Cheatsheet for Survival Analysis

Irving Gómez Méndez

Main reference: Collett (2023).

Contents

1 Survivor, hazard, and cumulative hazard functions

Distribution function

$$F(t) = \mathbb{P}(T < t) = \int_0^t f(u)du$$

Survivor function

$$S(t) = P(T \geq t) = 1 - F(t)$$

Hazard function

$$h(t) = \frac{f(t)}{S(t)}$$

Cumulative hazard function

$$H(t) = \int_0^t h(u)du$$

1.1 Useful relations

$$f(t) = -\frac{d}{dt}S(t); \quad h(t) = -\frac{d}{dt}\log S(t); \quad S(t) = \exp\{-H(t)\}$$

1.2 Kaplan-Meier estimator

- n individuals.
- r death times, $t_{(1)} < t_{(2)} < \cdots, t_{(r)}$.
- n_j individuals alive just before $t_{(j)}$.
- d_j deaths at $t_{(j)}$.

For $t_{(k)} \leq t < t_{(k+1)}$:

$$\begin{aligned} S(t) &= \mathbb{P}(T \geq t) \\ &= \mathbb{P}(T \geq t | T \geq t_{(k)})\mathbb{P}(T \geq t_{(k)}) \\ &= \mathbb{P}(T \geq t | T \geq t_{(k)})\mathbb{P}(T \geq t_{(k)} | T \geq t_{(k-1)})\mathbb{P}(T \geq t_{(k-1)}) \\ &= \vdots \end{aligned}$$

$$\widehat{S}(t) = \prod_{j=1}^{k} \left(\frac{n_j - d_j}{n_j} \right)$$

1.3 Nelson-Aalen estimator

Because $S(t) = \exp\{-H(t)\}$, then

$$\tilde{S}(t) = \exp\left\{ -\sum_{j=1}^{k} \frac{d_j}{n_j} \right\}$$

1.4 Greenwood's formula (standard deviation of the Kaplan-Meier estimator)

$$\widehat{\text{sd}}\left(\widehat{S}(t)\right) = \widehat{S}(t) \left(\sum_{j=1}^{k} \frac{d_j}{n_j(n_j - d_j)} \right)^{1/2}$$

1.5 Standard deviation of the Nelson-Aalen estimator

(Grambsch and Therneau, 2000)

$$\widehat{\text{sd}}\left(\tilde{S}(t)\right) = \tilde{S}(t) \left(\sum_{j=1}^{k} \frac{d_j}{n_j^2} \right)^{1/2}$$

1.6 Confidence intervals for the survivor function

Practical, yet not very realistic

$$\widehat{S}(t) \dot{\sim} \mathcal{N}\left(S(t), \widehat{\mathrm{sd}}\left(\widehat{S}(t)\right)\right)$$

Logit transformation

$$\log\left(\frac{\widehat{S}(t)}{1-\widehat{S}(t)}\right) \dot{\sim} \mathcal{N}$$

Log-log transformation

$$\log\left(-\log \widehat{S}(t)\right) \dot{\sim} \mathcal{N}$$

1.7 Expected life time

$$\mu = \int_0^\infty S(t)dt; \quad \hat{\mu} = \sum_{j=1}^{k} \tau_j \widehat{S}(t_{(j)}), \text{ if } \widehat{S}(t_{(r)}) = 0, \quad \tau_j = t_{(j)} - t_{(j-1)}.$$

1.8 Quantile survival time

$$S(t(p)) = 1 - p$$

$$\hat{t}(p) = \min\left\{t_{(j)} : \widehat{S}(t_{(j)}) < 1 - p\right\}$$

$$\widehat{\mathrm{sd}}\left(\hat{t}(p)\right) = \frac{1}{\hat{f}\left(\hat{t}(p)\right)} \widehat{\mathrm{sd}}\left(\widehat{S}\left(\hat{t}(p)\right)\right)$$

where

$$\hat{f}\left(\hat{t}(p)\right) = \frac{\widehat{S}\left(\hat{u}(p)\right) - \widehat{S}\left(\hat{l}(p)\right)}{\hat{l}(p) - \hat{u}(p)},$$

$$\hat{u}(p) = \max\left\{t_{(j)} : \widehat{S}(t_{(j)}) \geq 1 - p + \varepsilon\right\}$$

$$\hat{l}(p) = \min\left\{t_{(j)} : \widehat{S}(t_{(j)}) \leq 1 - p - \varepsilon\right\}$$

1.9 Codes

00_KaplanMeierNelsonAalenQuantileInference

- Implement by hand the KM and NA estimators.
- Estimate by hand quantile survival time with its confidence interval.

01_KaplanMeierNelsonAalen

- Fit the KM and NA estimators using `lifelines`.
- To estimate median survival times: `kmf.median_survival_time_`.
- To get a CI for the median survival times:
  `median_survival_times(kmf.confidence_interval_)`.
  It turns out not to be very useful nor accurate.
- Quantile survival time: `qth_survival_time(p,kmf)`.

01_KaplanMeierNelsonAalen_ScikitSurvival

- Fit the KM and NA estimators using `sksurv`.
- `time, survival_prob, conf_int = kaplan_meier_estimator(...)`.
- `time[survival_prob=p].min()`.
- `u = time[survival_prob>=1-p+epsilon].max()`.
- `l = time[survival_prob<=1-p-epsilon].min()`.

2 Comparison of two groups of survival data

- r distinct death times across the two groups.
- Hypothesis, H: no difference between the groups.

Group	# deaths at $t_{(j)}$	# surviving at $t_{(j)}$	# risk just before $t_{(j)}$
I	d_{1j}	$n_{1j} - d_{1j}$	n_{1j}
II	d_{2j}	$n_{2j} - d_{2j}$	n_{2j}
Total	d_j	$n_j - d_j$	n_j

2.1 Log-rank test

$$U_L = \sum_{j=1}^{r} (d_{1j} - e_{1j}), \quad V_L = \sum_{j=1}^{r} v_{1j},$$

where

$$v_{1j} = \frac{n_{1j} n_{2j} d_j (n_j - d_j)}{n_j^2 (n_j - 1)}, \quad e_{1j} = n_{1j} \frac{d_j}{n_j}$$

Under H,

$$\frac{U_L}{\sqrt{V_L}} \dot{\sim} \mathcal{N}(0,1) \Rightarrow \frac{U_L^2}{V_L} \dot{\sim} \chi_1^2.$$

2.2 Wilcoxon test

$$U_W = \sum_{j=1}^{r} n_j (d_{1j} - e_{1j}), \quad V_L = \sum_{j=1}^{r} n_j^2 v_{1j}$$

2.3 Peto-Peto test

$$U_P = \sum_{j=1}^{r} \widehat{S}(t_{(j)})(d_{1j} - e_{1j}), \quad V_L = \sum_{j=1}^{r} \widehat{S}(t_{(j)})^2 v_{1j}$$

$\widehat{S}$ is the estimated survivor function for the combined data.

2.4 Codes

00_ComparisonTwoGroups

- Implement by hand the log-rank, Wilcoxon, and Peto-Peto tests.

3 Cox regression model

Proportional hazard assumption

$$h_{\text{Group}_1}(t) = \psi h_{\text{Group}_0}(t),\ \psi \geq 0.$$

ψ: hazard ratio.

If there are covariables $X^{(1)}, \dots, X^{(p)}$, then $\psi = \psi(\vec{X})$.

Because $\psi(\vec{X}) \geq 0$, it is convenient to make

$$\psi(\vec{X}) = \exp\{\eta\},$$

where

$$\eta = \beta_1 X^{(1)} + \beta_2 X^{(2)} + \cdots + \beta_p X^{(p)} = \vec{\beta}^T \vec{X}.$$

η: risk score or prognostic index.

$$h(t) = \psi(\vec{X}) h_0(t)$$

$$= \exp\{\vec{\beta}^T \vec{X}\} h_0(t).$$

$h_0(t)$: baseline hazard function.

If $X = \mathbb{1}\{\text{Individual is in Group 1}\}$, then

$$\begin{aligned} h_{\text{Group}_1}(t) &= \psi h_0(t) \\ &= e^{\beta} h_0(t). \end{aligned}$$

Thus, $\beta = 0 \Rightarrow \psi = e^{\beta} = 1$.

3.1 Partial Likelihood

$$L(\vec{\beta}) = \prod_{j=1}^{r} \frac{\exp\left\{\vec{\beta}^T \vec{X}_{(j)}\right\}}{\sum_{\ell \in R(t_{(j)})} \exp\left\{\vec{\beta}^T \vec{X}_{\ell}\right\}}.$$

L is maximized using Newton-Raphson.

3.2 Hypothesis testing

Under the hypothesis $H : \beta_j = 0$,

$$\frac{\hat{\beta}_j}{\widehat{\mathrm{sd}}(\hat{\beta}_j)} \dot{\sim} \mathcal{N}(0,1).$$

$$p_{value} = 2\texttt{norm.sf}\left(\left|\frac{\hat{\beta}_j}{\widehat{\mathrm{sd}}(\hat{\beta}_j)}\right|\right).$$

3.3 Breslow approximation

$$\tilde{\xi}_j = \exp\left\{\frac{-d_j}{\sum_{\ell \in R(t_{(j)})} \exp\{\vec{\beta}^T \vec{X}_\ell\}}\right\}$$

$$\hat{h}_0(t) = \frac{1-\tilde{\xi}_j}{t_{(j+1)} - t_{(j)}}$$

3.4 Comparing models

3.4.1 Nested models, log-likelihood test

- Model 1: p covariables.
- Model 2: $p+q$ covariables.
- Model 1$\subset$Model 2.

$$-2\log\frac{\hat{L}(\text{Model 1})}{\hat{L}(\text{Model 2})} \dot{\sim} \chi^2_q$$

3.4.2 Non-nested models, Akaike information criterion

$$\text{AIC} = -2\log\hat{L} + 2p$$

3.5 Codes

02_ComparisonTwoGroupsUsingCoxNoVariables

- Use `logrant_test` from `lifelines.statistics` to perform log-rank, Wilcoxon, and Peto-Peto tests.
- Use `CoxPHFitter()` from `lifelines` to compare two groups.
- `cph.print_summary()`

coef	exp(coef)	CI(β)	CI(ψ)	z-stat	p_{value}
$\hat{\beta}$	$\hat{\psi} = e^{\hat{\beta}}$				$H : \beta = 0$

- `-2*cph.log_likelihood_`$= -2\log\hat{L}$.
- `cph.log_likelihood_ratio_test()`$= -2\log\hat{L}_{\text{Null}} + 2\log\hat{L}_{\text{Model}}$.

4 Risk-adjusted survivor function

Because

$$S(t) = \exp\left\{-\int_0^t h(u)du\right\},$$

and

$$h(t) = \exp\{\vec{\beta}^T \vec{X}\} h_0(t),$$

then

$$S(t) = \exp\left\{\exp\{\vec{\beta}^T \vec{X}\}\left(-\int_0^t h_0(u)du\right)\right\}$$

$$= S_0(t)^{\exp\{\vec{\beta}^T \vec{X}\}}$$

Let be $\widehat{S}_i(t)$ the estimated survivor function for individual i, then

$$\widehat{S}(t) = \frac{1}{n}\sum_{i=1}^{n} \widehat{S}_i(t).$$

$\widehat{S}$: average survivor function or risk-adjusted survivor function.

4.1 Codes

09_CoxModelTwoFactors

- Fit Cox model using two factors.
- Estimate the average survivor function.

5 Stratified Cox regression

If proportional hazards cannot be assumed between groups

$$h_j(t) = \exp\{\vec{\beta}^T \vec{X}\} h_{0j}(t), \quad \text{for Group}_j$$

5.1 Codes

10_CoxModelTwoStrata

- Fit Cox model using a strata with two levels.

6 Concordance, standard deviation of risk score, and explained variation

6.1 Concordance indices

C-index

- It is a generalization of the ROC AUC, including censoring.
- $$C_{\text{index}} = \widehat{P}(\eta_i > \eta_j | T_i < T_j)$$
- It can only be calculated when there is no uncertainty of who lived longer.

K-statistic Gönen and Heller (2005)

$$K = \mathbb{P}(T_i < T_j | \eta_i > \eta_j)$$

$$\widehat{K} = \frac{2}{n(n-1)} \sum\sum_{i<j} \frac{\mathbb{1}\left(\hat{\eta}_i \neq \hat{\eta}_j\right)}{1 + \exp\left\{-\left|\hat{\eta}_j - \hat{\eta}_i\right|\right\}}$$

6.2 Standard deviation of risk score

D-statistic Royston and Sauerbrei (2004)

$$\widehat{\mathrm{sd}}(\eta) = \frac{D}{K}, \quad K = \sqrt{8/\pi}$$

Process to calculate D:

1. Order $\eta_{(1)} < \eta_{(2)} < \cdots < \eta_{(n)}$
2. Let be
$$z_{(i)} = \Phi^{-1}\left(\frac{i - 3/8}{n + 1/4}\right)$$
3. Fit the Cox model using
$$\frac{z_{(i)}}{\sqrt{8/\pi}}$$
as the predictive variable.
$$D = \hat{\beta}.$$

6.3 Explained variation

R^2_P Kent and O'Quigley (1988)

$$R^2_P = \frac{\widehat{V}_P}{\widehat{V}_P + \pi^2/6}, \quad \widehat{V}_P = \mathbb{V}(e\hat{t}a_1, \ldots, \hat{\eta}_n)$$

R^2_D Royston and Sauerbrei (2004)

$$R^2_D = \frac{D_0^2}{D_0^2 + \pi^2/6}, \quad D_0 = \frac{D}{\sqrt{8/\pi}}$$

6.4 Codes

11_ConcordanceAndExplainedVariance

- Implement by hand the statistics K, D, R^2_P, R^2_D.
- C_{index} =`CoxPHFitter().concordance_index_`.
- η =`CoxPHFitter().predict_log_partial_hazard(dat)`.

7 Time-dependent ROC curves

1. Fit the KM estimator for all the observations, $\widehat{S}(t)$.
2. Fix some time τ.
3. For c in $\hat{\eta}_{(1)}, \ldots, \hat{\eta}_{(n)}$:
   (a) Let be
   $$\text{Dat}_1 = \text{Dat}[\hat{\eta}_i > c], \quad \text{Dat}_2 = \text{Dat}[\hat{\eta}_i \leq c], \quad p = \frac{n_2}{n}.$$
   (b)
   $$\widehat{S}_1 = \text{KM estimator of Dat}_1, \quad \widehat{S}_2 = \text{KM estimator of Dat}_2$$
   (c)
   $$\widehat{\text{sens}}(\tau, c) = \frac{(1 - \widehat{S}_1(\tau))(1-p)}{1 - \widehat{S}(\tau)}$$
   (d)
   $$\widehat{\text{spec}}(\tau, c) = \frac{\widehat{S}_2(\tau)p}{\widehat{S}(\tau)}$$
4. Plot the ROC curve form by the points $(1 - \widehat{\text{spec}}(\tau, c), \widehat{\text{sens}}(\tau, c))$.
5. Calculate the AUC for the ROC curve.

7.1 Codes

12_TimeDependentRoc

- Implement time-dependent ROC curves by hand.
- `from sklearn.metrics import auc`.
- `auc(Roc[‘‘1-Spec’’]),Roc[‘‘Sens’’])`.

8 Weibull model

8.1 Exponential distribution

$$\begin{aligned}
h(t) &= \lambda, \\
H(t) &= \lambda t, \\
S(t) &= e^{-\lambda t}, \\
f(t) &= \lambda e^{-\lambda t}, \\
t(p) &= \frac{1}{\lambda} \log \left(\frac{1}{1-p} \right).
\end{aligned}$$

8.2 Weibull distribution

$$
\begin{aligned}
h(t) &= \lambda\gamma t^{\gamma-1},\\
H(t) &= \lambda t^{\gamma},\\
S(t) &= \exp\left\{-\lambda t^{\gamma}\right\},\\
f(t) &= \lambda\gamma t^{\gamma-1}\exp\left\{-\lambda t^{\gamma}\right\}.
\end{aligned}
$$

It is common the reparametrization:

$$
\beta = \frac{1}{\lambda^{1/\gamma}} \Leftrightarrow \lambda = \frac{1}{\beta^{\gamma}},
$$

$$
\alpha = \gamma.
$$

$$
\begin{aligned}
h(t) &= \frac{\alpha}{\beta^{\alpha}} t^{\alpha-1},\\
H(t) &= \left({}^{t}\!/\!_{\beta}\right)^{\alpha},\\
S(t) &= \exp\left\{-\left({}^{t}\!/\!_{\beta}\right)^{\alpha}\right\},\\
f(t) &= \frac{\alpha}{\beta^{\alpha}} t^{\alpha-1}\exp\left\{-\left({}^{t}\!/\!_{\beta}\right)^{\alpha}\right\},\\
t(p) &= \beta\left[\log\left(\frac{1}{1-p}\right)\right]^{1/\alpha}.
\end{aligned}
$$

- Weibull($\gamma = 1, \lambda$) = Exponential(λ).
- Hazard function increases monotonically if $\gamma > 1$.
- Hazard function decreases monotonically if $\gamma < 1$.

8.3 Log-likelihood

Likelihood

$$
\begin{aligned}
L &= \prod_{i=1}^{n} f(t_i)^{\delta_i} S(t_i)^{1-\delta_i}\\
&= \prod_{i=1}^{n} h(t_i)^{\delta_i} S(t_i)
\end{aligned}
$$

Log-likelihood

$$
\begin{aligned}
l &= \sum_{i=1}^{n} \delta_i \log f(t_i) + (1-\delta_i)\log S(t_i)\\
&= \sum_{i=1}^{n} \delta_i \log h(t_i) + \log S(t_i)
\end{aligned}
$$

8.4 Suitability of Weibull model

$$\log(-\log S(t)) = -\alpha \log \beta + \alpha \log t$$

- The points $(\log t_i, \log(-\log \widehat{S}(t_i)))$ should follow a straight line.
- We can implement linear regression:

$$\log(-\log \widehat{S}(t_i)) = \beta_0 + \beta_1 \log t_i + \varepsilon \Rightarrow \hat{\beta},$$

$$\exp\left\{-\hat{\beta}_0/\hat{\beta}_1\right\}, \quad \hat{\alpha} = \hat{\beta}_1.$$

8.5 Weibull model, Bayesian approach

$$T \sim \text{Weibull}(\alpha, \beta)$$

$$\begin{aligned} p(\alpha) &\propto \mathbb{1}_{(0,\infty)}(\alpha) \\ \lambda &\sim \text{Gamma}(\alpha = 1, \beta \to 0) \\ \beta &= 1/\lambda^{1/\alpha} \end{aligned}$$

8.6 Codes

15_BayesianWeibullNoCovariates

- Use `WeibullFitter` to fit Weibull model using `lifelines`.
- `lifeline` parameters are λ and ρ.
  λ corresponds with β.
  ρ corresponds with α.
- Implement Weibull model using a Bayesian approach.

9 Weibull regression model

9.1 Weibull proportional hazards model

Proportional hazard model

$$h(t) = \underbrace{\exp\left\{\vec{\beta}^T \vec{X}\right\}}_{\psi} h_0(t)$$

Let be $h_0(t) = \lambda\gamma t^{\gamma-1}$, then $h(t) = \psi\lambda\gamma t^{\gamma-1}$. That is, if

$$T_{\text{Group}_0} \sim \text{Weibull}(\gamma, \lambda) \Rightarrow T_{\text{Group}_1} \sim \text{Weibull}(\gamma, \psi\lambda)$$

9.2 Weibull accelerated failure time, Weibull AFT

$$h(t) = \exp\left\{\vec{\beta}^T \vec{X}\right\} \lambda\gamma t^{\gamma-1}$$

Making the reparametrization:

$$\vec{\beta} = -\gamma\vec{\zeta}, \quad \lambda = e^{-\gamma\mu}, \quad \gamma = \alpha,$$

then

$$h(t) = \frac{\alpha}{\beta^\alpha} t^{\alpha-1},$$

where

$$\beta = \exp\left\{\mu + \vec{\zeta}^T \vec{X}\right\}.$$

That is,

$$\begin{aligned} T &\sim \text{Weibull}(\alpha, \beta) \\ \beta &= \exp\{\eta\} \\ \eta &= \mu + \vec{\zeta}^T \vec{X} \end{aligned}$$

9.3 Weibull AFT for comparing two groups

$$\begin{aligned} T &\sim \text{Weibull}(\alpha, \beta) \\ \beta &= \exp\{\eta\} \\ \eta &= \zeta_{[0]} + \zeta_{[1]} \end{aligned}$$

$$\psi = e^{-\alpha(\zeta_{[1]} - \zeta_{[0]})}$$

9.4 Weibull AFT for comparing two groups, Bayesian approach

$$\begin{aligned} \zeta_{[j]} &\sim \mathcal{N}(0, \sigma_\zeta^2) \\ \beta_{[j]} &= e^{\zeta_{[j]}} \\ \alpha &\sim \text{HalfNormal}(\sigma_\alpha^2) \end{aligned}$$

9.5 Codes

17_BayesianWeibullRegression

- Implement WeibullAFT from a Bayesian approach.
- Use `WeibullAFTFitter` from `lifelines`.

10 Frailty models

z: Frailty level.

10.1 Conditional hazard, conditional cumulative hazard, and conditional survivor function

$$h(t|z) = zh_{\text{unfrail}}(t),$$

$$\begin{aligned} H(t|z) &= zH_{\text{unfrail}}(t) \\ &= z(-\log S_{\text{unfrail}}(t)), \end{aligned}$$

$$\begin{aligned} S(t|z) &= \exp\{-H(t|z)\} \\ &= \exp\{z \log S_{\text{unfrail}}(t)\}, \end{aligned}$$

$$\begin{aligned} \Rightarrow \log S(t|z) &= z \log S_{\text{unfrail}}(t) \\ &= -zH_{\text{unfrail}}(t) \end{aligned}$$

10.2 Likelihood

$$L = \prod_{i=1}^{n} \int_{-\infty}^{\infty} h(t_i|z_i)^{\delta_i} S(t_i|z_i) f_Z(z_i) dz_i$$

10.3 Gamma frailty

$$z \sim \Gamma(\theta, \theta)$$

The unfrailty model is achieved when $\theta \to \infty$.

$$h(t) = \frac{h_{\text{unfrail}}(t)}{1 + \theta^{-1} H_{\text{unfrail}}(t)},$$

$$H(t) = \theta \log\left(1 + \theta^{-1} H_{\text{unfrail}}(t)\right)$$

$$S(t) = \left(1 + \theta^{-1} H_{\text{unfrail}}(t)\right)^{-\theta}$$

$$f(t) = h_{\text{unfrail}}(t) \left(1 + \theta^{-1} H_{\text{unfrail}}(t)\right)^{-\theta-1}$$

$$t(p) = H_{\text{unfrail}}^{-1} \left(\theta(1-p)^{-1/\theta} - 1\right)$$

10.4 Log-likelihood

$$l = \sum_{i=1}^{n} \theta \log \theta - \log \Gamma(\theta) + \log \Gamma(\theta + \delta_i) + \delta_i h_{\text{unfrail}}(t_i) - (\theta + \delta_i) \log(\theta + H_{\text{unfrail}}(t_i))$$

10.5 Posterior distribution of Z_i

$$Z_i|\vec{t}, \vec{\delta} \sim \Gamma(\theta + \delta_i, \theta + H_{\text{unfrail}}(t_i))$$

$$\Rightarrow \mathbb{E}(Z_i|\vec{t}, \vec{\delta}) = \frac{\theta + \delta_i}{\theta + H_{\text{unfrail}}(t_i)}$$

10.6 Weibull-Gamma frailty model, Bayesian approach

$$T_{\text{unfrail}} \sim \text{Weibull}(\alpha, \beta)$$

$$\theta \sim \text{Exponential}(1)$$

We can use the same priors as before for the other parameters.

10.7 Codes

18_BayesianWeibullGammaFrailtyRegression

References

Collett, David (2023). *Modelling survival data in medical research.* Chapman and Hall/CRC.

Gönen, Mithat and Glenn Heller (2005). "Concordance probability and discriminatory power in proportional hazards regression". In: *Biometrika* 92.4, pp. 965–970.

Grambsch, Patricia M and Terry M Therneau (2000). "Modeling survival data: extending the Cox model". In: *Statistics for Biology and Health.*

Kent, John T and John O'Quigley (1988). "Measures of dependence for censored survival data". In: *Biometrika* 75.3, pp. 525–534.

Royston, Patrick and Willi Sauerbrei (2004). "A new measure of prognostic separation in survival data". In: *Statistics in medicine* 23.5, pp. 723–748.