PH 716 Applied Survival Analysis

Part IV: Accelerated Failure Time Model

Zhiyang Zhou (zhou67@uwm.edu, zhiyanggeezhou.github.io)

2024/Feb/14 22:56:30

Assumptions

- T_i are independent across i
  - NO longer assumed to share the identical distribution
  - i.e., "personalized" or "individualized"
- log-linear model: $\ln T_i = \beta_0 + \sum_{j=1}^p x_{ij}\beta_j + \sigma\varepsilon_i$
  - Unknown parameters $\sigma > 0$ and $\beta_j \in \mathbb{R}$
  - Error terms ε_i are iid
- Equiv. $T_i = \exp(\beta_0 + \varepsilon_i)\prod_{j=1}^p \exp(x_{ij}\beta_j)$
  - (Why is called "accelerated failure time model"?) The effect of covariates acts multiplicatively on the survival time and accelerates or decelerates the progress along the time axis.

Survival function

- If $\varepsilon_i \overset{iid}{\sim} N(0,1)$,
  - $S_{T_i}(t) = \Pr(\ln T_i > \ln t) = \Pr\{\varepsilon_i > \sigma^{-1}(\ln t - \beta_0 - \sum_{j=1}^p x_{ij}\beta_j)\} = 1 - \Phi\{\sigma^{-1}(\ln t - \beta_0 - \sum_{j=1}^p x_{ij}\beta_j)\}$
    - $\Phi(\cdot)$: the cdf of $N(0,1)$
  - i.e., $T_i \sim \text{log-normal}(\beta_0 + \sum_{j=1}^p x_{ij}\beta_j, \sigma^2)$
- If $\varepsilon_i \overset{iid}{\sim}$ the standard Gumbel distribution for minimum (i.e., $F_{\varepsilon_i}(\epsilon) = 1 - \exp(-\exp\epsilon)$),
  - P.S. $\min(X_1, X_2, \ldots, X_n) - \ln n \xrightarrow{d}$ standard Gumbel distribution (for minimum) as $n \to \infty$ if $X_i \overset{iid}{\sim} \exp(1)$
  - $S_{T_i}(t) = \Pr\{\varepsilon_i > \sigma^{-1}(\ln t - \beta_0 - \sum_{j=1}^p x_{ij}\beta_j)\} = 1 - F_{\varepsilon_i}\{\sigma^{-1}(\ln t - \beta_0 - \sum_{j=1}^p x_{ij}\beta_j)\} = \exp[-t^{1/\sigma}\exp\{-(\beta_0 + \sum_{j=1}^p x_{ij}\beta_j)/\sigma\}] = \exp[-\{t/\exp(\beta_0 + \sum_{j=1}^p x_{ij}\beta_j)\}^{1/\sigma}]$
  - i.e., $T_i \sim$ Weibull with $1/\sigma$ as the "shape" and $\exp(\beta_0 + \sum_{j=1}^p x_{ij}\beta_j)$ as the "scale"
    - Widely used in practice, with a hazard descending or ascending with respect to t
    - Specifically, $T \sim$ exponential if $\sigma = 1$, with a hazard constant with respect to hazard

Likelihood principles (for uncensored data)

- Observed $T_1 = t_1, \ldots, T_n = t_n$
- Joint density of $\mathbf{T} = [T_1, \ldots, T_n]^\top$ evaluated at $[t_1, \ldots, t_n]^\top$: $f_{\mathbf{T}}(t_1, \ldots, t_n; \boldsymbol{\theta})$
  - $\boldsymbol{\theta}$: a p-vector of unknown parameters
- Observed-data likelihood $L(\boldsymbol{\theta}) = f_{\mathbf{T}}(t_1, \ldots, t_n; \boldsymbol{\theta})$
  - Taken as a function of $\boldsymbol{\theta}$
  - $L(\boldsymbol{\theta}) = \prod_{i=1}^n f_{T_i}(t_i; \boldsymbol{\theta})$ if T_i is independent across i
- Maximum likelihood estimator (MLE): $\hat{\boldsymbol{\theta}}_{\text{ML}} = \max_{\boldsymbol{\theta}} L(\boldsymbol{\theta}) = \max_{\boldsymbol{\theta}} \ell(\boldsymbol{\theta})$

- $\ell(\boldsymbol{\theta}) = \ln L(\boldsymbol{\theta})$
- A closed-form solution for $\hat{\boldsymbol{\theta}}_{\mathrm{ML}}$ usually not available
  - Resorting to numerical optimization techniques, e.g., Newton's method

- Confidence interval (CI) of $\boldsymbol{\theta}$
  - $\hat{\boldsymbol{\theta}}_{\mathrm{ML}} \approx N(\boldsymbol{\theta}, I(\hat{\boldsymbol{\theta}}_{\mathrm{ML}})^{-1})$ for iid T_i
    - Because $\sqrt{n}(\hat{\boldsymbol{\theta}}_{\mathrm{ML}} - \boldsymbol{\theta}) \xrightarrow{d} N(0, nI(\boldsymbol{\theta})^{-1})$ for iid T_i
    - Fisher information (the expectation of Hessian matrix of $\ell(\boldsymbol{\theta})$): $I(\boldsymbol{\theta}) = -\mathrm{E}\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \approx -\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top}$

- Likelihood ratio test (LRT)
  - H_0 vs H_1
  - Test statistic: $-2\ln\frac{L(\hat{\boldsymbol{\theta}}_{\mathrm{ML},H_0})}{L(\hat{\boldsymbol{\theta}}_{\mathrm{ML}})} = 2\{\ell(\hat{\boldsymbol{\theta}}_{\mathrm{ML}}) - \ell(\hat{\boldsymbol{\theta}}_{\mathrm{ML},H_0})\}$
    - $\hat{\boldsymbol{\theta}}_{\mathrm{ML},H_0}$: the (constrained) MLE under H_0
    - $\hat{\boldsymbol{\theta}}_{\mathrm{ML}}$: the MLE under $H_0 \bigcup H_1$
  - Reject H_0 if the value of $-2\ln\frac{L(\hat{\boldsymbol{\theta}}_{\mathrm{ML},H_0})}{L(\hat{\boldsymbol{\theta}}_{\mathrm{ML}})}$ is over $\chi^2_{p,1-\alpha}$
    - $\chi^2_{p,1-\alpha}$: the $1-\alpha$ quantile of $\chi^2(p)$
    - Because $-2\ln\frac{L(\hat{\boldsymbol{\theta}}_{\mathrm{ML},H_0})}{L(\hat{\boldsymbol{\theta}}_{\mathrm{ML}})} \approx \chi^2(p)$
      - p: the difference of free parameters with and without H_0

Ex. 4.1 (uncensored exponential-distributed observations)

- The following $n = 10$ iid failure times are assumed to arise from $\exp(\lambda)$, i.e., $f_T(t) = \lambda \exp(-\lambda t)$.

i	1	2	3	4	5	6	7	8	9	10
t_i	10	12	8	7	2	4	15	6	5	19

- Computing MLE
  1. $f(t_i;\lambda) = \lambda\exp(-\lambda t_i)$, $i = 1, \ldots, 10$
  2. $L(\lambda) = \prod_{i=1}^{10} f(t_i;\lambda) = \lambda^{10}\exp(-\lambda\sum_{i=1}^{10} t_i)$
  3. $\ell(\lambda) = \sum_{i=1}^{10}\ln f(t_i;\lambda) = 10 \times (\ln\lambda) - \lambda\sum_{i=1}^{10} t_i$
     - $\ell'(\lambda) = 10/\lambda - \sum_{i=1}^{10} t_i$
  4. $\hat{\lambda}_{\mathrm{ML}} = \arg\max_{\lambda\in(0,\infty)} \ell(\lambda)$
     - $\hat{\lambda}_{\mathrm{ML}} = 10/\sum_{i=1}^{10} t_i = 10/88$ by solving the score equation $\ell'(\lambda) = 0$
- 95% CI of λ
  1. $\ell''(\lambda) = -10/\lambda^2$
  2. $I(\lambda) = -\mathrm{E}\ell''(\lambda) = 10/\lambda^2$
  3. 95% CI of λ: $\hat{\lambda}_{\mathrm{ML}} \pm 1.96 \times I(\hat{\lambda}_{\mathrm{ML}})^{-1/2}$, i.e., $10/88 \pm 1.96 \times \sqrt{10}/88$
     - Because $\lambda \approx N(\hat{\lambda}_{\mathrm{ML}}, I(\hat{\lambda}_{\mathrm{ML}})^{-1}) = N(10/88, 10/88^2)$
  4. Interpretation
- Testing $H_0 : \lambda = .1$ vs $H_1 : \lambda \neq .1$ at the significance level $\alpha = .05$
  1. Test statistic: $2\{\ell(\hat{\lambda}_{\mathrm{ML}}) - \ell(\hat{\lambda}_{\mathrm{ML},H_0})\} \approx .16$
     - $\hat{\lambda}_{\mathrm{ML},H_0} = .1$
  2. Compare the value of test statistic with $\chi^2_{p,1-\alpha}$
     - $\chi^2_{p,1-\alpha} \approx 3.84$ with $p = 1$
  3. Or, the p-value may be calculated via `pchisq(.16, 1)`
  4. Conclusion

Likelihood principles (for right-censored data)

- Observed $\widetilde{T}_i = \tilde{t}_i$ and $\Delta_i = \delta_i$ (event indicator),
  - $\widetilde{T}_i$: the smaller one between T_i (event time) and C_i (right-cencoring time)

  - Assuming the independence across i
  - Assuming the independent and noninformative censoring, i.e.,
    * $T_i \perp C_i$ (conditional on covariates)
    * $S_{T_i}(t \mid \boldsymbol{\theta})$ and $S_{C_i}(t \mid \boldsymbol{\eta})$ have NO common parameter
- Joint density of $\widetilde{T}_i$ and Δ_i: $f_{\widetilde{T}_i,\Delta_i}(\tilde{t}_i, \delta_i) =$
  - $f_{T_i}(\tilde{t}_i \mid \boldsymbol{\theta}) S_{C_i}(\tilde{t}_i \mid \boldsymbol{\eta})$ if $\delta_i = 1$
  - $S_{T_i}(\tilde{t}_i \mid \boldsymbol{\theta}) f_{C_i}(\tilde{t}_i \mid \boldsymbol{\eta})$ if $\delta_i = 0$
    * Because
      · $\Pr(\widetilde{T}_i > t, \Delta_i = 1) = \Pr(C_i \geq T_i, T_i > t) = \int_t^\infty \Pr(C_i \geq u, T_i = u) \mathrm{d}u = \int_t^\infty S_{C_i}(u \mid \boldsymbol{\eta}) f_{T_i}(u \mid \boldsymbol{\theta}) \mathrm{d}u$
      · $\Pr(\widetilde{T}_i > t, \Delta_i = 0) = \Pr(T_i \geq C_i, C_i > t) = \int_t^\infty \Pr(T_i \geq u, C_i = u) \mathrm{d}u = \int_t^\infty S_{T_i}(u \mid \boldsymbol{\theta}) f_{C_i}(u \mid \eta) \mathrm{d}u$
- Observed-data likelihood: $L(\boldsymbol{\theta}, \boldsymbol{\eta}) = \prod_{i=1}^n f_{\widetilde{T}_i,\Delta_i}(\tilde{t}_i, \delta_i) = \prod_{i=1}^n \{f_{T_i}(\tilde{t} \mid \boldsymbol{\theta}) S_{C_i}(\tilde{t} \mid \boldsymbol{\eta})\}^{\delta_i} \{S_{T_i}(\tilde{t} \mid \boldsymbol{\theta}) f_{C_i}(\tilde{t} \mid \boldsymbol{\eta})\}^{1-\delta_i}$
  - Reducing to $\prod_{i=1}^n f_{T_i}(\tilde{t}_i \mid \boldsymbol{\theta})^{\delta_i} S_{T_i}(\tilde{t}_i \mid \boldsymbol{\theta})^{1-\delta_i} = \prod_{i=1}^n \lambda_{T_i}(\tilde{t}_i \mid \boldsymbol{\theta})^{\delta_i} S_{T_i}(\tilde{t}_i \mid \boldsymbol{\theta})$ if we are only concerned about the MLE of $\boldsymbol{\theta}$

Likelihood principles (for general censored data)

- Assuming the independence across i and independence and noninformative censoring
- Observed-data likelihood:

$$\prod_{i \in \mathfrak{D}} f_{T_i}(\tilde{t}_i) \prod_{i \in \mathfrak{R}} S_{T_i}(\tilde{t}_i) \prod_{i \in \mathfrak{L}} \{1 - S_{T_i}(\tilde{t}_i)\} \prod_{i \in \mathfrak{I}} \{S_{T_i}(\tilde{t}_{iL}) - S_{T_i}(\tilde{t}_{iR})\}$$

  - $\mathfrak{D}$: the set of **uncensored** subjects
  - $\mathfrak{R}$: the set of **right-censored** subjects
  - $\mathfrak{L}$ the set of **left-censored** subjects
  - $\mathfrak{I}$: the set of **interval-censored** subjects

Exponential regression for right-censored data

- Observed $\{\widetilde{T}_i = \tilde{t}_i, \Delta_i = \delta_i, x_{i1}, \ldots, x_{ip}\}$
  - $\widetilde{T}_i = \min(T_i, C_i)$
  - $\Delta_i = 1$ if $\widetilde{T}_i = T_i$ and zero otherwise
- Assuming independent and non-informative censoring
- Assuming $T_i \sim \exp(\lambda_i)$
  - $\lambda_i = \lambda(x_{i1}, \ldots, x_{ip} \mid \boldsymbol{\beta}) = \exp(\beta_0 + \sum_{j=1}^p x_{ij}\beta_j)$ (Why using the exponential form?)
  - Two distinct forms of parameterization used by different `R` functions
    * (`survival::survreg`) density $f_{T_i}(t) = \lambda_i^{-1} \exp(-t/\lambda_i)$, hazard rate $\lambda_{T_i}(t \mid \boldsymbol{\beta}) = 1/\lambda_i$, and survival function $S_{T_i}(t \mid \boldsymbol{\beta}) = \exp(-t/\lambda_i)$
    * (`flexsurv::flexsurvreg`) density $f_{T_i}(t) = \lambda_i \exp(-\lambda_i t)$, hazard rate $\lambda_{T_i}(t \mid \boldsymbol{\beta}) = \lambda_i$, and survival function $S_{T_i}(t \mid \boldsymbol{\beta}) = \exp(-\lambda_i t)$
- Likelihood function $L(\boldsymbol{\beta}) = \prod_i \lambda_{T_i}(\tilde{t}_i \mid \boldsymbol{\beta})^{\delta_i} S_{T_i}(\tilde{t}_i \mid \boldsymbol{\beta})$
  - $\boldsymbol{\beta} = [\beta_0, \beta_1, \ldots, \beta_p]^\top$
- Log-likelihood function $\ell(\boldsymbol{\beta}) = \sum_i \{\delta_i \ln \lambda_{T_i}(\tilde{t}_i \mid \boldsymbol{\beta}) + \ln S_{T_i}(\tilde{t}_i \mid \boldsymbol{\beta})\}$
  - Score function $U(\boldsymbol{\beta}) = \frac{\partial \ell(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = [\frac{\partial \ell(\boldsymbol{\beta})}{\partial \beta_0}, \frac{\partial \ell(\boldsymbol{\beta})}{\partial \beta_1}, \ldots, \frac{\partial \ell(\boldsymbol{\beta})}{\partial \beta_p}]^\top$
    * In general no closed-form for the solution of score equations $U(\boldsymbol{\beta}) = 0$
  - Fisher information $I(\boldsymbol{\beta}) = -\mathrm{E} \frac{\partial \ell(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top}$

    * $\frac{\partial \ell(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} = [\frac{\partial \ell(\boldsymbol{\beta})}{\partial \beta_i \partial \beta_j}]_{(p+1)\times(p+1)}$
  - Newton's method (for maximization)
    1. Start with an initial guess $\hat{\boldsymbol{\beta}}_{(0)}$
    2. Update the current estimate with $\hat{\boldsymbol{\beta}}_{(k+1)} = \hat{\boldsymbol{\beta}}_{(k)} + I(\hat{\boldsymbol{\beta}}_{(k)})^{-1} U(\boldsymbol{\beta}_{(k)})$ until $\hat{\boldsymbol{\beta}}_{(k)}$ and $\hat{\boldsymbol{\beta}}_{(k+1)}$ are close enough

- Interpretation of β_j, $j \neq 0$ (after fixing all covariates other than the jth one)
  - (`survival::survreg`) one-unit increase in the jth covariate change the survival time by $(\exp(\beta_j)-1) \times 100\%$ OR the hazard ratio (HR, i.e., the ratio of hazard rates) associated with a one-unit increase in the jth covariate is $\exp(-\beta_j)$
  - (`flexsurv::flexsurvreg`) one-unit increase in the jth covariate change the survival time by $(\exp(-\beta_j) - 1) \times 100\%$ OR the HR associated with a one-unit increase in the jth covariate is $\exp(\beta_j)$.

- Ex 4.3. ([DM] pp.147): The purpose of Steinberg et al. (2009) was to evaluate extended duration of a triple-medication combination versus therapy with the nicotine patch alone in smokers with medical illnesses.

```
head(asaur::pharmacoSmoking)
data.ex42 = asaur::pharmacoSmoking
data.ex42 = data.ex42[data.ex42$ttr != 0,] # ttr=0 not allowed in AFT models
is.factor(data.ex42$grp)
aft.ex42.1 = survival::survreg(
  survival::Surv(ttr, relapse) ~ grp,
  data = data.ex42,
  dist="weibull",
  scale = 1
)
summary(aft.ex42.1)
# Or
aft.ex42.2 = survival::survreg(
  survival::Surv(ttr, relapse) ~ grp,
  data = data.ex42,
  dist="exponential"
)
summary(aft.ex42.2)
# Or using flexsurv::flexsurvreg
aft.ex42.3 = flexsurv::flexsurvreg(
  survival::Surv(ttr, relapse) ~ grp,
  data = data.ex42,
  dist = "exponential"
)
aft.ex42.3
survminer::ggflexsurvplot(aft.ex42.3, data=data.ex42[data.ex42$grp=='patchOnly',])

# prediction for grp='combination'
shape = 1/aft.ex42.1$scale
scale = unname(exp(aft.ex42.1$coefficients[1])) # scale
(ET = scale*gamma(1+1/shape)) # expectation of T
(medT = scale*log(2)^(1/shape)) # median of T
surv.fun = function(t){ # survival function
  return(
    1-pweibull(t, shape = shape, scale = scale)
```

```r
  )
}
curve(surv.fun, from = 0, to = 1e3) # plot the survival curve for grp='combination'
```

- Interpretation of β_1
  - Campared to the "triple-medication-combination", the "patch-alone" therapy change the survival time by $(\exp(-0.723) - 1) \times 100\%$, i.e., reduce the survival time by 51.5%. OR, the HR of the "patch-alone" therapy to the "triple-medication-combination" is 2.06, i.e., the hazard of the "patch-alone" therapy is twice as high as the "triple-medication-combination".

Weibull regression for right-censored data

- Observed $\{\widetilde{T}_i = \tilde{t}_i, \Delta_i = \delta_i, x_{i1}, \ldots, x_{ip}\}$
- Assuming the independence across i and the independent and non-informative censoring
- Recall that if $\ln T_i = \beta_0 + \sum_{j=1}^p x_{ij}\beta_j + \sigma\varepsilon_i$ and $\varepsilon_i \overset{iid}{\sim} F_{\varepsilon_i}(\epsilon) = 1 - \exp(-\exp\epsilon)$, then
  - $S_{T_i}(t) = \exp[-\{t/\exp(\beta_0 + \sum_{j=1}^p x_{ij}\beta_j)\}^{1/\sigma}]$
    * This parameterization is honored by both `survival::survreg` and `flexsurv::flexsurvreg`
    * BUT there is an inconsistency when the two `R` functions name some parameters
      · β_0: called "(Intercept)" in the output of `survival::survreg`
      · $\exp(\beta_0)$: called "scale" in the output of `flexsurv::flexsurvreg`
      · σ: called "Scale" in the output of `survival::survreg`
      · $1/\sigma$: called "shape" in the output of `flexsurv::flexsurvreg`
- Interpretation of β_j, $j \neq 0$ (after fixing all covariates other than the jth one) one-unit increase in the jth covariate change the survival time by $(\exp(\beta_j) - 1) \times 100\%$
  - Inconvenient to Interpret from the perspective of HR (why?)
- Ex 4.4. (revisit to Ex. 4.3.)

```r
head(asaur::pharmacoSmoking)
data.ex44 = asaur::pharmacoSmoking
data.ex44 = data.ex44[data.ex44$ttr != 0,] # ttr=0 not allowed in AFT models
is.factor(data.ex44$grp)
aft.ex44.1 = survival::survreg(
  survival::Surv(ttr, relapse) ~ grp,
  data = data.ex44,
  dist="weibull"
)
summary(aft.ex44.1)
# OR using flexsurv::flexsurvreg
aft.ex44.2 = flexsurv::flexsurvreg(
  survival::Surv(ttr, relapse) ~ grp,
  data = data.ex44,
  dist = "weibull"
)
aft.ex44.2
survminer::ggflexsurvplot(aft.ex44.2)

# prediction for grp='combination'
shape = 1/aft.ex42$scale
scale = unname(exp(aft.ex42$coefficients[1])) # scale
(ET = scale*gamma(1+1/shape)) # expectation of T
(medT = scale*log(2)^(1/shape)) # median of T
```

```r
surv.fun = function(t){ # survival function
  return(
    1-pweibull(t, shape = shape, scale = scale)
  )
}
curve(surv.fun, from = 0, to = 1e3) # plot the survival curve for grp='combination'
```

- Interpretation of β_1
  - Campared to the "triple-medication-combination", the "patch-alone" therapy reduce the survival time by $1 - \exp(-1.0325) = 64.4\%$.

log-normal regression for right-censored data

- Observed $\{\widetilde{T}_i = \tilde{t}_i, \Delta_i = \delta_i, x_{i1}, \ldots, x_{ip}\}$
- Assuming the independence across i and the independent and non-informative censoring
- Recall that if $\ln T_i = \beta_0 + \sum_{j=1}^{p} x_{ij}\beta_j + \sigma\varepsilon_i$ and $\varepsilon_i \overset{\text{iid}}{\sim} N(0,1)$, then
  - $S_{T_i}(t) = 1 - \Phi\{\sigma^{-1}(\ln t - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)\}$
    * This parameterization is honored by both `survival::survreg` and `flexsurv::flexsurvreg`
    * BUT there is an inconsistency when the two `R` functions name some parameters
      · β_0: called "(Intercept)" in the output of `survival::survreg` but "meanlog" in the output of `flexsurv::flexsurvreg`
      · σ: called "Scale" in the output of `survival::survreg` but "sdlog" in the output of `flexsurv::flexsurvreg`
- Interpretation of β_j, $j \neq 0$ (after fixing all covariates other than the jth one) one-unit increase in the jth covariate change the survival time by $(\exp(\beta_j) - 1) \times 100\%$

- Ex. 4.5. Revisit the data of bladder cancer recurrences which contain three treatment arms for 118 subjects.

```r
data.ex45 = survival::bladder1[
  complete.cases(
    survival::bladder1[,c('id', 'treatment', 'start', 'stop', 'status')]
  ),
  c('id', 'treatment', 'start', 'stop', 'status')
]
data.ex45$status = 1*(data.ex45$status %in% c(1,2,3)) # merging status 1, 2,3
data.ex45$tte = data.ex45$stop - data.ex45$start
data.ex45 = data.ex45[data.ex45$tte != 0,] # ttr=0 not allowed in AFT models
is.factor(data.ex45$treatment)
aft.ex45.1 = survival::survreg(
  survival::Surv(tte, status) ~ treatment,
  data = data.ex45,
  dist="lognormal"
)
summary(aft.ex45.1)
# OR using flexsurv::flexsurvreg
aft.ex45.2 = flexsurv::flexsurvreg(
  survival::Surv(tte, status) ~ treatment,
  data = data.ex45,
  dist = "lognormal"
)
```

```r
aft.ex45.2
survminer::ggflexsurvplot(aft.ex45.2)

# prediction for treatment='pyridoxine'
sigma = aft.ex43$scale
mu = sum(aft.ex43$coefficients[1:2]) # scale
(ET = exp(mu+sigma^2/2)) # expectation of T
(medT = exp(mu)) # median of T
surv.fun = function(t){ # survival function for treatment='pyridoxine'
  return(
    1-pnorm((log(t)-mu)/sigma)
  )
}
curve(surv.fun, from = 0, to = 1e2) # plot the survival curve
```

Pros and cons

- Likelihood principles
  - Clear pathway
  - Exact inference only available for selected (and really simple) cases, i.e., approximations usually employed
  - MLE considered (approximately) the most efficient in regular cases
  - LRT optimal for simple cases but well accepted even in complex cases
- AFT model
  - Easy to interprete coeffcients: effects on the failure time directly
  - Distribution assumptions may be too strong
  - Can handle non-standard situations such interval censoring
  - Yields estimates of functions like hazard and survival for all times (even beyond the scope of follow-up)
    - Also dangerous since the extrapolation beyond the observed data range is not reliable