Survival Models

Lecture III. Fitting parametric distributions to survival data

Exponential distribution (1)

Main characteristics:

- parameter: rate $\lambda > 0$
- support: $t \geq 0$
- pdf: $f(t) = \lambda \exp(-\lambda t)$
- survivor function: $S(t) = \exp(-\lambda t)$
- hazard function: $h(t) = \lambda$
- mean: $E(T) = \frac{1}{\lambda}$
- variance: $Var(T) = \frac{1}{\lambda^2}$

For cumulative hazard, we can write:

$$\ln H(t) = \ln(-\ln(S(t))) = \ln(\lambda) + \ln(t)$$

or, equivalently:

$$\ln(t) = -\ln(\lambda) + \ln(-\ln(S(t)))$$

In other words, the plot of $\ln(t)$ vs $\ln(-\ln(S(t)))$ is a straight line with slope 1 and intercept $-\ln(\lambda)$.

Exponential distribution (2)

Let $t_1, t_2, \ldots, t_r$ be the (ordered) observed death/failure times.

Two simple graphical tests can be done to visually test the goodness of fit of the expolential model to the data:

1. QQ-plot with $\ln(t_i)$ (uncensored sample quantiles in log-scale) against parametric candidate quantile $\ln(-\ln(\hat{S}(t_i)))$
2. Plots of empirical hazards against time. Let us recall that the empirical hazard can be estimated in two ways:
   - estimate at an observed death time t_i

$$\tilde{h}(t_i) = \frac{d_i}{n_i}$$

   - estimate for death rate per unit time in the interval $t_i \leq t < t_{i+1}$

$$\hat{h}(t) = \frac{d_i}{n_i(t_{i+1} - t_i)}$$

Weibull distribution (1)

Main characteristics:

- parameters: rate (scale) $\lambda > 0$, shape $\alpha > 0$
- support: $t \geq 0$
- pdf: $f(t) = \lambda\alpha(\lambda t)^{\alpha-1}\exp(-(\lambda t)^{\alpha})$
- survivor function: $S(t) = \exp(-(\lambda t)^{\alpha})$
- hazard function: $h(t) = \lambda\alpha(\lambda t)^{\alpha-1}$
- mean: $E(T) = \frac{1}{\lambda}\Gamma(1 + \frac{1}{\alpha})$
- median: $\text{med}\, T = \frac{1}{\lambda}(\ln 2)^{\frac{1}{\alpha}}$
- variance: $Var(T) = \frac{1}{\lambda^2}\Gamma(1 + \frac{2}{\alpha}) - \frac{1}{\lambda^2}\Gamma(1 + \frac{1}{\alpha})^2$

Note that the hazard function is

- monotonely increasing when $\alpha > 1$
- monotonely decreasing when $\alpha < 1$
- constant when $\alpha = 1$

Weibull distribution (2)

For cumulative hazard, we can write:

$$\ln H(t) = \ln(-\ln(S(t))) = \alpha(\ln(\lambda) + \ln(t))$$

or, equivalently:

$$\ln(t) = -\ln(\lambda) + \frac{1}{\alpha}\ln(-\ln(S(t)))$$

Thus, the plot of $\ln(t)$ vs $\ln(-\ln(S(t)))$ is a straight line with slope $\frac{1}{\alpha}$ and intercept $-\ln(\lambda)$.

This linear relationship can be used to construct QQ-plot (similarly to exponential distribution)!

The extreme (minimum) value distribution (1)

Main characteristics:

- parameters: location $\mu \in \mathbb{R}$, scale $\sigma > 0$
- support: $y \in \mathbb{R}$
- pdf: $f(y) = \frac{1}{\sigma} \exp\left(\frac{y-\mu}{\sigma} - \exp\left(\frac{y-\mu}{\sigma}\right)\right)$
- survivor function: $S(y) = \exp\left(-\exp\left(\frac{y-\mu}{\sigma}\right)\right)$
- mean: $E(Y) = \mu - \gamma\sigma$
- variance: $Var(Y) = \frac{\pi^2}{6}\sigma^2$

Here γ denotes Euler's constant, $\gamma = 0.5772...$

The extreme (minimum) value distribution is also called the Gumbel (minimum) distribution

The extreme (minimum) value distribution (2)

NB!

If T is a Weibull rv with parameters α and λ, then $Y = \ln T$ follows an extreme (minimum) value distribution with $\mu = -\ln(\lambda)$ and $\sigma = \frac{1}{\alpha}$.

This representation allows us to write

$$Y = \mu + \sigma Z,$$

where Z is a standard extreme value random variable (with $\mu = 0$ and $\sigma = 1$)

The log-normal distribution (1)

- parameters (1): from corresponding normal distribution, mean $\mu \in \mathbb{R}$, std. dev. $\sigma > 0$
- parameters (2): $\lambda > 0$ and $\alpha > 0$ defined by $\mu = -\ln(\lambda)$ and $\sigma = \frac{1}{\alpha}$
- support: $t \geq 0$
- pdf: $f(t) = \frac{\alpha}{\sqrt{2\pi}t} \exp\left(\frac{-\alpha^2(\ln(\lambda t))^2}{2}\right)$
- survivor function: $S(t) = 1 - \Phi\left(\alpha \ln(\lambda t)\right)$
- hazard function: $h(t) = \frac{f(t)}{S(t)}$
- mean: $E(T) = \exp(\mu + \frac{\sigma^2}{2})$
- variance: $Var(T) = (\exp(\sigma^2) - 1)\exp(2\mu + \sigma^2)$

The log-normal distribution (2)

The hazard function $h(t)$

- has value 0 at $t = 0$
- increases to maximum then starts to decrease

Therefore,

- since the hazard is decreasing for large values of t, it is not plausible to model a lifetime in most situations
- log-normal distribution is still usable in situations where large values of t are not of interest (e.g., tuberculosis)

The log-logistic distribution (1)

- parameters: rate $\lambda > 0$ and shape $\alpha > 0$
- support: $t \geq 0$
- pdf: $f(t) = \lambda\alpha(\lambda t)^{\alpha-1}(1 + (\lambda t)^{\alpha})^{-2}$
- survivor function: $S(t) = \frac{1}{1+(\lambda t)^{\alpha}}$
- hazard function: $h(t) = \frac{\lambda\alpha(\lambda t)^{\alpha-1}}{1+(\lambda t)^{\alpha}}$
- mean: $E(T) = \exp(\mu + \frac{\sigma^2}{2})$
- variance: $Var(T) = (\exp(\sigma^2) - 1)\exp(2\mu + \sigma^2)$

The log-logistic distribution (2)

NB!

If T is a log-logistic rv with parameters α and λ, then $Y = \ln T$ follows a logistic distribution with $\mu = -\ln(\lambda)$ and $\sigma = \frac{1}{\alpha}$.

This representation allows us to write

$$Y = \mu + \sigma Z,$$

where Z is a standard logistic random variable variable with density

$$f_Z(z) = \frac{\exp(z)}{(1 + \exp(z))^2}$$

The log-logistic distribution (3)

The hazard function $h(t)$ of log-logistic distribution is similar to the hazard of Weibull distribution (aside from the denominator factor)

- for $\alpha < 1$ $(\sigma > 1)$ it is monotone decreasing from ∞
- for $\alpha = 1$ $(\sigma = 1)$ it is monotone decreasing from λ
- for $\alpha > 1$ $(\sigma < 1)$ it resembles the log-normal hazard:
  - has value 0 at $t = 0$
  - increases to maximum at $t = (\alpha - 1)^{\frac{1}{\alpha}}$, then starts to decrease

The log-logistic distribution (4)

The odds of survival beyond time t are

$$\frac{S(t)}{1 - S(t)} = (\lambda t)^{-\alpha}$$

From here we get a linear relationship

$$\ln(t) = \mu + \sigma \left(- \ln \frac{S(t)}{1 - S(t)} \right),$$

where $\mu = -\ln(\lambda)$ and $\sigma = \frac{1}{\alpha}$

This linear relationship between $\ln(t)$ and $\left(-\ln \frac{S(t)}{1-S(t)}\right)$ (slope σ, intercept μ) can can be used to construct QQ-plot.

The gamma distribution (1)

Main characteristics:

- parameters: rate (scale) $\lambda > 0$, shape $k > 0$
- support: $t \geq 0$
- pdf: $f(t) = \frac{\lambda^k t^{k-1}}{\Gamma(k)} \exp(-\lambda t)$
- survivor function: no simple form
- hazard function: no simple form
- mean: $E(T) = \frac{k}{\lambda}$
- variance: $Var(T) = \frac{k}{\lambda^2}$

The gamma distribution (2)

The hazard function $h(t)$ for gamma distribution is

- monotone increasing from 0 when $k > 1$
- monotone decreasing from ∞ when $k < 1$
- approaches λ as t increases (for both cases)
- constant (λ) when $k = 1$

Relationships between distributions. Location-scale family

We established that most of the distributions of lifetime T had the property that the distribution of log-transform $\ln(T)$ is the member of location-scale family.

Common features to remember:

- distribution of time T has two parameters, scale λ and shape α
- in log-time, $Y = \ln(T)$, the distribution has two parameters, location $\mu = -\ln(\lambda)$, scale $\sigma = \frac{1}{\alpha}$
- each rv Y can be expressed as $Y = \mu + \sigma Z$, where Z is the standard member, i.e. $\mu = 0$ $(\lambda = 1)$ and $\sigma = 1$ $(\alpha = 1)$
- the models are log-linear

Summary of relationships between distributions of interest:

T	$Y = \ln(T)$
Weibull	extreme minimum value
log-normal	normal
log-logistic	logistic

Construction of the QQ-plot (1)

Required notation:

- $\hat{S}(t)$ – the K-M estimator of survival probability beyond t
- t_i, $i = 1, \ldots, r \leq n$ – ordered uncensored failure times
- $\hat{p}_i = 1 - \hat{S}(t_i)$ – estimated failure probability

Recall also that for Weibull, log-normal and log-logistic distribution we derived certain "useful linear relationships" that are summarized in the following table

Table 3.1: *Relationships to exploit to construct a graphical check for model adequacy*

t_p quantile	$y_p = \log(t_p)$ quantile	form of standard quantile z_p
Weibull	extreme value	$\log(-\log(S(t_p))) = \log(H(t_p))$ $= \log(-\log(1-p))$
log-normal	normal	$\Phi^{-1}(p)$, where Φ denotes the standard normal d.f.
log-logistic	logistic	$-\log\left(\frac{S(t_p)}{1-S(t_p)}\right) = -\log(\text{odds})$ $= -\log\left(\frac{1-p}{p}\right)$

Construction of the QQ-plot (2)

Now, using these "useful linear relationships", we can determine parametric standard quantiles z_i from

$$F_{0,1}(z_i) = \mathbb{P}(Z \leq z_i) = \hat{p}_i,$$

where

- the value of $\hat{p}_i$ comes from our K-M estimate
- the form of z_i is determined by the model (distribution)
- $F_{0,1}$ is the standard parametric model ($\mu = 0$, $\sigma = 1$) under consideration.

As the K-M estimator is distribution free and estimates "true" survival function, for large sample sizes n, the z_i should reflect "true" standard quantiles, if F is the "true" lifetime distribution function.

Thus, if the proposed candidate distribution fits the data adequately, the points $(z_i, \ln(t_i))$ should lie close to a straight line with slope σ and intercept μ. Such plot is called quantile-quantile plot (QQ-plot)

Sidenote: comparison with "regular" QQ-plot (1)

Idea of QQ-plot:

- plot the sample quantiles against the theoretical quantiles
- if the proposed distribution fits data well, the points on the QQ-plot lie close to a straight line

For an ordered sample $(x_{(1)}, \ldots, x_{(n)})$ and theoretical candidate distribution H, we have

- $\frac{i}{n}$-th sample quantile $x_{(i)}$
- $\frac{i}{n}$-th theoretical quantile $H^{-1}\left(\frac{i}{n}\right)$

Thus we can plot (for technical reasons $\frac{i}{n+1}$-th theoretical quantile is used)

$$\left\{x_{(i)}, H^{-1}\left(\frac{i}{n+1}\right)\right\}, \quad i = 1, \ldots, n.$$

Sidenote: comparison with "regular" QQ-plot (2)

Problem

In case of censored data we CAN NOT order our sample!
We can only order the (observed) failure times.

Let us have (ordered) failure times t_i, $i = 1, \ldots, r \leq n$

Using KM estimator $\hat{S}$ and denoting $\hat{p}_i = 1 - \hat{S}(t_i)$, we have

- $\hat{p}_i$-th sample quantile (in log-scale) $\ln t_i$
- $\hat{p}_i$-th theoretical quantile $z_i = H^{-1}(\hat{p}_i)$

and we plot

$$\left\{\ln t_i, H^{-1}(\hat{p}_i)\right\}, \quad i = 1, \ldots, r.$$

Maximum likelihood estimation (1)

Let us recall the likelihood function

$$L(\theta) = \prod_{i=1}^{n} f(t_i|\theta)$$

and the log-likelihood function

$$\ln L(\theta) = \sum_{i=1}^{n} \ln f(t_i|\theta)$$

The maximum likelihood estimator $\hat{\theta}$ is the solution of

$$\frac{\partial \ln L(\theta)}{\partial \theta} = 0$$

Maximum likelihood estimation (2)

In case of survival data, we need to take the censoring into account, thus the likelihood function has the following form

$$L(\theta) = \prod_{i=1}^{n} f^{\delta_i}(y_i|\theta) S^{1-\delta_i}(y_i|\theta)$$

and the corresponding log-likelihood is

$$\ln L(\theta) = \ln \prod_{i=1}^{n} f^{\delta_i}(y_i|\theta) S^{1-\delta_i}(y_i|\theta) = \sum_{u} \ln f(y_i|\theta) + \sum_{c} \ln S(y_i|\theta),$$

where u and c mean sums over the uncensored and censored observations, respectively

MLE for exponential model (1)

For simplicity, let us first assume that all failures are observed, i.e. no censoring

The likelihood function and log-likelihood function for exponential distribution are

$$L(\lambda) = \prod_{i=1}^{n} \lambda \exp(-\lambda t_i) = \lambda^n \exp\left(-\lambda \sum_{i=1}^{n} t_i\right),$$

$$\ln L(\lambda) = n \ln \lambda - \lambda \sum_{i=1}^{n} t_i$$

MLE for exponential model (2)

Taking derivative of the last expression gives

$$\frac{\partial \ln L(\lambda)}{\partial \lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} t_i,$$

which implies that the MLE for λ is

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} t_i}$$

The confidence intervals can be constructed using the exact distribution theory because of the relation between exponential (gamma) and chi-square distributions. More precisely, let us have iid random variables $T_i \sim Exp(\lambda)$, then

- $\sum_{i=1}^{n} T_i \sim \Gamma(n, \lambda)$
- $2\lambda \sum_{i=1}^{n} T_i \sim \chi^2_{(2n)}$

MLE for exponential model (3)

In case of censoring, the likelihood can be expressed

$$\begin{aligned} L(\lambda) &= \prod_u f(y_i|\lambda) \prod_c S(y_i|\lambda) \\ &= \prod_u \lambda \exp(-\lambda y_i) \prod_c \exp(-\lambda y_i) \\ &= \lambda^{n_u} \exp\left(-\lambda \sum_u y_i\right) \exp\left(-\lambda \sum_c y_i\right) \\ &= \lambda^{n_u} \exp\left(-\lambda \sum_{i=1}^{n} y_i\right) \end{aligned}$$

from where the log-likelihood is

$$\ln L(\lambda) = n_u \ln(\lambda) - \lambda \sum_{i=1}^{n} y_i$$

MLE for exponential model (4)

Now, the derivative of log-likelihood is

$$\frac{\partial \ln L(\lambda)}{\partial \lambda} = \frac{n_u}{\lambda} - \sum_{i=1}^{n} y_i,$$

which results in the following MLE for λ:

$$\hat{\lambda} = \frac{n_u}{\sum_{i=1}^{n} y_i}$$

MLE for exponential model (5)

The estimate for variance of λ can be found (recall the Fisher information matrix!) using the second derivative of log-likelihood:

$$\frac{\partial^2 \ln L(\lambda)}{\partial \lambda^2} = -\frac{n_u}{\lambda^2},$$

which implies

$$Var(\hat{\lambda}) \approx \frac{\hat{\lambda}^2}{n_u}$$

Now the confidence intervals can be constructed using normal approximation and obtained estimates $\hat{\lambda}$ and $Var\hat{\lambda}$

Sidenote: asymptotic distribution of ML estimates (1)

Let us recall that a d-dimensional parameter estimate $\hat{\theta}$ follows asymptotically multivariate normal distribution:

$$\hat{\theta} \overset{a}{\sim} MVN(\theta^*, I^{-1}(\theta^*)),$$

where

- θ^* is the true value of the parameter vector
- $I(\theta)$ is the $(d \times d)$ Fisher information matrix:

$$I(\theta) = \left(-E\left(\frac{\partial^2}{\partial\theta_j \partial\theta_k} \ln L(\theta)\right) \right)$$

Sidenote: asymptotic distribution of ML estimates (2)

In practice, we usually do not know the (theoretical) Fisher information matrix and have to approximate it by the observed information matrix evaluated at ML estimate $\hat{\theta}$:

$$i(\hat{\theta}) = \left(-\frac{\partial^2}{\partial\theta_j \partial\theta_k} \ln L(\theta) \right) \Big|_{\theta=\hat{\theta}}$$

Thus, for a scalar parameter θ (think, e.g., of exponential distribution example):

$$Var(\hat{\theta}) \approx \frac{1}{i(\hat{\theta})} = \left(-\frac{\partial^2}{\partial\theta^2} \ln L(\theta) \Big|_{\theta=\hat{\theta}} \right)^{-1}$$