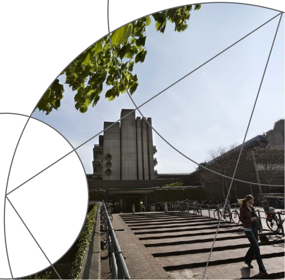
Survival Analysis
Week 2: Parametric models

Section of Biostatistics



Parametric survival models

Allows for

- easy calculation of selected quantiles of the survival distribution
- estimation of the expected survival time (usually by extrapolation)
- estimation of the hazard $\alpha(t)$ itself, not only the cumulated hazard $\int_0^t \alpha(u)du$
- incorporation of multiple time scales
- enhanced understanding of the failure mechanism
- straightforward handling of interval censored data

. . . but are not as flexible as semi- and nonparametric models.

Hazard rate

Let T be a continuous time to event

- Cumulative distribution function $F(t) = \mathrm{pr}(T \leq t)$
- Survival function $S(t) = \int_t^\infty f(u)du = 1 - F(t) = \mathrm{pr}(T > t)$
- Density $f(t) = \partial F(t)/\partial t$

The hazard rate $\alpha(t)$ is the conditional event rate at time t for those still alive at time t,

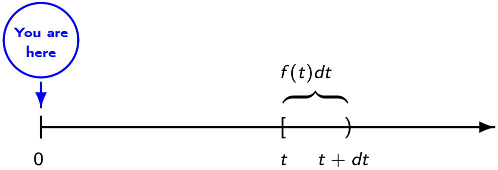
$$\begin{aligned}\alpha(t) &= \lim_{h\to\infty} \frac{\mathrm{pr}(t \leq T < t+h | T \geq t)}{h} \\ &= \frac{\lim_{h\to\infty} \mathrm{pr}(t \leq T < t+h)/h}{\mathrm{pr}(T \geq t)} \\ &= \frac{f(t)}{S(t-)}\end{aligned}$$

Using Leibniz notation, for infinitely small dt, we write

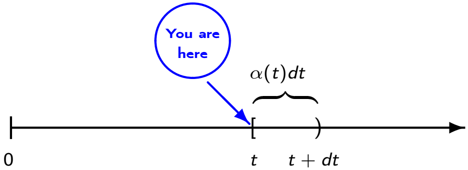
$$\alpha(t)dt = \mathrm{pr}(t \leq T < t + dt | T \geq t)$$

Density vs. hazard

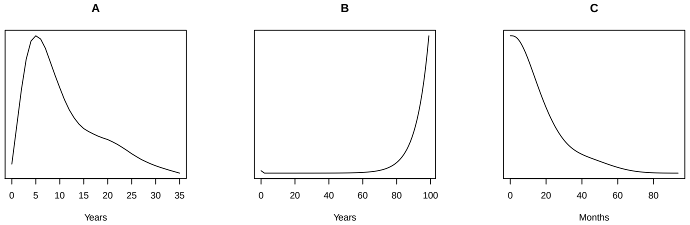
- The density is a margial rate: $f(t)dt = \text{pr}(t \leq T < t + dt)$



- The hazard is a conditional rate: $\alpha(t)dt = \text{pr}(t \leq T < t + dt | T \geq t)$



Examples of hazards

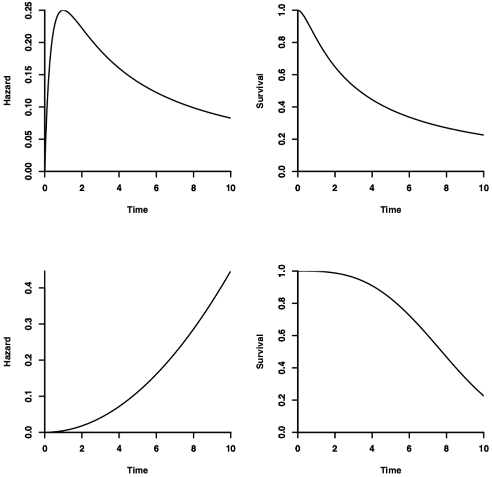


What could the hazards in A, B and C be?
What event and what time origin could give rise to such hazards?

Hazard and survival

$$\alpha(s) = \frac{f(s)}{S(s)} = -\frac{\partial}{\partial s} \log S(s) \Leftrightarrow \int_0^t \alpha(s) ds = -\log S(t)$$

$$\Leftrightarrow S(t) = \exp\left(-\int_0^t \alpha(s) ds\right)$$



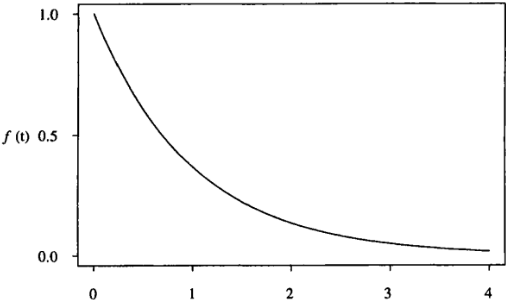
Exponential distribution

The exponential distribution corresponds to a constant hazard.

- Hazard $\alpha(t) = \lambda$, $\lambda > 0$

The hazard is the same regardless of how long the individual has been on study; it is memoryless.

- $S(t) = e^{-\lambda t}$
- Density $f(t) = \lambda e^{-\lambda t}$
- $E(T) = 1/\lambda$, $\text{var}(T) = 1/\lambda^2$



Exponential distribution

When T is exponentially distributed, the density of

$$Y = \log T$$

is

$$\begin{aligned}
\frac{\partial}{\partial y}\mathrm{pr}(Y \leq y) &= \frac{\partial}{\partial y}\mathrm{pr}(T \leq e^y) \\
&= f(e^y)e^y \\
&= \lambda e^{-\lambda e^y} e^y \\
&= \exp\left(\log\lambda - \lambda e^y + y\right) \\
&= \exp\left(\log\lambda - e^{y+\log\lambda} + y\right) \\
&= \exp\left(y - \alpha - e^{y-\alpha}\right), \quad -\infty < y < \infty,
\end{aligned}$$

where $\alpha = -\log\lambda$.

Letting $Y = \alpha + W$, the density is of W

$$\exp\left(w - e^w\right), \quad -\infty < w < \infty,$$

the extreme value distribution.

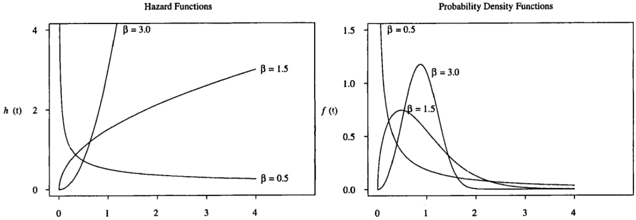
Weibull distribution

The Weibull distribution is a generalization of the exponential distribution allowing for a power dependence of the hazard on time.

The hazard with parameters λ and γ is

$$\alpha(t) = \lambda\gamma(\lambda t)^{\gamma-1}, \ \lambda > 0, \gamma > 0$$

- $S(t) = \exp\left(-(\lambda t)^{\gamma}\right)$
- $f(t) = \alpha\gamma(\lambda\gamma t)^{\gamma-1} \exp\left(-(\lambda t)^{\gamma}\right)$



γ is a shape parameter

- $\gamma < 1$: $\alpha(t)$ is decreasing
- $\gamma > 1$: $\alpha(t)$ is increasing
- $\gamma = 1$: $\alpha(t) = \lambda$, i.e., reduces to the exponential distribution

Weibull distribution

When T follows a Weibull distribution, $Y = \log T$ can we written

$$Y = \alpha + \sigma W,$$

where W follows the extreme value distribution. The relation between the parameters in the two representatons is $\alpha = -\log \lambda$ and $\sigma = \gamma^{-1}$. See Exercise C from week 1.

The parameters α and σ affect only the location and scaling of the distribution of Y, the shape of the density is fixed.

Log-normal distribution

When T is log-normally distributed, $Y = \log T$ is normally distributed, or

$$Y = \alpha + \sigma W$$

where W has a standard normal distribution. The density of Y is

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(y-\alpha)^2\right)$$

and the density of $T = \exp Y$ is

$$f_T(t) = \frac{1}{\sqrt{2\pi}t}\gamma \exp\left(-\frac{\gamma^2}{2}(\log(\lambda t))^2\right).$$

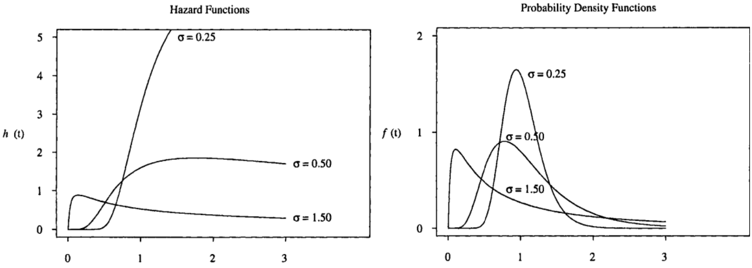
where $\alpha = -\log\lambda$ and $\sigma = \gamma^{-1}$.

The hazard and survival functions involve the normal distribution function

$$\Phi(w) = \int_{-\infty}^{w} \frac{e^{-u^2/2}}{\sqrt{2\pi}} du,$$

- Survival $S(t) = 1 - \Phi(\gamma \log(\lambda t))$
- Hazard $\alpha(t) = f_T(t)/S(t)$

Log-normal distribution



Log-logistic distribution

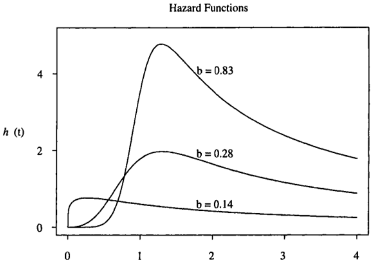
The log-normal is simple to apply if there is no censoring, but with censoring the computation becomes more difficult. A log-logistic distribution

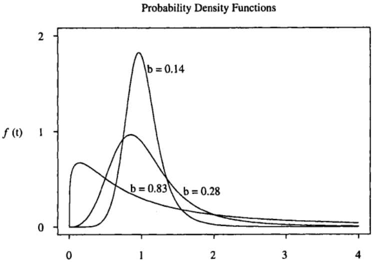
$$Y = \alpha + \sigma W,$$

where W follows a logistic distribution with density $e^w/(1+e^w)^2$, a symmetric mean zero densisty with slightly heavier tails than the standard normal distribution, provides a good approximation to the log-normal distribution. It is more convenient as it has closed form expression for the hazard and survival functions

$$\alpha(t) = \frac{\lambda\gamma(\lambda t)^{\gamma-1}}{1+(\lambda t)^\gamma}, \quad S(t) = \frac{1}{1+(\lambda t)^\gamma}$$

where, again, $\alpha = -\log\lambda$ and $\sigma = \gamma^{-1}$.





Parametric regression models

- In most studies there are covariates such as treatment or individual characteristics whose relationship to lifetime is of interest.
- Survival regression models specifiy the distribution of a lifetime T given a vector of covariates Z.
- Parametric models can be made into regression models by specifying a relationship between the model parameters and covariates.

Accelerated failure time (AFT) model

In the parametric AFT model, $Y = \log T$ is related to the covariate Z by the linear model,

$$Y = \beta^T Z + W$$

where W is an error variable
The covariates have a multiplicative effect on the survival time

$$T = e^{\beta^T Z} T_0 \text{ where } T_0 = e^W$$

The role of the covariates is to accelerate or decelerate the time to event.

- Let $\alpha_0(\cdot;\gamma)$ be the hazard function for T_0 specified by letting T_0 follow some parametric distribution indexed by γ.
- The distribution of T_0 doesn't involve β
- The hazard for T is

$$\alpha(t) = -\frac{\partial}{\partial t}\log \mathrm{pr}(T > t) = -\frac{\partial}{\partial t}\log \mathrm{pr}(T_0 > te^{-\beta^Z}) = \alpha_0(te^{-\beta^T Z};\gamma)e^{-\beta^T Z}.$$

- The survival function is

$$S(t|Z) = \mathrm{pr}(T > t) = \mathrm{pr}(T_0 > te^{-\beta^T Z}) = S_0(te^{-\beta^T Z}),$$

where $S_0(t;\gamma) = \mathrm{pr}(T_0 > t)$ is the survival function for T_0.

Proportional hazards (PH) model

- In the PH model, the covariates act multiplicatively on the hazard function

$$\alpha(t|Z) = \underbrace{\alpha_0(t;\gamma)}_{\text{doesn't involve Z}} \overbrace{e^{\beta^T Z}}^{\text{doesn't involve time}}$$

  where $\alpha_0(\cdot;\gamma)$ is a parametric baseline hazard indexed by γ.
- The covariates act on the baseline survival function (corresponding to a subject with all covariates zero) $S_0(t;\gamma) = \exp\left(-\int_0^t \alpha_0(u;\gamma)du\right)$, by raising it to a power

$$S(t|Z) = \exp\left(-\int_0^t \alpha_0(u;\gamma)e^{\beta^T Z}du\right) = S_0(t;\gamma)^{e^{\beta^T Z}}.$$

- The ratio for the hazards corresponding to covariates Z_1 and Z_2,

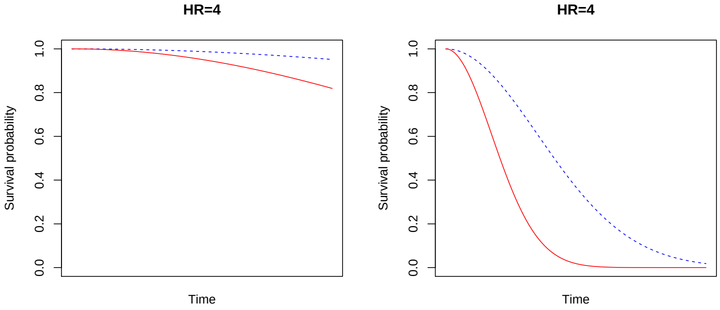
$$\frac{\alpha(t|Z_1)}{\alpha(t|Z_2)} = \frac{\alpha_0(t)e^{\beta^T Z_1}}{\alpha_0(t)e^{\beta^T Z_2}} = \exp(\beta^T(Z_1 - Z_2))$$

  doesn't depend on time. The hazards are proportional.

The HR does not give the whole picture

Hazard ratios cannot generally be extended to a corresponding ratio of survival probabilities or absolute risks

The hazard ratio is 4 in both plots:



The ratio of survival probabilities or absolute risks will depend on time even when the hazard ratio is constant over time.

Weibulll PH and AFT

Letting an individual without covarites have Weibull distributed with paramters (γ_1, λ_1) T corresponds to the baseline hazrd

$$\alpha_0(t; \gamma_1, \lambda_1) = \gamma_1 \lambda_1^{\gamma_1} t^{\gamma_1 - 1}$$

and thus the proportional hazards model

$$\alpha(t; Z) = \gamma_1 \lambda_1^{\gamma_1} t^{\gamma_1 - 1} e^{\beta_1^T Z}$$

Let

$$\alpha_0(t; \gamma_2, \lambda_2) = \gamma_2 \lambda_2^{\gamma_2} t^{\gamma_2 - 1}$$

then the hazard in the AFT Weibull model is

$$\begin{aligned}
\alpha(t; Z) &= \gamma_2 \lambda_2^{\gamma_2} \left(t e^{-\beta_2^T Z} \right)^{\gamma_2 - 1} e^{-\beta_2^T Z} \\
&= \gamma_2 \lambda_2^{\gamma_2} e^{-\beta_2^T Z(\gamma_2 - 1)} t^{\gamma_2 - 1} e^{-\beta_2^T Z} \\
&= \underbrace{\gamma_2 \lambda_2^{\gamma_2} t^{\gamma_2 - 1}}_{\text{doesn't involve Z}} e^{-\gamma_2 \beta_2^T Z}
\end{aligned}$$

The AFT Weibull model is of proportional hazards form. To see that the models are quivalent, set $\gamma_1 = \gamma_2$ and $\beta_1 = -\gamma_2 \beta_2$. It is only the Weibull model (and thus its special case the exponential) that can be both a PH and a AFT model. The log-normal hazard functions with different location parameters are not proportional.

Parametric likelihood

Let T^* and C be the event and censoring times, respectively. We observe

$$(T_i = T_i^* \wedge C_i, \Delta_i = I(T_i^* \leq C_i), Z_i), \ \ i = 1, \dots, n$$

from $i = 1, \dots, n$ independent subjects.

Assumptions

- The hazard is known up to $\theta^T = (\beta^T, \gamma^T)$ where β are the regression parameters and γ are the parameters for α_0
- Independent censoring: T^* and C are independent conditionally on Z
- The censoring is noninformative: The censoring distribution doesn't involve θ

The likelihood for θ is proportional to

$$\begin{aligned}
L(\theta) &= \prod_{i=1}^{n} f(T_i|Z_i;\theta)^{\Delta_i} S(T_i|Z_i;\theta)^{1-\Delta_i} \\
&= \prod_{i=1}^{n} \alpha(T_i|Z_i;\theta)^{\Delta_i} S(T_i|Z_i;\theta)^{\Delta_i} S(T_i|Z_i;\theta)^{1-\Delta_i} \\
&= \prod_{i=1}^{n} \alpha(T_i|Z_i;\theta)^{\Delta_i} S(T_i|Z_i;\theta) \\
&= \prod_{i=1}^{n} \alpha(T_i|Z_i;\theta)^{\Delta_i} \exp\left(-\int_0^{T_i} \alpha(u|Z_i;\theta) du\right)
\end{aligned}$$

Counting process notation

The log-likelihood is proportional to

$$\ell(\theta) = \log L(\theta) = \sum_{i=1}^{n} \left(\Delta_i \log \alpha(T_i|Z_i;\theta) - \int_0^{T_i} \alpha(u|Z_i;\theta)du \right)$$

- Define the counting process $N_i(t) = \Delta_i I(T \leq t)$, counting $(+1)$ the occurrence of an uncensored event for individual i.
- $dN_i(t) = N_i(t) - N_i(t-) = \Delta_i I(T_i = t)$ is one if N_i jumps at time t, otherwise zero.
- Integrating a function $f(t)$ with respect to $N_i(t)$ corresponds to evaluating f at the event time, if the event is observed to occur; if it is censored the integral is zero,

$$\int_s^t f(u)dN_i(u) = I(T_i \in [s,t])\Delta_i f(T_i).$$

- Also define the at-risk indicator $Y_i(t) = I(T_i \geq t)$ that is one if the individual is both still alive and uncensored at time t

With counting process notation,

$$\ell(\theta) = \sum_{i=1}^{n} \left(\int_0^\infty \log \alpha(u|Z_i;\theta)dN_i(u) - \int_0^\infty Y_i(u)\alpha(u|Z_i;\theta)du \right)$$

Scores and information

The score function is

$$
\begin{aligned}
U(\theta) &= \frac{\partial}{\partial\theta}\ell(\theta) \\
&= \sum_{i=1}^{n}\int_0^\tau \frac{\partial \log \alpha(t|Z_i;\theta)}{\partial\theta}dN_i(t) - \int_0^\tau Y_i(t)\overbrace{\frac{\partial\alpha(t|Z_i;\theta)}{\partial\theta}}^{\left(\frac{\partial \log \alpha(t|Z_i;\theta)}{\partial\theta}\right)\alpha(t|Z_i;\theta)}dt \\
&= \sum_{i=1}^{n}\int_0^\tau \frac{\partial \log \alpha(t|Z_i;\theta)}{\partial\theta}\underbrace{(dN_i(t) - Y_i(t)\alpha(t|Z_i;\theta)dt)}_{=dM_i(t;\theta)} \\
&= \sum_{i=1}^{n}\int_0^\tau \frac{\partial \log \alpha(t|Z_i;\theta)}{\partial\theta}dM_i(t;\theta)
\end{aligned}
$$

The processes

$$M_i(t;\theta_0) = \int_0^t dM_i(t;\theta_0),$$

evaluated at the true θ_0 have zero mean (and are continuous time martingales, the topic for next week).

The martingale residual

A heuristic argument that $E(dM_i(t;\theta_0)|\text{History prior to } t) = 0$ is

- For t such that individual i is at risk, $Y_i(t) = 1$,

$$E(dN_i(t)|T_i \geq t, Z_i) = \frac{\text{pr}(T^* \in [t, t+dt), C \geq t|Z_i)}{\text{pr}(T^* \geq t, C \geq t|Z_i)}$$
$$\underbrace{=}_{T^* \perp\!\!\!\perp C} \frac{\text{pr}(T^* \in [t, t+dt)|Z_i)}{\text{pr}(T^* \geq t|Z_i)} = \alpha(t|Z_i;\theta_0)dt$$

- For t such that subject i is no longer at risk, i.e., $Y_i(t) = 0$, we know that $dN_i(t) = 0$
- The probability that N_i jumps "now" given "the past" is the intensity

$$\lambda_i(t) = E(dN_i(t)|\text{History up to } t) = Y_i(t)\alpha(t|Z_i;\theta_0)dt$$

- Thus,

$$\begin{aligned} E(dM_i(t;\theta_0)|\text{History up to } t) &= E(dN_i(t) - Y_i(t)\alpha(t|Z_i;\theta_0)dt|\text{History up to } t) \\ &= E(dN_i(t)|\text{History up to } t) - Y_i(t)\alpha(t|Z_i;\theta_0)dt = 0 \end{aligned}$$

Intensities and history

- Later in the course "History prior to t" will be formalized as a filtration $\mathcal{H}_t$, a nested sequence of σ-fields containing increasing (in t) information. The intensity functions λ are stochastic, being functions of the history $\mathcal{H}_{t-}$.
- The filtration corresponds to what information is given. One may condition on more information than contained in $\mathcal{H}_t$, say $\mathcal{F}_t$ such that $\mathcal{H}_t \subseteq \mathcal{F}_t$.
- The innovation theorem: If the intensity with respect to $\mathcal{F}_{t-}$ is λ, the intensity with respect to $\mathcal{H}_t$ is

$$\tilde{\lambda}(t) = E(\lambda(t)|\mathcal{H}_{t-})$$

which is generally different since we condition on less information.

Score

Because the subjects are independent and we have assumed that $T_i^* \perp\!\!\!\perp C_i | Z_i$, the terms of the score function

$$\mathrm{U}(\theta) = \sum_{i=1}^{n} \int_0^{\tau} \frac{\partial \log \alpha(t|Z_i;\theta)}{\partial \theta} dM_i(t;\theta)$$

are i.i.d..

$\mathrm{U}(\theta)$ is an unbiased estimation equation,

$$E\left(\mathrm{U}(\theta_0)\right) = E\left(\sum_{i=1}^{n} \int_0^{\tau} \frac{\partial \log \alpha(t|Z_i;\theta_0)}{\partial \theta} dM_i(t;\theta_0)\right) = 0.$$

This follows as

$$E(dM_i(t,\theta_0)|\mathcal{H}_{t-}) = 0,$$

where $\mathcal{H}_{t-}$ denotes the history prior to t, and that the randomness in the hazard $\alpha(t|Z_i;\theta)$ is contained in the history,

$$E(\alpha(t|Z_i;\theta)|\mathcal{H}_{t-}) = \alpha(t|Z_i;\theta).$$

Score

Standard results apply for the asymptotics of the parametric MLE $\hat{\theta}$ solving

$$\mathrm{U}(\hat{\theta}) = 0,$$

e.g.

$$n^{1/2}\left(\hat{\theta} - \theta_0\right) \xrightarrow{\mathcal{L}} N\left(0, \mathcal{I}(\theta_0)^{-1}\right)$$

where

$$n\mathcal{I}(\theta) = E\left(-\frac{\partial^2}{\partial\theta\partial\theta^T}\ell(\theta)\right)$$

is the expected information matrix. The expectation involves the censoring distribution that often is unknown, nuisance and not modelled. However, $\mathcal{I}(\theta_0)$ can be consistently estimated by $I(\hat{\theta})/n$, where

$$I(\theta) = -\frac{\partial^2}{\partial\theta\partial\theta^T}\ell(\theta) = -\frac{\partial}{\partial\theta}\mathrm{U}(\theta)$$

is the observed information.

The likelihood ratio, score and Wald tests apply as usual.