

## Likelihood Function for Censored Data

Suppose we have  $n$  units, with unit  $i$  observed for a time  $t_i$ . If the unit died at  $t_i$ , its contribution to the likelihood function (under non-informative censoring) is

$$L_i = f(t_i) = S(t_i)\lambda(t_i)$$

If the unit is still alive at  $t_i$ , all we know under non-informative censoring is that the lifetime exceeds  $t_i$ . The probability of this event is

$$L_i = S(t_i),$$

which becomes the contribution of a censored observation to the likelihood.

Letting  $d_i$  be a death indicator, we have

$$L = \prod_{i=1}^n L_i = \prod_{i=1}^n \lambda(t_i)^{d_i} S(t_i).$$

Taking logs, we have

$$\log L = \sum_{i=1}^n \{d_i \log \lambda(t_i) - \Lambda(t_i)\}.$$

Suppose we have exponentially distributed survival times, so that  $\lambda(t) = \lambda$  for all  $t$ . Then, the log likelihood follows the form

$$\log L = \sum_{i=1}^n \{d_i \log \lambda - \lambda t_i\}.$$

Letting  $D = \sum d_i$  be the total number of deaths and  $T = \sum t_i$  be the total time at risk, we have

$$\log L = D \log \lambda - \lambda T.$$

Differentiating this with respect to  $\lambda$ , the score function is

$$u(\lambda) = D/\lambda - T.$$

Setting this equal to 0, we obtain the maximum likelihood estimate

$$\hat{\lambda} = D/T,$$

which is simply the total number of deaths divided by the time at risk.

The observed information is minus the second derivative of the score, which is  $I(\lambda) = D/\lambda^2$ . Taking the inverse and plugging in the mle, we have

$$\widehat{\text{var}}(\hat{\lambda}) = D/T^2.$$

A useful observation is that the log-likelihood for exponential survival data is exactly the same (up to a proportionality constant) as the likelihood that would have been obtained by treating  $D$  as a Poisson random variable with mean  $\lambda T$ .

This will be useful in extending the methods to account for predictors by defining Poisson generalized linear models.

Until this point, we have focused on the homogeneous case where everyone has the same survival distribution. In most applications, interest will focus on inference on predictors

# Survival Modeling

## Approach I. Accelerated Life Models

We briefly discussed the Cox proportional hazards model earlier in the course. The Cox model assumes multiplicative covariate effects on the hazard function.

Alternative:

Let  $T_i$  denote the (possibly unobserved) survival time for subject  $i$ .

Since  $T_i$  is positive, we might consider the model:

$$\log T_i = \mathbf{x}'_i \boldsymbol{\beta} + \epsilon_i,$$

where  $\epsilon_i$  is a suitable error term

Exponentiating, we have

$$T_i = \exp(\mathbf{x}'_i \boldsymbol{\beta}) T_{0i},$$

where  $T_{0i}$  is the exponentiated error term and we let  $\gamma_i = \exp(\mathbf{x}'_i \boldsymbol{\beta})$  as shorthand.

Let  $S_0(t)$  denote the survival function in a reference group (group zero).

Let  $S_1(t) = S_0(t/\gamma)$ , which implies that a member of group one will be alive at age  $t$  with the same probability that a member of group zero will be alive at age  $t/\gamma$ .

For  $\gamma = 2$ , this would be half the age, so the probability that a member of group one would be alive at age 40 would be the same as the probability that a member of group zero would be alive at age 20.

Therefore, model of this time as known as **Accelerated Failure Time (AFT)** models.

Different kinds of parametric AFT models are obtained by assuming different error distributions.

For example, if the  $\epsilon_i \sim N(0, \sigma^2)$ , then we have a log-normal model for  $T_i$ . For censored data, this model is known (primarily in the economics literature) as the **Tobit** model.

Alternatively, if the  $\epsilon_i$  have an extreme value distribution with density

$$f(\epsilon) = \exp \{ \epsilon - \exp(\epsilon) \},$$

then  $T_{0i}$  has an exponential distribution and we obtain the exponential regression model, where  $T_i$  is exponential with hazard  $\lambda_i$  satisfying the log-linear model

$$\log \lambda_i = \mathbf{x}'_i \boldsymbol{\beta}.$$

**Parametric** accelerated failure time models, are just standard linear regression models applied to the log of the survival times. The only technical hurdle, which is typically trivial to deal with in the parametric case, is the censoring.

## Parametric Proportional Hazards Models

Recall that the proportional hazards model can be expressed as:

$$\lambda_i(t; \mathbf{x}_i) = \lambda_0(t) \exp(\mathbf{x}'_i \boldsymbol{\beta}).$$

By making different parametric assumptions on the baseline hazard, we can formulate different kinds of proportional hazards models.

The simplest case is to assume exponentially distributed survival times in the baseline group, which implies  $\lambda_0(t) = \lambda_0$  and hence

$$\lambda_i(t; \mathbf{x}_i) = \lambda_0 \exp(\mathbf{x}'_i \boldsymbol{\beta}).$$

Note that this model is both a proportional hazards model and an accelerated failure time model.

The only other case where the two families coincide is when the baseline survival times follow a Weibull distribution,

$$S(t) = \exp \{ - (\lambda t)^p \},$$

which results in the hazard function

$$\lambda(t) = p\lambda(\lambda t)^{p-1},$$

for parameters  $\lambda > 0$  and  $p > 0$ .

If  $p = 1$ , then the Weibull model reduces to the exponential model and the hazard is constant over time.

If  $p > 1$ , then the risk increases over time

If  $p < 1$ , then the risk decreases over time

## Other Properties of Weibull model:

- The logarithm of the hazard is a linear function of log time with slope  $p - 1$ ,

$$\log \lambda(t) = \log p + p \log \lambda + (p - 1) \log t.$$

- If the baseline survival distribution is Weibull, then multiplying the hazard by a constant results in a Weibull distribution. For example, if

$$\lambda_0(t) = p\lambda(\lambda t)^{p-1},$$

then, for  $\gamma_i = \exp(\mathbf{x}'_i \boldsymbol{\beta})$ , we have

$$\lambda_i(t; \mathbf{x}_i) = \lambda_0(t) \gamma_i = p(\lambda \gamma_i^{1/p})(\lambda \gamma_i^{1/p} t)^{p-1}$$

and hence the Weibull family is closed under proportional hazards

- If the baseline survival distribution is Weibull and we modify time by a multiplicative constant in the AFT model, the resulting distribution is still a Weibull, so the family is also closed under accelerated failure times.

## Strategies for Model Fitting

1. The most straightforward approach is to assume a parametric model for the baseline hazard  $\lambda_0(t)$  and then proceed with maximum likelihood or posterior computation. Common choices of parametric models include the exponential, Weibull, gamma, log-normal and generalized F distributions.
2. Another strategy is to use a flexible model, where we make mild assumptions about the baseline hazard  $\lambda_0(t)$ . For example, a common approach is to assume constant hazards within pre-specified time intervals, resulting in a piecewise exponential model.
3. A final strategy is to follow a non-parametric approach in which the baseline hazard  $\lambda_0(t)$  is left completely unspecified. This approach relies on the partial likelihood proposed by Cox (1972). From a Bayesian perspective, one can potentially choose a flexible prior for  $\lambda_0(t)$  (e.g., by using a Gamma process).

## Piecewise Exponential for Baseline Hazards in Cox Proportional Hazards Model

Suppose that we partition time into  $J$  intervals with cutpoints

$$0 = \tau_0 < \tau_1 < \dots < \tau_J = \infty,$$

with the  $j$ th interval defined as  $[\tau_{j-1}, \tau_j)$ .

We assume that the baseline hazard is constant within each interval:

$$\lambda_0(t) = \lambda_j \quad \text{for } t \in [\tau_{j-1}, \tau_j),$$

and hence the baseline hazard is characterized using  $J$  parameters,  $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_J)'$ .

Clearly, we can choose a large number of cutpoints in order to approximate any baseline hazard function, though care should be taken to avoid choosing too many intervals since we may not have enough data.

Introducing predictors,  $\mathbf{x}_i$ , the hazard for the  $i$ th person in the  $j$ th interval is

$$\lambda_{ij} = \lambda_j \exp(\mathbf{x}'_i \boldsymbol{\beta}),$$

and taking the logarithm we have

$$\log \lambda_{ij} = \alpha_j + \mathbf{x}'_i \boldsymbol{\beta},$$

where  $\alpha_j = \log \lambda_j$  and it is trivial to incorporate time-dependent covariates and coefficients.

This model can be shown to be equivalent to a Poisson regression model

Let  $t_{ij}$  denote the time spent by individual  $i$  in interval  $j$

Let  $d_{ij} = 1$  if individual  $i$  dies in interval  $j$  and  $d_{ij} = 0$  otherwise

Assuming that  $d_{ij} \sim \text{Poisson}(t_{ij}\lambda_{ij})$  results in the correct likelihood for estimating  $\boldsymbol{\lambda}$  and  $\boldsymbol{\beta}$  (Show as homework exercise)

Hence, we can simply re-format the data and fit a Poisson log linear model.