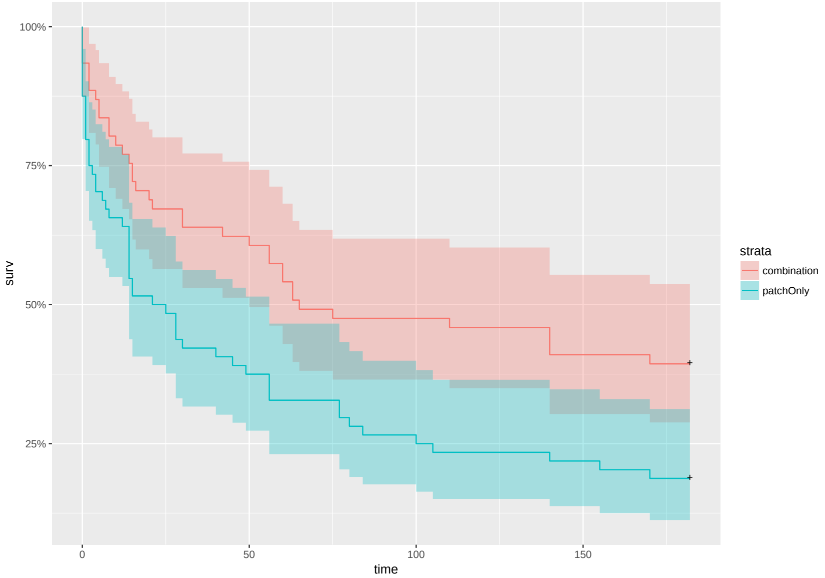
Survival and longitudinal data analysis
Chapter 2: tests and the Cox model

Agathe Guilloux
Professeure au LaMME - Université d'Évry - Paris Saclay

Comparing survival distributions

The pharmacoSmoking dataset

```
autoplot(survfit(Surv(ttr,relapse)~grp, data = pharmacoSmoking))
```



Construction of the log-rank test (1)

We consider

- two durations
  - T_1, with survival function $\bar{F}_1$ and
  - T_2, with survival function $\bar{F}_2$ and
- possibly censored by C_1 and C_2, independent of T_1 and T_2
- and that we have access to 2 groups of realizations
  - n_1 i.i.d. copies of $(T_1^C = \min(T_1, C_1), \delta_1 = \mathbb{1}_{T_1 \leq C_1})$ and
  - n_2 i.i.d. copies of $(T_2^C = \min(T_2, C_2), \delta_2 = \mathbb{1}_{T_2 \leq C_2})$:

$$\{(t_{1,1}^C, \delta_{1,1}), \ldots, (t_{1,n_1}^C, \delta_{1,n_1})\} \text{ and } \{(t_{2,1}^C, \delta_{2,1}), \ldots, (t_{2,n_2}^C, \delta_{2,n_2})\}$$

```
##     id ttr relapse         grp  ##      id ttr relapse       grp
## 3   39   5       1 combination  ## 1    21 182       0 patchOnly
## 4   80  16       1 combination  ## 2   113  14       1 patchOnly
## 5   87   0       1 combination  ## 7    16  14       1 patchOnly
```

Construction of the log-rank test (2)

Let $\tau_1 < \tau_2 < \ldots < \tau_D$ be the distinct times of event and, for each $k = 1, \ldots, D$

	At risk at τ_k	Dead at τ_k	At risk at τ_{k+1}
Group 1	$Y_{1,k}$	$d_{1,k}$	$Y_{1,k} - d_{1,k}$
Group 2	$Y_{2,k}$	$d_{2,k}$	$Y_{2,k} - d_{2,k}$
Total	Y_k	d_k	$Y_k - d_k$

Suppose that $\mathcal{H}_0 : \bar{F}_1 = \bar{F}_2$ holds, then the probability of observing $d_{1,k}$ deaths in group 1 at time τ_k is given by

$$\frac{\binom{d_k}{d_{1,k}} \binom{Y_k - d_k}{Y_{1,k} - d_{1,k}}}{\binom{Y_k}{Y_{1,k}}}$$

Construction of the log-rank test (3)

This defines a hypergeometric distribution with mean

$$E_k = \frac{Y_{1,k}}{Y_k} d_k$$

and variance

$$V_k = \frac{Y_{1,k} Y_{2,k} d_k (Y_k - d_k)}{Y_k^2 (Y_k - d_1)}.$$

The log-rank test

Now, it suffices to compare the observed number of deaths in group 1 to the expected one for each disctinct times $d_{1,k} - E_k$ and divide by the total variance

$$\frac{\sum_{k=1}^{D} d_{1,k} - E_k}{\sqrt{\sum_{k=1}^{D} V_k}}$$

The log-rank test

Under assumption $\mathcal{H}_0 : \bar{F}_1 = \bar{F}_2$, when n_1 and n_2 tend to infinity

$$\left(\frac{\sum_{k=1}^{D} d_{1,k} - E_k}{\sqrt{\sum_{k=1}^{D} V_k}}\right)^2 \xrightarrow{\mathcal{L}} \chi^2(1).$$

Remark: this is equivalent to the Cochran-Mantel-Haenzel test for testing the independence of two factors.

Example on the pharmocoSmoking dataset

```r
survdiff(Surv(ttr,relapse)~grp, data = pharmacoSmoking)
```

```
## Call:
## survdiff(formula = Surv(ttr, relapse) ~ grp, data = pharmacoSmoking)
##
##                    N Observed Expected (O-E)^2/E (O-E)^2/V
## grp=combination 61       37     49.9      3.36      8.03
## grp=patchOnly   64       52     39.1      4.29      8.03
##
##  Chisq= 8  on 1 degrees of freedom, p= 0.00461
```

Generalizations of the log-rank test

A generalization of the log-rank test has been proposed in Harrington and Fleming 1982, it introduces weights:

$$\frac{\sum_{k=1}^{D} \omega_k (d_{1,k} - E_k)}{\sqrt{\sum_{k=1}^{D} \omega_k^2 V_k}}$$

of the form

- $\omega_k = Y_k$ for an equivalent of the Mann-Withney-Wilcoxon test.
- $\omega_k = \widehat{F}^{\rho}(\tau_k)$ for the G-rho family of Harrington and Fleming 1982 (coded in function `survdiff`)

The idea is to give more weight to times points where there is the most data.

Tests for more than two samples

Now, suppose that they are L subgroups for which we want to test whether $\bar{F}_1 = \ldots = \bar{F}_L$. For example, this is the case where there are more than 2 possible treatments. For each subgroup l, define

$$E_{l,k} = \frac{Y_{l,k}}{Y_k} d_k \text{ and}$$
$$\hat{\Sigma} = \left(V_k^{l_1,l_2} = \frac{Y_{l_1,k}}{Y_k} d_k \left(\mathbb{1}_{l_1=l_2} - \frac{Y_{l_2,k}}{Y_k} \right) \frac{Y_k - d_k}{Y_k - 1} \right).$$

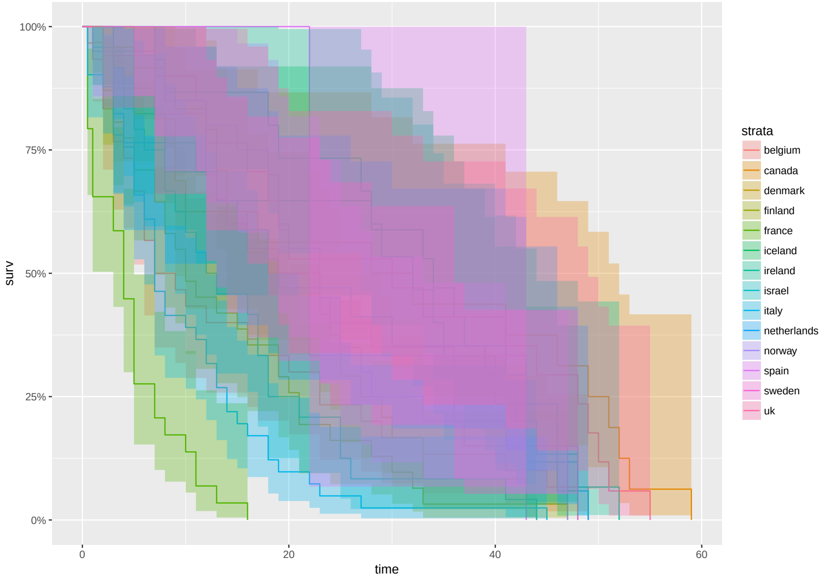
The k-sample log-rank test

Under assumption $\mathcal{H}_0 : \bar{F}_1 = \ldots = \bar{F}_L$, when $n_1, \ldots, n_L$ tend to infinity

$$\begin{pmatrix} \sum_{k=1}^{D} d_{1,k} - E_{1,k} \\ \ldots \\ \sum_{k=1}^{D} d_{L,k} - E_{L,k} \end{pmatrix}^{\top} \hat{\Sigma}^{-1} \begin{pmatrix} \sum_{k=1}^{D} d_{1,k} - E_{1,k} \\ \ldots \\ \sum_{k=1}^{D} d_{L,k} - E_{L,k} \end{pmatrix} \xrightarrow{\mathcal{L}} \chi^2(L-1).$$

Coalition data King et al. 1990

This dataset contains survival data on government coalitions in parliamentary democracies for the period 1945-1987.



Coalition data King et al. 1990

```r
survdiff(Surv(duration,rep(1,n))~country, data=coalition)
```

```
##                        N Observed Expected (O-E)^2/E (O-E)^2/V
## country=belgium      30       30    24.68   1.14911   1.34631
## country=canada       16       16    31.94   7.95299  11.07080
## country=denmark      24       24    24.37   0.00554   0.00643
## country=finland      31       31    21.73   3.95077   4.54284
## country=france       29       29     7.10  67.48666  75.15721
## country=iceland      17       17    23.66   1.87235   2.18122
## country=ireland      15       15    24.66   3.78615   4.45779
## country=israel       24       24    17.77   2.18045   2.46753
## country=italy        41       41    20.67  19.98748  23.32714
## country=netherlands 17       17    22.26   1.24259   1.44947
## country=norway       20       20    24.62   0.86860   1.00660
## country=spain         3        3     4.21   0.34671   0.37127
## country=sweden       20       20    25.51   1.18965   1.39553
## country=uk           17       17    30.82   6.19431   7.82142
##
##  Chisq= 142  on 13 degrees of freedom, p= 0
```

Semi-parametric proportional hazard model

Covariates in the pharmocoSmoking dataset

```r
head(pharmacoSmoking)
```

```
##    id ttr relapse         grp age gender    race employment yearsSm
## 1  21 182       0   patchOnly  36   Male   white         ft
## 2 113  14       1   patchOnly  41   Male   white      other
## 3  39   5       1 combination  25 Female   white      other
## 4  80  16       1 combination  54   Male   white         ft

##   levelSmoking ageGroup2 ageGroup4 priorAttempts longestNoSmoke
## 1        heavy     21-49     35-49             0              0
## 2        heavy     21-49     35-49             3             90
## 3        heavy     21-49     21-34             3             21
## 4        heavy       50+     50-64             0              0
```

We observe for each $i = 1, \ldots, n$

$$(T_i^C, \delta_i) \text{ AND } X_i^\top \in \mathbb{R}^p \text{ (here } p = 11)$$

The proportional hazards model or Cox 1972 model (2)

The proportional hazards model

Let $\lambda(t|X)$ be the hazard rate at time t for an individual with covariates $X = (X^1, \ldots, X^p)$ (vector of size $1 \times p$). In the proportional hazards model, this hazard rate takes the form

$$\begin{aligned}\lambda(t|X) &= \lambda_0^\star(t) \exp\big(X\beta^\star\big) \\ &= \lambda_0^\star(t) \exp\Big(\sum_{j=1}^{p} X^j \beta_j^\star\Big)\end{aligned}$$

where

- $\lambda_0^\star$ is an unknown function, called "baseline hazard rate" (or "baseline intensity function")
- $\beta^\star$ is an unknown vector of regression parameters in $\mathbb{R}^p$.

The proportional hazards model or Cox 1972 model (2)

Key relation of the Cox model

Let i_1 and i_2 be two individuals with covariates X_{i_1} and X_{i_2} respectively, then

$$\frac{\lambda(t|X_{i_1})}{\lambda(t|X_{i_2})} = \frac{\lambda_0^\star(t)\exp\left(X_{i_1}\beta^\star\right)}{\lambda_0^\star(t)\exp\left(X_{i_2}\beta^\star\right)} = \exp\left((X_{i_1} - X_{i_2})\beta^\star\right)$$

Hazard ratio

Let us assume that X_{i_1} and X_{i_2} only differ on the jth covariate ($X_{i_1}^k = X_{i_2}^k$ for $k \neq j$ and $X_{i_1}^j \neq X_{i_2}^j$. In this case,

$$\frac{\lambda(t|X_{i_1})}{\lambda(t|X_{i_2})} = \exp\left((X_{i_1} - X_{i_2})\beta^\star\right) = \exp\left((X_{i_1}^j - X_{i_2}^j)\beta_k^\star\right).$$

Now suppose that the jth covariate encodes a treatment. For example, individual i_1 has recived a treatment $X_{i_1}^j = 1$ and i_2 did not $X_{i_2}^j = 0$, then

$$\frac{\lambda(t|X_{i_1})}{\lambda(t|X_{i_2})} = \exp\left(\beta_k^\star\right).$$

The value $\exp\left(\beta_k^\star\right)$ is also called the **relative risk.**

Cox model with treatment groups in the pharmocoSmoking dataset

```r
summary(coxph(Surv(ttr,relapse)~grp, data = pharmacoSmoking))
```

```
## Call:
## coxph(formula = Surv(ttr, relapse) ~ grp, data = pharmacoSmoking)
##
##   n= 125, number of events= 89
##
##                 coef exp(coef) se(coef)   z Pr(>|z|)
## grppatchOnly 0.6050    1.8313   0.2161 2.8  0.00511 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##              exp(coef) exp(-coef) lower .95 upper .95
## grppatchOnly     1.831     0.5461     1.199     2.797
```

Hazard ratio

For the jth covariate, the value $\exp(\beta_k^\star)$ is called the hazard ratio. When

- $X_{i_1}^j = X_{i_2}^j + 1$
- and other things being equal, ($X_{i_1}^k = X_{i_2}^k$ for $k \neq j$)

it equals

$$\frac{\lambda(t|X_{i_1})}{\lambda(t|X_{i_2})} = \exp\left((X_{i_1} - X_{i_2})\beta^\star\right) = \exp\left((X_{i_1}^j - X_{i_2}^j)\beta_j^\star\right) = \exp(\beta_k^\star)$$

It is interpreted as the constant by which the hazard function is multiplied when X^j increases of 1 unit.

Cox model with treatment groups and age in the pharmocoSmoking dataset

```r
summary(coxph(Surv(ttr,relapse) ~ grp + age  , data = pharmacoSmoking))
```

```
## Call:
## coxph(formula = Surv(ttr, relapse) ~ grp + age,
##^^I^^I^^I^^I^^I^^Idata = pharmacoSmoking)
##
##   n= 125, number of events= 89
##
##                   coef exp(coef)  se(coef)      z Pr(>|z|)
## grppatchOnly  0.558663  1.748334  0.216674  2.578  0.00993 **
## age          -0.023018  0.977245  0.009605 -2.397  0.01655 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##              exp(coef) exp(-coef) lower .95 upper .95
## grppatchOnly    1.7483      0.572     1.143    2.6734
## age             0.9772      1.023     0.959    0.9958
```

Derivation of the partial likelihood (1)

We just saw estimates of the true regression parameter $\beta^\star$, we now describe how they are derived.

Let us come back to the likelihood for n independent individuals, independently right-censored data. We observe

$$(T_1^C, \delta_1, X_1), (T_2^C, \delta_2, X_2), \ldots, (T_n^C, \delta_n, X_n).$$

The likelihood is proportional to:

$$\prod_{i=1}^n f(T_i^C)^{\delta_i} \bar{F}(T_i^C)^{1-\delta_i} = \prod_{i=1}^n \left(\frac{f(T_i^C)}{\bar{F}(T_i^C)} \right)^{\delta_i} \bar{F}(T_i^C) = \prod_{i=1}^n \lambda(T_i^C | X_i)^{\delta_i} \bar{F}(T_i^C)$$

$$= \prod_{i=1}^n \left(\lambda_0(T_i^C) \exp\left(X_i \beta\right) \right)^{\delta_i} \exp\left(- \Lambda_0(T_i^C) \exp\left(X_i\beta\right) \right).$$

Derivation of the partial likelihood (2)

To find the maximum likelihood estimator, we start by optimizing with respect to each $\hat{\lambda}_0(T_i^C)$ at a fixed value of β. To that end, notice that

$$\sum_{i=1}^{n} \hat{\Lambda}_0(T_i^C) \exp\left(X_i\beta\right) = \sum_{i=1}^{n} \hat{\lambda}_0(T_i^C) \sum_{j:\, T_j^C \geq T_i^C} \exp\left(X_j\beta\right)$$

(when $\hat{\Lambda}$ is a step function) which gives

$$\hat{\lambda}_0(T_i^C, \beta) = \frac{\delta_i}{\sum_{j:\, T_j^C \geq T_i^C} \exp\left(X_j\beta\right)}.$$

Derivation of the partial likelihood (3)

Notice that

$$\sum_{i=1}^{n} \hat{\lambda}_0(T_i^C) \sum_{j:\, T_j^C \geq T_i^C} \exp\left(X_j\beta\right) = \sum_{i=1}^{n} \delta_i$$

replace then λ_0 by $\hat{\lambda}_0$ in the equation above:

$$\prod_{i=1}^{n} \left(\lambda_0(T_i^C|X_i) \exp\left(X_i\beta\right)\right)^{\delta_i} \exp\left(-\sum_{i=1}^{n} \lambda_0(T_i^C) \sum_{j:\, T_j^C \geq T_i^C} \exp\left(X_j\beta\right)\right)$$

$$= \prod_{i=1}^{n} \left(\frac{\delta_i}{\sum_{j:\, T_j^C \geq T_i^C} \exp\left(X_j\beta\right)} \exp\left(X_i\beta\right)\right)^{\delta_i} \exp(\sum_{i=1}^{n} \delta_i)$$

with the convention $0^0 = 1$.

Derivation of the partial likelihood (4)

The Cox partial likelihood

The Cox partial likelihood is defined as

$$\mathcal{L}^{\text{partial}}(\beta) = \prod_{i=1}^{n} \left(\frac{\exp\left(X_i \beta\right)}{\sum_{j:\, T_j^C \geq T_i^C} \exp\left(X_j \beta\right)} \right)^{\delta_i}. \tag{1}$$

The maximum estimator of $\beta^\star$ is defined as

$$\hat{\beta} = \text{argmin}_{\beta \in \mathbb{R}^p} \mathcal{L}^{\text{partial}}(\beta).$$

Prediction via the Breslow estimator

Recall that $\bar{F}^*(t|X_i) = \exp\Big(-\int_0^t \lambda_0^\star(s)\exp(X_i\beta^\star)ds\Big)$.

The Breslow estimator

Once $\hat{\beta}$ computed, the Breslow estimator of $\Lambda_0^\star(t)$ is defined as

$$\hat{\Lambda}_0(t) = \sum_{i:\, T_i^C \leq t} \frac{\delta_i}{\sum_{j:\, T_j^C \geq T_i^C} \exp\big(X_j\hat{\beta}\big)}.$$

All this can be defined (with few differences) in the case where T has a discrete distribution. Methods to handle such ties include Breslow's and Efron's methods (see Klein and Moeschberger 2005 page 259 for more details).

Asymptotic distributions

Let $I_n(\beta)$ be the information matrix associated with the Cox partial likelihood defined in Equation (1) (you can compute it, it is ugly...).

Asymptotic distributions of $\hat{\beta}$

As n tends to infty

$$I_n(\hat{\beta})^{-1/2}\left(\hat{\beta} - \beta^\star\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0,1).$$

Asymptotic distributions of the likelihood ratio

As n tends to infty

$$-2\left(\log \mathcal{L}^{\text{partial}}(\hat{\beta}) - \log \mathcal{L}^{\text{partial}}(\beta^\star)\right) \xrightarrow{\mathcal{L}} \chi^2(p).$$

Univariate Wald tests

Let $\hat{\sigma}_j^2$ be the jth diagonal element of $I_n(\hat{\beta})$.

The univariate Wald test for $\beta_j^\star = 0$

To test $\mathcal{H}_0 : \beta_j^\star = 0$ at level α, use the Wald test statistic

$$\frac{\hat{\beta}_j^2}{\hat{\sigma}_j^2}$$

and reject $\mathcal{H}_0$ when it is greater than $q_{\chi^2(1)}(1-\alpha)$.

Univariate Wald tests in the pharmocoSmoking dataset

```
summary(coxph(Surv(ttr,relapse) ~ grp + age  , data = pharmacoSmoking))
```

```
## Call:
## coxph(formula = Surv(ttr, relapse) ~ grp + age,
##^^I^^I^^I^^I^^I^^Idata = pharmacoSmoking)
##
##   n= 125, number of events= 89
##
##                    coef exp(coef)  se(coef)      z Pr(>|z|)
## grppatchOnly  0.558663  1.748334  0.216674  2.578  0.00993 **
## age          -0.023018  0.977245  0.009605 -2.397  0.01655 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##              exp(coef) exp(-coef) lower .95 upper .95
## grppatchOnly    1.7483      0.572     1.143    2.6734
## age             0.9772      1.023     0.959    0.9958
```

Tests for $\beta^\star = 0$

Wald test

We known that, as n tends to infty

$$I_n(\hat\beta)^{-1/2}\big(\hat\beta - \beta^\star\big) \xrightarrow{\mathcal{L}} \mathcal{N}(0,1),$$

it implies that

$$\big(\hat\beta - \beta^\star\big)^\top I_n(\hat\beta)^{-1}\big(\hat\beta - \beta^\star\big) \xrightarrow{\mathcal{L}} \chi^2(p).$$

To test $\mathcal{H}_0 : \beta_1^\star = \ldots = \beta_p^\star = 0$ at level α, use the Wald test statistic

$$\hat\beta^\top I_n(\hat\beta)^{-1}\hat\beta$$

and reject $\mathcal{H}_0$ when it is greater than $q_{\chi^2(p)}(1-\alpha)$.

Likelihood ratio test

To test $\mathcal{H}_0 : \beta_1^\star = \ldots = \beta_p^\star = 0$ at level α, use the likelihood ratio test statistic

$$-2\Big(\log \mathcal{L}^{\text{partial}}(\hat\beta) - \log \mathcal{L}^{\text{partial}}(0)\Big)$$

and reject $\mathcal{H}_0$ when it is greater than $q_{\chi^2(p)}(1-\alpha)$.

Tests in the pharmocoSmoking dataset

```r
summary(coxph(Surv(ttr,relapse) ~ grp + age  , data = pharmacoSmoking))
```

```
## Call:
## coxph(formula = Surv(ttr, relapse) ~ grp + age, data = pharmacoSmoki
##
##   n= 125, number of events= 89
##
##                    coef exp(coef)  se(coef)      z Pr(>|z|)
## grppatchOnly  0.558663  1.748334  0.216674  2.578  0.00993 **
## age          -0.023018  0.977245  0.009605 -2.397  0.01655 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##              exp(coef) exp(-coef) lower .95 upper .95
## grppatchOnly    1.7483      0.572     1.143    2.6734
## age             0.9772      1.023     0.959    0.9958
##
## Concordance= 0.625  (se = 0.034 )
## Rsquare= 0.105   (max possible= 0.998 )
## Likelihood ratio test= 13.82  on 2 df,   p=0.0009956
## Wald test            = 13.48  on 2 df,   p=0.001183
## Score (logrank) test = 13.74  on 2 df,   p=0.00104
```

Concordance index

A common concordance measure that does not depend on time is the C-index (see Harrell, Lee, and Mark 1996) defined by

$$C_{\text{Harrell}} = \mathbb{P}[M_i > M_j | T_i < T_j],$$

with $i \neq j$ two independent patients, and $M_i = X_i\hat{\beta}$ and $M_j = X_j\hat{\beta}$ are the marker value in a given Cox model. In Heagerty and Zheng 2005, is proposed an estimation of the C_{Harrell} in the Cox model and under censoring.

Comparing survival distributions

The 2-sample log-rank test

Generalization to k-sample tests

k-sample tests

Semi-parametric proportional hazard model

The proportional hazards model or Cox 1972 model

Hazard ratio

Partial likelihood

Asymptotic distributions and tests

References I

David R. Cox. "Regression models and life tables (with discussion)". In: *Journal of the Royal Statistical Society* 34 (1972), pp. 187–220.

David P Harrington and Thomas R Fleming. "A class of rank test procedures for censored survival data". In: *Biometrika* (1982), pp. 553–566.

Frank E Harrell, Kerry L Lee, and Daniel B Mark. "Tutorial in biostatistics multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors". In: *Statistics in medicine* 15 (1996), pp. 361–387.

Patrick J Heagerty and Yingye Zheng. "Survival model predictive accuracy and ROC curves". In: *Biometrics* 61.1 (2005), pp. 92–105.

Gary King et al. "A unified model of cabinet dissolution in parliamentary democracies". In: *American Journal of Political Science* (1990), pp. 846–871.

John P Klein and Melvin L Moeschberger. *Survival analysis: techniques for censored and truncated data*. Springer Science & Business Media, 2005.