PH 716 Applied Survival Analysis

Part 4: Accelerated Failure Time Model

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

2026/02/22 18:44:44

Motivating example 1: accelerated life testing of motorettes

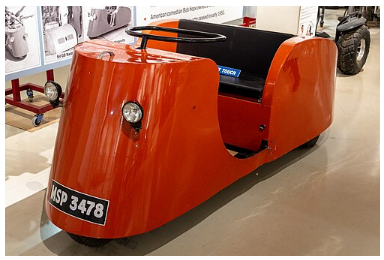

Figure 1: Motorette in the Gedee Car Museum, Coimbatore, TN, India (From Wikipedia)

We often cannot wait years to observe failures under normal use. Instead, we run accelerated life tests by increasing stress (e.g., temperature) to make failures happen sooner, then model how stress changes the time-to-failure distribution.

The `MASS::motors` dataset contains an accelerated life testing of 10 motorettes at each of four temperatures. The recorded `time` is the time (hours) to failure, or censoring at 8064 hours.

```
head(MASS::motors)
```

```
##   temp time cens
## 1  150 8064    0
## 2  150 8064    0
## 3  150 8064    0
## 4  150 8064    0
## 5  150 8064    0
## 6  150 8064    0
```

Scientific questions:

- Does higher temperature accelerate failure time (shorten lifetime)?
- If so, by how much?

Motivating example 2: capacitor failure under different stress levels

Capacitors are tested under various stress conditions (such as voltage or temperature) to study time to failure.

The dataset `survival::capacitor` contains failure times of capacitors tested under different temperature and voltage levels. Some units did not fail during the testing period, resulting in right-censored observations.

```
head(survival::capacitor)
```

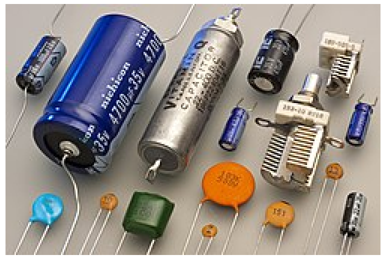

Figure 2: Capacitors (From Wikipedia)

```
##   temperature voltage fail time status
## 1         170     200   NA  439      1
## 2         170     200   NA  904      1
## 3         170     200   NA 1092      1
## 4         170     200   NA 1105      1
## 5         170     250   NA  572      1
## 6         170     250   NA  690      1
```

Scientific questions:

- Does higher voltage accelerate failure time?
- By what factor does lifetime change under increased stress?

Assumptions

- 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 $\varepsilon_i \overset{\text{iid}}{\sim} N(0,1)$,
  - $S_{T_i}(t) = \Pr(\ln T_i > \ln t) = \Pr\{\varepsilon_i > \sigma^{-1}(\ln t - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)\} = 1 - \Phi\{\sigma^{-1}(\ln t - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)\}$
    - $\Phi(\cdot)$: 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 $\varepsilon_i \overset{\text{iid}}{\sim}$ 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 with a hazard constant if $\sigma = 1$

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 $L(\boldsymbol{\theta}) = f_{\mathbf{T}}(t_1, \dots, t_n; \boldsymbol{\theta})$
  - 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 $\hat{\boldsymbol{\theta}}_{\text{ML}}$ usually not available
    * Resorting to numerical optimization techniques, e.g., Newton's method

- Confidence interval (CI) of $\boldsymbol{\theta}$
  - $\hat{\boldsymbol{\theta}}_{\text{ML}} \approx N(\boldsymbol{\theta}, I(\hat{\boldsymbol{\theta}}_{\text{ML}})^{-1})$ for iid T_i
    * Because $\sqrt{n}(\hat{\boldsymbol{\theta}}_{\text{ML}} - \boldsymbol{\theta}) \xrightarrow{d} N(0, nI(\boldsymbol{\theta})^{-1})$ 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: $-2 \ln \frac{L(\hat{\boldsymbol{\theta}}_{\text{ML},H_0})}{L(\hat{\boldsymbol{\theta}}_{\text{ML}})} = 2\{\ell(\hat{\boldsymbol{\theta}}_{\text{ML}}) - \ell(\hat{\boldsymbol{\theta}}_{\text{ML},H_0})\}$
    * $\hat{\boldsymbol{\theta}}_{\text{ML},H_0}$: the (constrained) MLE under H_0
    * $\hat{\boldsymbol{\theta}}_{\text{ML}}$: the MLE under $H_0 \bigcup H_1$
  - Reject H_0 if the value of $-2 \ln \frac{L(\hat{\boldsymbol{\theta}}_{\text{ML},H_0})}{L(\hat{\boldsymbol{\theta}}_{\text{ML}})}$ is over $\chi^2_{p,1-\alpha}$
    * $\chi^2_{p,1-\alpha}$: the $1-\alpha$ quantile of $\chi^2(p)$
    * Because $-2 \ln \frac{L(\hat{\boldsymbol{\theta}}_{\text{ML},H_0})}{L(\hat{\boldsymbol{\theta}}_{\text{ML}})} \approx \chi^2(p)$
      · 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. $f(t_i; \lambda) = \lambda \exp(-\lambda t_i)$, $i = 1, \dots, 10$
  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. $\hat{\lambda}_{\text{ML}} = \arg\max_{\lambda \in (0,\infty)} \ell(\lambda)$
     - $\hat{\lambda}_{\text{ML}} = 10/\sum_{i=1}^{10} t_i = 10/88$ 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: $\hat{\lambda}_{\text{ML}} \pm 1.96 \times I(\hat{\lambda}_{\text{ML}})^{-1/2}$, i.e., $10/88 \pm 1.96 \times \sqrt{10}/88$
     - Because $\lambda \approx N(\hat{\lambda}_{\text{ML}}, I(\hat{\lambda}_{\text{ML}})^{-1}) = N(10/88, 10/88^2)$
  4. Interpretation
- Testing $H_0 : \lambda = .1$ vs $H_1 : \lambda \neq .1$ at the significance level $\alpha = .05$
  1. Test statistic: $2\{\ell(\hat{\lambda}_{\text{ML}}) - \ell(\hat{\lambda}_{\text{ML},H_0})\} \approx .16$
     - $\hat{\lambda}_{\text{ML},H_0} = .1$
  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: $L(\boldsymbol{\theta}, \boldsymbol{\eta}) = \prod_{i=1}^n f_{\widetilde{T}_i,\Delta_i}(\tilde{t}_i, \delta_i) = \prod_{i=1}^n \{f_{T_i}(\tilde{t}_i \mid \boldsymbol{\theta}) S_{C_i}(\tilde{t}_i \mid \boldsymbol{\eta})\}^{\delta_i} \{S_{T_i}(\tilde{t}_i \mid \boldsymbol{\theta}) f_{C_i}(\tilde{t}_i \mid \boldsymbol{\eta})\}^{1-\delta_i}$
  - 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 if $\widetilde{T}_i = C_i$
- Assuming
  - Independence across i
  - Independent and non-informative censoring
  - $\ln T_i = \beta_0 + \sum_{j=1}^p x_{ij}\beta_j + \sigma\varepsilon_i$ with
    - $\varepsilon_i \overset{\text{iid}}{\sim} F_{\varepsilon_i}(\epsilon) = 1 - \exp(-\exp\epsilon)$
    - $\sigma = 1$
- Accordingly
  - $T_i = \exp(\varepsilon_i)\exp(\beta_0)\prod_{j=1}^p \exp(x_{ij}\beta_j)$
  - $S_{T_i}(t \mid \boldsymbol{\beta}) = \exp[-t/\exp(\beta_0 + \sum_{j=1}^p x_{ij}\beta_j)] = \exp\{-t\exp(-\beta_0 - \sum_{j=1}^p x_{ij}\beta_j)\}$ (as derived when introducing the log-linear model)
    - $\Rightarrow \lambda_{T_i}(t \mid \boldsymbol{\beta}) = \exp(-\beta_0 - \sum_{j=1}^p x_{ij}\beta_j)$

- 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 $I(\boldsymbol{\beta}) = -\mathrm{E}\frac{\partial \ell(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top}$
        - $\frac{\partial \ell(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} = [\frac{\partial \ell(\boldsymbol{\beta})}{\partial \beta_i \partial \beta_j}]_{(p+1)\times(p+1)}$
    - Maximization via, e.g, Newton's method
        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 parameters
    - β_0
        - $\exp(\beta_0)$: the baseline expected survival time
        - $\exp(-\beta_0)$: the baseline hazard rate
    - β_j, $j \neq 0$ (after fixing all covariates other than the jth one)
        - A one-unit increase in the jth covariate results in an expected survival time that is around $\exp(\beta_j)$ times the expected survival time at the baseline level.
        - A one-unit increase in the jth covariate results in a hazard rate that is $\exp(-\beta_j)$ times the hazard at the baseline level.
- Graphically check the correctness of model assumption
    1. Collect residuals $\ln T_i - \hat{\beta}_0 - \sum_j x_{ij}\hat{\beta}_j$ for uncensored subjects
    2. Compare residuals to a gumbel random sample via the Q-Q plot.
- Difference in the outputs of `R` functions
    - Due to different ways of parameterization
    - `survival::survreg`: "Intercept" (i.e., $\hat{\beta}_0$) and $\hat{\beta}_j$, $j = 1, \ldots, p$
    - `flexsurv::flexsurvreg`: "rate" (i.e., $\exp(-\hat{\beta}_0)$) and $-\hat{\beta}_j$, $j = 1, \ldots, p$

Ex 4.2. Revisit the accelerated life testing of 10 motorettes

```r
head(MASS::motors)
data.ex42 = MASS::motors
data.ex42 = data.ex42[data.ex42$time != 0,] # ttr=0 not allowed in AFT models
is.factor(data.ex42$temp)
data.ex42$temp = factor(data.ex42$temp)
aft.ex42.1 = survival::survreg(
  survival::Surv(time, cens) ~ temp,
  data = data.ex42,
  dist="weibull",
  scale = 1,
  x = T
)
summary(aft.ex42.1)
# Or
aft.ex42.2 = survival::survreg(
  survival::Surv(time, cens) ~ temp,
  data = data.ex42,
  dist="exponential"
)
summary(aft.ex42.2)
# Or using flexsurv::flexsurvreg
aft.ex42.3 = flexsurv::flexsurvreg(
```

```r
  survival::Surv(time, cens) ~ temp,
  data = data.ex42,
  dist = "exponential"
)
aft.ex42.3

# The survival time for temp=150
ET_tmp150 = unname(exp(aft.ex42.1$coefficients[1])) # temp=150 is the reference level
(medT = log(2)*ET_tmp150) # median of T; this formula working only for exp dist
surv.fun = function(t){ # survival function
  return(
    1-pexp(t, rate = 1/ET_tmp150)
  )
}
curve(surv.fun, from = 0, to = 1e3)

# The survival time for temp=220
ET_tmp220 = unname(exp( # expectation of T
  aft.ex42.1$coefficients[1] +
  aft.ex42.1$coefficients['temp220']
))
(medT = log(2)*ET_tmp220) # median of T; this formula working only for exp dist
surv.fun = function(t){ # survival function
  return(
    1-pexp(t, rate = 1/ET_tmp22)
  )
}
curve(surv.fun, from = 0, to = 1e3)

# Graphically check the correctness of exponential assumption
set.seed(718)
g.rnd = ordinal::rgumbel(10000,max = F) # gumbel random sample
lnTs.uncen = log(as.vector(data.ex42$time[data.ex42$cens==1]))
res = (
  lnTs.uncen - aft.ex42.1$x[data.ex42$cens==1,] %*% as.matrix(aft.ex42.1$coefficients)
)
qqplot(
  x = g.rnd,
  y = res,
  xlab = "Theoretical Quantiles",
  ylab = "Sample Quantiles"
)
qqline(res, distribution = function(p){ordinal::qgumbel(p,max=F)})
```

- Interpretation of $\hat{\beta}_1$
  - That means `temp=170` has an expected survival time that is around $\exp(\hat{\beta}_1)$ times that at `temp=150`;
  - Campared with `temp=150`, `temp=170` multiplies the expected survival time by $\exp(\hat{\beta}_1)$;
  - The hazard at `temp=170` is $\exp(-\hat{\beta}_1)$ times the hazard at `temp=150`.

Weibull regression for right-censored data

- Observed $\{\widetilde{T}_i = \tilde{t}_i, \Delta_i = \delta_i, x_{i1}, \ldots, x_{ip}\}$
- Assuming

- Independence across i
- Independent and non-informative censoring
- $\ln T_i = \beta_0 + \sum_{j=1}^p x_{ij}\beta_j + \sigma\varepsilon_i$ with
  - $\varepsilon_i \overset{\text{iid}}{\sim} F_{\varepsilon_i}(\epsilon) = 1 - \exp(-\exp\epsilon)$

- Accordingly
  - $T_i = \exp(\sigma\varepsilon_i)\exp(\beta_0)\prod_{j=1}^p \exp(x_{ij}\beta_j)$
  - $S_{T_i}(t) = \exp[-\{t/\exp(\beta_0+\sum_{j=1}^p x_{ij}\beta_j)\}^{1/\sigma}] = \exp[-t^{1/\sigma}\exp\{(-\beta_0-\sum_{j=1}^p x_{ij}\beta_j)/\sigma\}]$ (as derived when introducing the log-linear model)
    - $\Rightarrow \lambda_{T_i}(t) = \sigma^{-1}t^{1/\sigma-1}\exp\{(-\beta_0 - \sum_{j=1}^p x_{ij}\beta_j)/\sigma\}$

- Interpretation of parameters
  - β_0: $\exp(\beta_0)$ is the baseline expected survival time.
  - β_j, $j \neq 0$ (after fixing all covariates other than the jth one): A one-unit increase in the jth covariate results in an expected survival time that is around $\exp(\beta_j)$ times the expected survival time at the baseline level.
    - Inconvenient to interpret β_j from the perspective of hazards (why?)

- Graphically check the correctness of model assumption
  1. Collect residuals $(\ln T_i - \hat\beta_0 - \sum_j x_{ij}\hat\beta_j)/\hat\sigma$ for uncensored subjects
  2. Compare residuals to a gumbel random sample via the Q-Q plot.

- Difference in the outputs of `R` functions
  - `survival::survreg`: "Intercept" ($\hat\beta_0$), "scale" ($\hat\sigma$), "log(scale)" ($\ln\hat\sigma$), and $\hat\beta_j$, $j = 1, \ldots, p$
  - `flexsurv::flexsurvreg`: "shape" ($1/\hat\sigma$), "scale" ($\exp(\hat\beta_0)$), and $\hat\beta_j$, $j = 1, \ldots, p$

Ex 4.3. Revisit the accelerated life testing of 10 motorettes

```r
head(MASS::motors)
data.ex43 = MASS::motors
data.ex43 = data.ex43[data.ex43$time != 0,] # ttr=0 not allowed in AFT models
is.factor(data.ex43$temp)
data.ex43$temp = factor(data.ex43$temp)
aft.ex43.1 = survival::survreg(
  survival::Surv(time, cens) ~ temp,
  data = data.ex43,
  dist="weibull",
  x = T
)
summary(aft.ex43.1)
# OR using flexsurv::flexsurvreg
aft.ex43.2 = flexsurv::flexsurvreg(
  survival::Surv(time, cens) ~ temp,
  data = data.ex43,
  dist = "weibull"
)
aft.ex43.2

# The survival time for temp=150
shape = 1/aft.ex43.1$scale
scale = unname(exp(aft.ex43.1$coefficients[1]))
(ET_tmp150 = 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)

# The survival time for temp=220
shape = 1/aft.ex43.1$scale
scale = unname(exp(
  aft.ex43.1$coefficients[1]+
  aft.ex43.1$coefficients['temp220']
))
(ET_tmp220 = 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)
  )
}
curve(surv.fun, from = 0, to = 1e3)

# Graphically check the correctness of weibull assumption
set.seed(718)
g.rnd = ordinal::rgumbel(10000, max = F) # gumbel random sample
lnTs.uncen = log(as.vector(data.ex43$time[data.ex43$cens==1]))
res = (
  lnTs.uncen - aft.ex43.1$x[data.ex43$cens==1,] %*% as.matrix(aft.ex43.1$coefficients)
)/aft.ex43.1$scale
qqplot(
  x = g.rnd,
  y = res,
  xlab = "Theoretical Quantiles",
  ylab = "Sample Quantiles"
)
qqline(res, distribution = function(p){ordinal::qgumbel(p,max=F)})
```

- Interpretation of $\hat{\beta}_1$: Campared with `temp=150`, `temp=170` multiplies the expected survival time by $\exp(\hat{\beta}_1)$.

Log-normal regression for right-censored data

- Observed $\{\widetilde{T}_i = \tilde{t}_i, \Delta_i = \delta_i, x_{i1}, \ldots, x_{ip}\}$
- Assuming
  - Independence across i
  - Independent and non-informative censoring
  - $\ln T_i = \beta_0 + \sum_{j=1}^{p} x_{ij}\beta_j + \sigma\varepsilon_i$ with
    - $\varepsilon_i \overset{\text{iid}}{\sim} N(0,1)$
- Accordingly
  - $T_i = \exp(\sigma\varepsilon_i)\exp(\beta_0)\prod_{j=1}^{p}\exp(x_{ij}\beta_j)$
  - $S_{T_i}(t) = 1 - \Phi\{\sigma^{-1}(\ln t - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)\}$ (as derived when introducing the log-linear model)

    * $\Rightarrow \lambda_{T_i}(t) = (\sigma t)^{-1}\phi\{\sigma^{-1}(\ln t - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)\}/S_{T_i}(t)$
        · $\phi(\cdot)$: the pdf of $N(0,1)$

- Interpretation of parameters
  - β_0: $\exp(\beta_0)$ is the baseline mean survival time.
  - β_j, $j \neq 0$ (after fixing all covariates other than the jth one): A one-unit increase in the jth covariate results in an expected survival time that is around $\exp(\beta_j)$ times the expected survival time at the baseline level.
- Graphically check the correctness of model assumption
  1. Collect residuals $(\ln T_i - \hat{\beta}_0 - \sum_j x_{ij}\hat{\beta}_j)/\hat{\sigma}$ for uncensored subjects
  2. Compare residuals to a $N(0,1)$ random sample via the Q-Q plot.
- Difference in the outputs of `R` functions
  - `survival::survreg`: "Intercept" ($\hat{\beta}_0$), "scale" ($\hat{\sigma}$), and $\hat{\beta}_j$, $j = 1, \ldots, p$
  - `flexsurv::flexsurvreg`: "meanlog" ($\hat{\beta}_0$), "sdlog" ($\hat{\sigma}$), and $\hat{\beta}_j$, $j = 1, \ldots, p$

Ex. 4.4. Revisit the data of bladder cancer recurrences which contain three treatment arms for 118 subjects.

```
data.ex44 = survival::capacitor
data.ex44 = data.ex44[data.ex44$time != 0,] # ttr=0 not allowed in AFT models
is.factor(data.ex44$temperature)
is.factor(data.ex44$voltage)
data.ex44$temperature = factor(data.ex44$temperature)
data.ex44$voltage = factor(data.ex44$voltage)
aft.ex44.1 = survival::survreg(
  survival::Surv(time, status) ~ temperature + voltage,
  data = data.ex44,
  dist="lognormal",
  x = T
)
summary(aft.ex44.1)
# OR using flexsurv::flexsurvreg
aft.ex44.2 = flexsurv::flexsurvreg(
  survival::Surv(time, status) ~ temperature + voltage,
  data = data.ex44,
  dist = "lognormal"
)
aft.ex44.2

# The survival time for temp=180 and voltage=300
sigma = aft.ex44.1$scale
mu = aft.ex44.1$coefficients[1] +
  aft.ex44.1$coefficients['temperature180'] +
  aft.ex44.1$coefficients['voltage300']
(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)
  )
}
```

```r
curve(surv.fun, from = 0, to = 2e3L) # plot the survival curve

# Graphically check the correctness of log-normal assumption
set.seed(718)
lnTs.uncen = log(as.vector(data.ex44$time[data.ex44$status==1]))
res = (
  lnTs.uncen - aft.ex44.1$x[data.ex44$status==1,] %*% as.matrix(aft.ex44.1$coefficients)
)
qqnorm(
  y = res,
  xlab = "Theoretical Quantiles",
  ylab = "Sample Quantiles"
)
qqline(res)
# Shapiro-Wilk test for normality
shapiro.test(res)
```

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 in terms of the inflation of failure time
  - 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