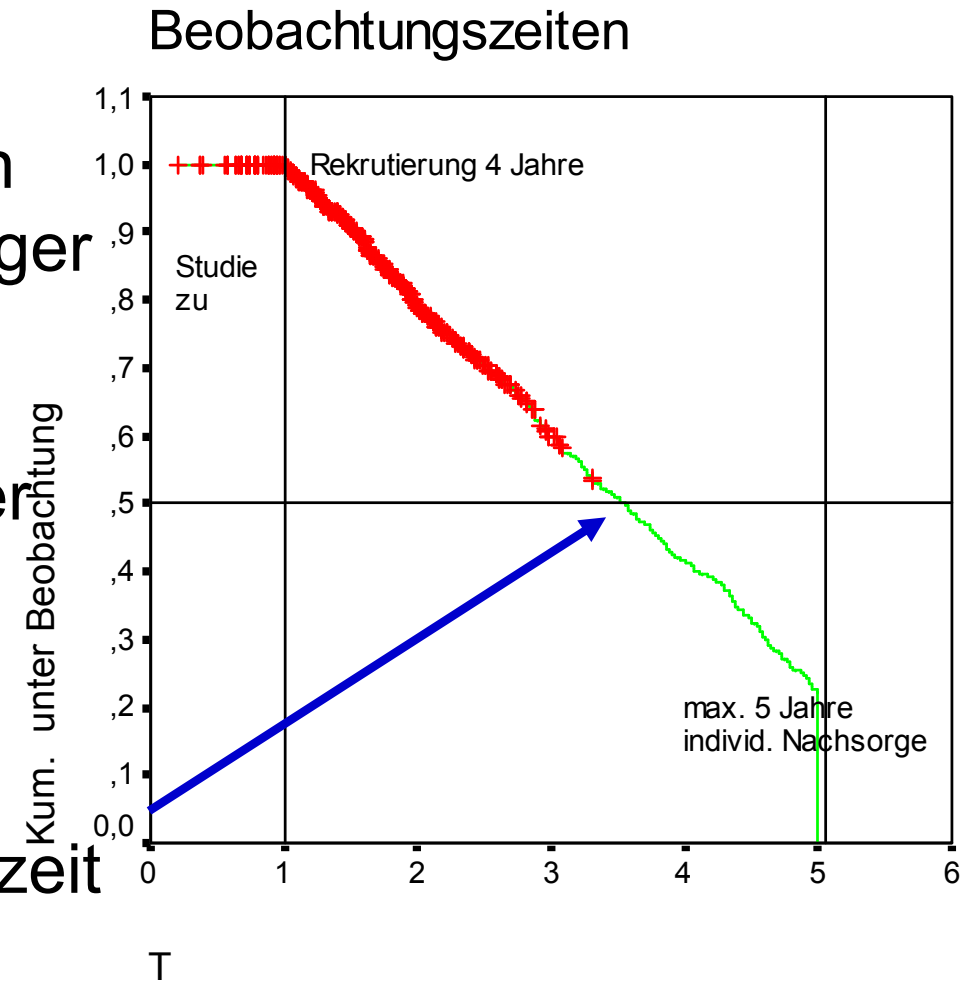
Für die Likelihood zu bestimmen sind :

$\text{prDichte}(t, 0)$ und $\text{prDichte}(t, 1)$

Verteilung von Beobachtungszeiten

Verteilung Beobachtungszeiten einer seit 1 Jahr geschlossenen Studie mit 4 Jahren gleichmäßiger Rekrutierung und individuell auf 5 Jahre beschränkter Nachbeobachtung und perfekter Nachsorge.

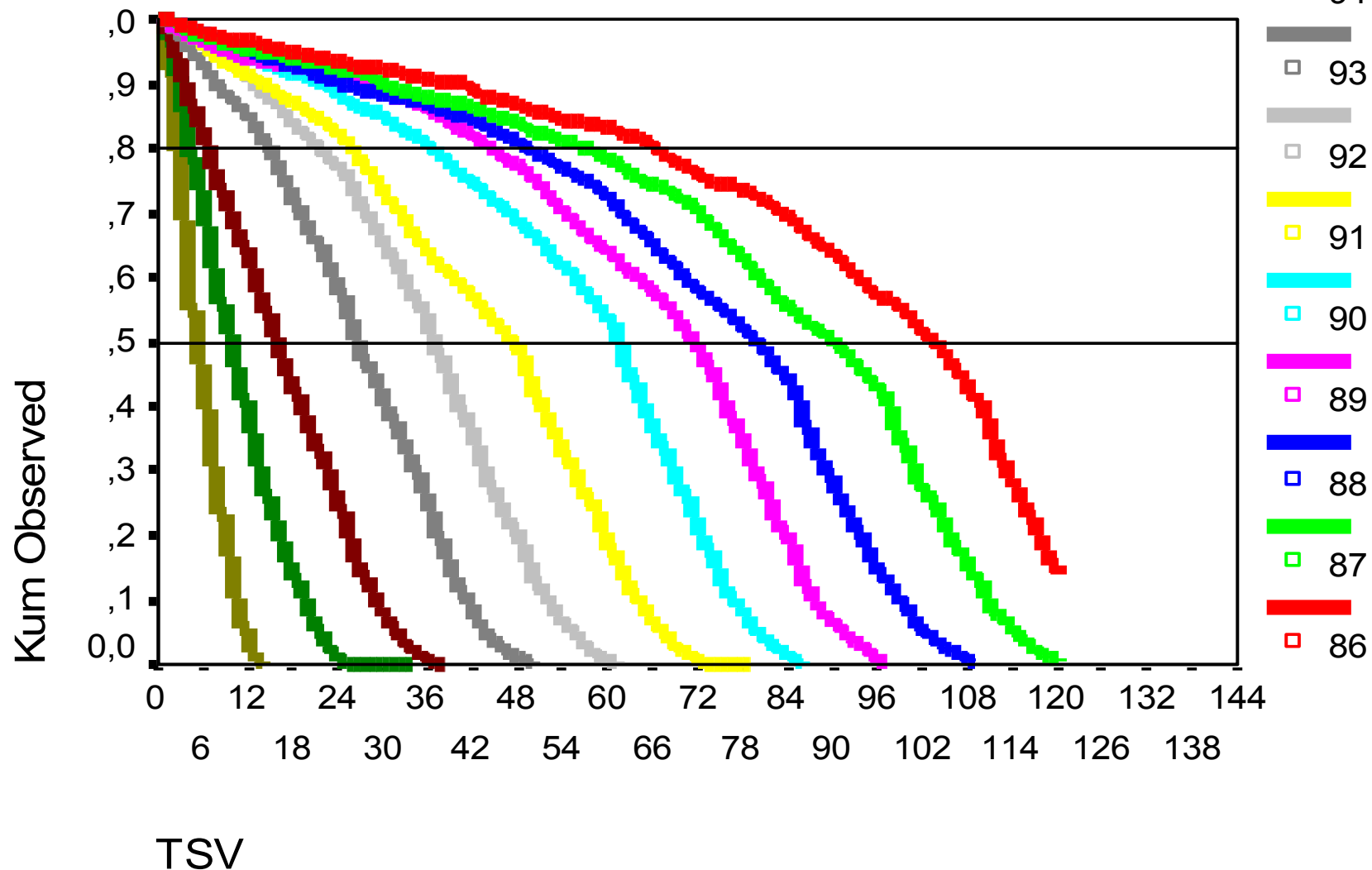
Mediane Beobachtungszeit
Hier ablesen!



Qualität des Follow-ups
des Europäischen
Knochenmark-
transplantations-Registers
der EBMT nach
Jahreskohorten

Im Vergleich zur
optimalen Nach-
beobachtung
deutliche Defizite bei
etwa 50% der Fälle.

Distribution of observation times by year of transplant 1986-96



Likelihoodfunktion für zensierte Daten II

$$\text{prDichte}(t,0) = \text{prDichte}(C = t \wedge T > C) =$$

$$\frac{d}{dt} [pr(C \leq t \wedge T > C)] dt = \frac{d}{dt} \left[\int_0^t \int_t^\infty h(\tau, \zeta) d\tau d\zeta \right] dt =$$

$$\frac{d}{dt} \left[\int_0^t \int_t^\infty f(\tau) * g(\zeta) d\tau d\zeta \right] dt = \frac{d}{dt} \left[\int_0^t S(t) * g(\zeta) d\zeta \right] dt$$

$$= S(t) \cdot g(t) dt$$

Analog (vertauschte Rollen von T und C):

$$\text{prDichte}(t,1) = f(t) \cdot G(t) dt,$$

wobei $G(t) := \int_c^\infty g(\xi) d\xi =$ Inverse Verteilungsfunktion der Beobachtungszeiten

Annahme : nicht - informative Zensur

$$h(\tau, \zeta) d\tau d\zeta = f(\tau) \cdot g(\zeta) d\tau d\zeta$$

$$S(t) := \int_t^\infty f(\tau) d\tau$$

Invers kummulierte Verteilungsfunktion

Likelihoodfunktion für zensierte Daten III

Likelihoodfunktion Überlebenszeiten mit zensierten Daten :

$$L = \prod_{i=1}^n \text{prDichte}(t_i, \delta_i) = \prod_{i=1}^n \underbrace{[f(t_i) \cdot G(t_i)]^{\delta_i}}_{\text{Ereignisse}} \underbrace{[g(t_i) \cdot S(t_i)]^{1-\delta_i}}_{\text{Zensierungen}}$$

$$\propto \prod_{i=1}^n [f(t_i)]^{\delta_i} [S(t_i)]^{1-\delta_i}$$

(konditionell auf Verteilung der Beobachtungszeiten,
die keine Information über Verteilung der Ereigniszeiten enthält.
[bloßer konstanter Faktor])

Likelihoodfunktion für zensierte Daten IV

Parametrische Schätzer erhält man durch Maximieren der Likelihoodfunktion über eine geeignete Familie von invers kummulierten Verteilungsfunktionen S_θ

$$\hat{\theta} = \operatorname{argmax}(L(\theta)):$$

a) zweiparametrische Weibull - Verteilung : Schätze $\hat{\lambda}, \hat{\alpha}$

b) Familie "Stetige Funktionen mit n Sprungstellen"

⇒ Kaplan Meier Schätzer ist $2n$ - parametrischer Maximum Likelihood Schätzer.