Descriptive measures of survival experience

Goals of Survival Analysis:

1) To estimate and interpret survivor and/or hazard functions from survival data.
2) To compare survivor and/or hazard functions.
3) To assess the relationship of explanatory variables to survival time. Use math modeling, e.g., Cox proportional hazards

AS475 Survival Models for Actuaries Formula

\[
\Gamma(\alpha; x) = \frac{1}{\Gamma(\alpha)} \int_0^x t^{\alpha-1} e^{-t} dt \quad \Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt \quad \Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)
\]

<table>
<thead>
<tr>
<th>Continuous Dsn</th>
<th>pdf ( f(x) )</th>
<th>mgf ( M_X(t) )</th>
<th>Mean ( E[X] )</th>
<th>Var ( X )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uniform((a, b))</td>
<td>( f(x) = \begin{cases} (b - a)^{-1} &amp; a &lt; x &lt; b \ 0 &amp; \text{otherwise} \end{cases} )</td>
<td>( \frac{e^{bx} - e^{ax}}{t(b - a)} )</td>
<td>( \frac{b + a}{2} )</td>
<td>( \frac{(b - a)^2}{12} )</td>
</tr>
<tr>
<td>Exponential((\lambda)) ( \lambda &gt; 0 )</td>
<td>( f(x) = \begin{cases} \lambda e^{-\lambda x} &amp; x \geq 0 \ 0 &amp; x &lt; 0 \end{cases} )</td>
<td>( \frac{\lambda}{\lambda - t} )</td>
<td>( \frac{1}{\lambda} )</td>
<td>( \frac{1}{\lambda^2} )</td>
</tr>
<tr>
<td>Gamma((\alpha, \lambda))</td>
<td>( f(x) = \begin{cases} \lambda e^{-\lambda x} (\lambda x)^{\alpha-1} &amp; x \geq 0 \ 0 &amp; x &lt; 0 \end{cases} )</td>
<td>( \left( \frac{\lambda}{\lambda - t} \right)^\alpha )</td>
<td>( \frac{\alpha}{\lambda} )</td>
<td>( \frac{\alpha}{\lambda^2} )</td>
</tr>
<tr>
<td>Normal((\mu, \sigma^2)) (-\infty &lt; x &lt; \infty)</td>
<td>( f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)} )</td>
<td>( \exp \left( \mu t + \frac{\sigma^2 t^2}{2} \right) )</td>
<td>( \mu )</td>
<td>( \sigma^2 )</td>
</tr>
<tr>
<td>Pareto((\alpha, \theta))</td>
<td>( f(x) = \frac{\theta^\alpha}{(x + \theta)^{\alpha+1}} )</td>
<td>( F(x) = 1 - \left( \frac{\theta}{x + \theta} \right)^\alpha )</td>
<td>( M_X(t) ) not given</td>
<td>( \frac{\theta}{\alpha - 1} )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Discrete Dsn</th>
<th>pmf ( p(x) )</th>
<th>mgf ( M(t) )</th>
<th>Mean ( E[X] )</th>
<th>Variance ( Var(X) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Binomial((n, p)) ( 0 \leq p \leq 1 )</td>
<td>( \begin{pmatrix} n \ x \end{pmatrix} p^x (1-p)^{n-x} ) ( x = 0, 1, \ldots, n )</td>
<td>( (pe^t + 1 - p)^n )</td>
<td>( np )</td>
<td>( np(1-p) )</td>
</tr>
<tr>
<td>Poisson((\lambda)) ( \lambda &gt; 0 )</td>
<td>( e^{-\lambda} \lambda^x / x! ) ( x = 0, 1, \ldots )</td>
<td>( \exp[\lambda(e^{-t} - 1)] )</td>
<td>( \lambda )</td>
<td>( \lambda )</td>
</tr>
<tr>
<td>Geometric((p))</td>
<td>( p^x (1-p)^{r-1} ) ( x = 0, 1, \ldots )</td>
<td>( \frac{pe^t}{1 - (1-p)e^t} )</td>
<td>( \frac{1}{p} )</td>
<td>( \frac{1 - p}{p^2} )</td>
</tr>
<tr>
<td>Negative Binomial((r, p))</td>
<td>( \begin{pmatrix} n - 1 \ r - 1 \end{pmatrix} p^r (1-p)^{n-r} ) ( n = r, r+1, \ldots )</td>
<td>( \left[ \frac{pe^t}{1 - (1-p)e^t} \right]^r )</td>
<td>( \frac{r}{p} )</td>
<td>( \frac{1 - p}{p^2} )</td>
</tr>
<tr>
<td>Hypergeometric((n, K, N))</td>
<td>( \begin{pmatrix} K \ x \end{pmatrix} \begin{pmatrix} N-K \ n-x \end{pmatrix} ) ( x = 0, 1, \ldots, \min(n, K) )</td>
<td>special function</td>
<td>( np^* = \frac{K}{N} )</td>
<td>( np^<em>(1-p^</em>) \frac{N-n}{N-1} )</td>
</tr>
</tbody>
</table>

KK1 Introduction to Survival Analysis

Time = survival time
Event = failure
Left-censored: true survival time \( \leq \) the observed survival time
Right-censored: true survival time \( \geq \) observed survival time
Interval-censored: true survival time is within a known time interval

Left censoring \( \Rightarrow t_1 = 0; t_2 = \) upper bound Right censoring \( \Rightarrow t_1 = \) lower bound; \( t_2 = \infty \)

\( d = \begin{cases} 1 & \text{if failure} \\ 0 & \text{censored} \end{cases} \)

\( S(t) = \text{survivor function} \quad h(t) = \text{hazard function} \)

Hazard function = conditional failure rate \( h(t) = \text{instantaneous potential} \)

\[ S(t) = P(T > t) \quad h(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} \]

Relationship of \( S(t) \) and \( h(t) \): If you know one, you can determine the other.

\[ h(t) = \lambda \text{ iff } S(t) = e^{-\lambda t} \quad h(t) = - \left[ \frac{dS(t)/dt}{S(t)} \right] \quad S(t) = \exp \left[ - \int_0^t h(u) du \right] \quad \hat{S}(t) = \text{observed survivor function} \]

Goals of Survival Analysis: 1) To estimate and interpret survivor and/or hazard functions from survival data.
2) To compare survivor and/or hazard functions.
3) To assess the relationship of explanatory variables to survival time. Use math modeling, e.g., Cox proportional hazards

Descriptive measures of survival experience
Average survival time: \( T = \frac{1}{n} \sum_{i=1}^{n} t_i \)

<table>
<thead>
<tr>
<th>Measure of effect:</th>
<th>Linear regression</th>
<th>Logistic regression</th>
<th>Survival analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>regression coefficient ( \beta )</td>
<td>odds ratio ( e^\beta )</td>
<td>hazard ratio ( e^\beta )</td>
<td></td>
</tr>
</tbody>
</table>

Censoring Assumptions: Three assumptions
a) Independent (vs non-independent) censoring
b) Random (vs. non-random) censoring
c) Non-informative (vs. informative) censoring

**KK2: Kaplan-Meier Curves and the Log-Rank Test**

*Kaplan Meier* curves (see also KPW12). \( S(t_{(f)}) = S(t_{(f-1)})P(T > t_{(f)})P(T \geq t_{(f)}) = \prod_{i=1}^{f} P(T > t_{(i)})P(T \geq t_{(i)}) \)

Note: Kaplan-Meier product limit estimator comes from the probability rule \( P(A \cap B) = P(A) \times P(B | A) \)

**Log-Rank Test** for no difference in survival curves of Several Groups: \( d'V^{-1}d \sim \chi^2_{G-1}, i = 1, 2, \cdots, G \) where

\[
\begin{align*}
    d &= (O_1 - E_1, O_2 - E_2, \cdots, O_{G-1} - E_{G-1})' \\
    V &= (v_{ij}) \\
    v_{ii} &= \text{Var}(O_i - E_i) = \sum_{f=1}^{k} n_{if}(n_{if} - m_{if})m_{if}(n_{if} - m_{if})/n_i^2(n_i - 1) \\
    v_{ij} &= \text{Cov}(O_i - E_i, O_j - E_j) = \sum_{f=1}^{k} -n_{if}n_{jf}m_{if}(n_{if} - m_{if})/n_i^2(n_i - 1)
\end{align*}
\]

**Log-Rank Test** for no difference in survival curves of 2 Groups: \( \frac{(O_i - E_i)^2}{\text{Var}(O_i - E_i)} \sim \chi^2_1, i = 1, 2 \) where

\[
\begin{align*}
    O_i - E_i &= \sum_f (m_{if} - e_{if}), \; \text{Var}(O_i - E_i) = \sum_f n_{if}n_{2f}(m_{1f} + m_{2f})(n_{1f} + n_{2f} - m_{1f} - m_{2f})/n_{1f}n_{2f}(n_{1f} + n_{2f} - 1) \\
    e_{if} &= \left( \frac{n_{if}}{n_{1f} + n_{2f}} \right)(m_{1f} + m_{2f}) = \text{expected counts} = (\text{proportion in risk set}) \times (\text{#failures over both groups}) \\
    m_{if} &= \text{observed counts for the } i^{th} \text{ group at time } f.
\end{align*}
\]

*Approximate formula: \( \sum_{i=1}^{G} \frac{(O_i - E_i)^2}{E_i} \sim \chi^2_1, i = 1, 2. \)

**Alternative tests** for 2 groups: Test statistic: \( \frac{\left( \sum_f w(t_{(f)}) (m_{if} - e_{if}) \right)^2}{\text{Var} \left( \sum_f w(t_{(f)}) (m_{if} - e_{if}) \right)} \) where \( w(t_{(f)}) = \text{weights at the } f^{th} \text{ failure time.} \)

<table>
<thead>
<tr>
<th>( f^{th} ) failure time.</th>
<th>LogRank</th>
<th>Wilcoxon</th>
<th>Tarone-Ware</th>
<th>Peto</th>
<th>Flamington - Harrington</th>
</tr>
</thead>
<tbody>
<tr>
<td>( w(t_{(f)}) )</td>
<td>( 1 )</td>
<td>( n_f )</td>
<td>( \sqrt{n_f} )</td>
<td>( s(t_{(f)}) )</td>
<td>( S(t_{(f-1)})^p[1 - S(t_{(f-1)})]^q )</td>
</tr>
</tbody>
</table>

\( p = 0 \rightarrow \text{LogRank} \)

**Cox Models: KK3-KK6**
**De...nition 1**
The empirical distribution is a set of distribution functions, each member of which is determined by specifying one or more values called parameters. The number of parameters is...cted as at least as complex as the data or knowledge that produced it, and the number of "parameters" increases as the number of data points or amount of knowledge increases.

**De...nition 2**
A parametric distribution is a set of distribution functions, each member of which is determined by specifying one or more values called parameters. The number of parameters is...ed and finite.

**De...nition 3**
The empirical distribution is obtained by assigning probability 1/n to each data point.

**De...nition 4**
A kernel smoothed distribution is obtained by replacing each data point with a continuous random variable and then assigning probability 1/n to each such random variable. The random variables used must be identical except for a location or scale change that is related to its associated data point.

**De...nition 5**
The empirical distribution function is defined as $F_n(x) = \frac{\text{number of observations} \leq x}{n}$, when $n$ is the total number of observations.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>HR: $h(t, X)$</td>
<td>$h_0(t) \exp(\sum_{i=1}^{p} \beta_i X_i)$</td>
<td>$h_{0g}(t) \exp(\sum_{i=1}^{p} \beta_i X_i)$</td>
<td>$h_0(t) \exp(\sum_{i=1}^{p_1} \beta_i X_i + \sum_{j=1}^{p_2} \delta_j X_j)$</td>
</tr>
<tr>
<td>Mean... $\theta$</td>
<td>PH not satisfied</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| General Model | $h_{0g}(t) \exp(\sum_{i=1}^{p} \beta_i X_i)$ | $h_0(t) \exp(\sum_{i=1}^{p} \beta_i X_i + \sum_{i=1}^{p} \delta_i g_i(t))$ |
| Likelihood Ratio (LR) test | $-2 \ln L_R - (-2 \ln L_F)$ | $-2 \ln L_R - (-2 \ln L_F)$ |

95% Confidence Interval for Hazard Ratio, $HR = \exp(\ell)$ where $\ell = \hat{\beta}_1 + \sum_{i=1}^{k} \hat{\delta}_i W_i$:

$\exp(\ell + 1.96 \sqrt{\text{Var}(\ell)})$ where $\text{Var}(\ell) = \text{Var}(\hat{\beta}_1 + \sum_{i=1}^{k} \hat{\delta}_i W_i)$

Adjusted survival curve.

$S(t, X) = \exp \left[ - \int_{0}^{t} h(u) du \right] = \exp \left[ - \int_{0}^{t} h_0(u) \exp(\sum_{i=1}^{p} \beta_i X_i) du \right] = \exp \left[ - \exp(\sum_{i=1}^{p} \beta_i X_i) \int_{0}^{t} h_0(u) du \right]$

$= \exp \left[ - \int_{0}^{t} h_0(u) du \right] \exp(\sum_{i=1}^{p} \beta_i X_i) = [S_0(t)]^{\exp(\sum_{i=1}^{p} \beta_i X_i)}$

**KK4. Methods for checking PH assumptions**

<table>
<thead>
<tr>
<th>Method</th>
<th>Ideas</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Graphical</td>
<td>a) $\ln(-\ln S(t))$ vs t</td>
<td>$\ln(-\ln S(t)) = \sum_{i=1}^{p} \beta_i X_i + \ln(-\ln S_0(t))$ a linear function</td>
</tr>
<tr>
<td>b) Obs vs predicted S(t)</td>
<td>$h_0(t) \exp(\sum_{i=1}^{p} \beta_i X_i + \sum_{i=1}^{p} \delta_i g_i(t))$ Test for $H_0$: $\delta_1 = \delta_2 = \cdots = \delta_p = 0$ using LR with $\chi^2_p$</td>
<td></td>
</tr>
<tr>
<td>2) Time dependent covariate</td>
<td>interaction terms: $X \times g(t)$</td>
<td></td>
</tr>
<tr>
<td>3) Goodness of fit</td>
<td>large sample $Z$ test</td>
<td>Schoenfeld Residuals. Use $p$-values</td>
</tr>
</tbody>
</table>

If PH assumption not met, use stratified Cox or Cox with time-dependent covariates.

**KW11. Estimation of Complete Data**

**Definition 1** (D1.11) A data-dependent distribution is at least as complex as the data or knowledge that produced it, and the number of "parameters" increases as the number of data points or amount of knowledge increases.

**Definition 2** (D1.12) A parametric distribution is a set of distribution functions, each member of which is determined by specifying one or more values called parameters. The number of parameters is fixed and finite.

**Definition 3** (D1.13) The empirical distribution is obtained by assigning probability 1/n to each data point.

**Definition 4** (D1.14) A kernel smoothed distribution is obtained by replacing each data point with a continuous random variable and then assigning probability 1/n to each such random variable. The random variables used must be identical except for a location or scale change that is related to its associated data point.

**Definition 5** (D1.15) The empirical distribution function is $F_n(x) = \frac{\text{number of observations} \leq x}{n}$, when $n$ is the total number of observations.
Definition 6 (11.6) The cumulative hazard rate function is defined as \( H(x) = -\ln S(x) \). The name comes from the fact that, if \( S(x) \) is differentiable, \( H'(x) = -\frac{S'(x)}{S(x)} = \frac{f(x)}{S(x)} = h(x) \), and then \( H(x) = \int_{-\infty}^{x} h(y)dy \).

Definition 7 (11.7) The Nelson-Åalen estimate of the cumulative hazard rate function is

\[
\hat{H}(x) = \begin{cases} 
0, & x < y_1 \\
\sum_{i=1}^{j-1} \frac{s_i}{r_i}, & y_{j-1} \leq x < y_j, \ j = 2, \ldots, k \\
\sum_{i=1}^{k} \frac{s_i}{r_i}, & x \geq y_k
\end{cases}
\]

where the risk set \( r_i = \sum_{j=i}^{k} s_j \) = number of observations \( \geq y_i \).

Definition 8 (11.8) For grouped data, the distribution function obtained by connecting the values of the empirical distribution function at the group boundaries with straight lines is called the ogive. The formula is

\[
F_n(x) = \frac{c_j - x}{c_j - c_{j-1}} F_n(c_{j-1}) + \frac{x - c_{j-1}}{c_j - c_{j-1}} F_n(c_j), \quad c_{j-1} \leq x \leq c_j.
\]

Definition 9 (11.9) For grouped data, the empirical density function can be obtained by differentiating the ogive. The resulting function is called a histogram. The formula is

\[
f_n(x) = \frac{F_n(c_j) - F_n(c_{j-1})}{c_j - c_{j-1}} = \frac{n_j}{n(c_j - c_{j-1})}, \quad c_{j-1} \leq x \leq c_j.
\]

KPW12. Estimation of Modified Data (See also KK2)

Definition 10 (12.1) An observation is truncated from below (also called left truncated) at \( d \) if when it is at or below \( d \) it is not recorded, but when it is above \( d \) it is recorded at its observed value.

- An observation is truncated from above (also called right truncated) at \( u \) if when it is at or above \( u \) it is not recorded, but when it is below \( u \) it is recorded at its observed value.

- An observation is censored from above (also called left censored) at \( d \) if when it is at or below \( d \) it is recorded as being equal to \( d \), but when it is above \( d \) it is recorded at its observed value.

- An observation is censored from above (also called right censored) at \( u \) if when it is at or above \( u \) it is recorded as being equal to \( u \), but when it is below \( u \) it is recorded at its observed value.

\[
r_j = (\text{number of } d_i s < y_j) - (\text{number of } x_i s < y_j) - (\text{number of } u_i s < y_j) \quad \quad (12.1)
\]

\[
r_j = r_{j-1} + (\text{number of } d_i s \text{ between } y_{j-1} \text{ and } y_j) \\
- (\text{number of } x_i s \text{ equal to } y_{j-1}) \\
- (\text{number of } u_i s \text{ between } y_{j-1} \text{ and } y_j) \quad \quad (12.2)
\]

\[
s_j = \# \text{ of time the uncensored event } y_j \text{ occurs in the sample.}
\]

Kaplan-Meier estimate \( S_n(x) = \sum_{i=1}^{j-1} \left( \prod_{i=1}^{j-1} \frac{r_i - s_i}{r_i} \right) \), \( y_{j-1} \leq x < y_j, \ j = 2, \ldots, k, \)

\[
\sum_{i=1}^{k} \left( \frac{r_i - s_i}{r_i} \right) \text{ or } 0, \quad t \geq y_k
\]

Greenwood’s approximation formula: \( \text{Var}[S_n(y_j)] = S_n(y_j)^2 \sum_{i=1}^{j} \frac{s_i}{r_i(r_i - s_i)}. \quad (12.3) \)

Definition 11 (12.2) A kernel density estimator of a distribution function is \( \hat{F}(x) = \sum_{j=1}^{k} p(y_j) K_{y_j}(x) \)

and the estimator of the density function is \( \hat{f}(x) = \sum_{j=1}^{k} p(y_j) k_{y_j}(x) \),
There exists a function $H$ such that the location where the derivative is zero is a maximum.

Then the following results hold:

(i) $\ln (\text{the population})$ makes sure that the probability is not overpopulated.

(ii) $\text{the equations can also be written as}$

(iii) $\text{the vector that maximizes the likelihood function}$ is three times differentiable with respect to $\theta$.

(iv) $\text{the equation is}$

(v) $\text{the population is not overpopulated}$ with regard to extreme values.

Then the following results hold:

(a) As $n \to \infty$, the probability that the likelihood equation $|L'(\theta) = 0|$ has a solution goes to 1.

(b) As $n \to \infty$, the distribution of the mle $\hat{\theta}_n$ converges to a normal distribution with mean $\theta$ and variance $\text{of a percentile is calculated as}$ $\hat{\pi}_k = \frac{1}{\text{the sample}}$.

### Definition 12 (12.3) The following defines 3 popular kernel smoothing methods:

<table>
<thead>
<tr>
<th>Uniform kernel</th>
<th>Triangular kernel</th>
<th>Gamma kernel</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_y(x)$</td>
<td>$k_y(x)$</td>
<td>$k_y(x)$</td>
</tr>
<tr>
<td>$\begin{cases} 0, &amp; x &lt; y - b, \ 1, &amp; \frac{1}{2b}, &amp; y - b \leq x \leq y + b, \ 0, &amp; x &gt; y + b, \end{cases}$</td>
<td>$\begin{cases} 0, &amp; x &lt; y - b, \ \frac{x - y + b}{b^2}, &amp; y - b \leq x \leq y, \ \frac{y + b - x}{b^2}, &amp; y \leq x \leq y + b, \ 0, &amp; x &gt; y + b, \end{cases}$</td>
<td>$\begin{cases} x^{a - 1}e^{-x^a/y} \ (y/\alpha)^a \Gamma(\alpha) \end{cases}$</td>
</tr>
<tr>
<td>$K_y(x)$</td>
<td>$K_y(x)$</td>
<td>$K_y(x)$</td>
</tr>
<tr>
<td>$\begin{cases} 0, &amp; x &lt; y - b, \ \frac{x - y + b}{2b^2}, &amp; y - b \leq x \leq y + b, \ 1, &amp; x &gt; y + b, \end{cases}$</td>
<td>$\begin{cases} 0, &amp; x &lt; y - b, \ \frac{(x - y + b)^2}{2b^2}, &amp; y - b \leq x \leq y, \ \frac{2b}{(y + b - x)^2}, &amp; y \leq x \leq y + b, \ 1, &amp; x &gt; y + b, \end{cases}$</td>
<td>$\begin{cases} \text{Gamma kernel has mean of } \alpha (y/\alpha) = y \text{variance of } \alpha (y/\alpha)^2 = y^2/\alpha \end{cases}$</td>
</tr>
</tbody>
</table>

**KPW13. Frequentist Estimation**

**Definition 13 (13.1)** A method-of-moments estimate of $\theta$ is any solution of the $p$ equations $\mu'_k(\theta) = \hat{\mu}'_k$, $k = 1, 2, ..., p$.

**Definition 14 (13.2)** A percentile matching estimate of $\theta$ is any solution of the $p$ equations $\pi_{g_k}(\theta) = \hat{\pi}_{g_k}$, $k = 1, 2, ..., p$, where $g_1, g_2, ..., g_p$ are $p$ arbitrarily chosen percentiles. From the definition of percentile, the equations can also be written as $F(\hat{\pi}_{g_k}|\theta) = \pi_k$, $k = 1, 2, ..., p$.

**Definition 15 (13.3)** The smoothed empirical estimate of a percentile is calculated as $\hat{\pi}_g = (1-h)x_{(j)} + hx_{(j+1)}$, where $j = \lfloor(n + 1)g\rfloor$ and $h = (n + 1)g - j$. Here $\lfloor\cdot\rfloor$ indicates the greatest integer function and $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$ are the order statistics from the sample.

**Definition 16 (13.4)** The likelihood function is $L(\theta) = \prod_{j=1}^{n} \text{Pr}(X_j \in A_j|\theta)$ and the maximum likelihood estimate of $\theta$ is the vector that maximizes the likelihood function.

**Theorem 17 (T13.5)** Assume that the pdf (pf in the discrete case) $f(x; \theta)$ satisfies the following for $\theta$ in an interval containing the true value (replace integrals by sums for discrete variables):

(i) $\ln f(x; \theta)$ is three times differentiable with respect to $\theta$.

(ii) $\int \frac{\partial}{\partial \theta} f(x; \theta)dx = 0$. This formula implies that the derivatives may be taken outside the integral and so we are just differentiating the constant 1.

(iii) $\int \frac{\partial^2}{\partial \theta^2} f(x; \theta)dx = 0$. This formula is the same concept for the second derivative.

(iv) $-\infty < \int f(x; \theta) \frac{\partial^2}{\partial \theta^2} \ln f(x; \theta)dx < 0$. This inequality establishes that the indicated integral exists and that the location where the derivative is zero is a maximum.

(v) There exists a function $H(x)$ such that $\int H(x) f(x; \theta)dx < \infty$ with $\left| \frac{\partial^3}{\partial \theta^3} \ln f(x; \theta) \right| < H(x)$. This inequality makes sure that the population is not overpopulated with regard to extreme values.

Then the following results hold:

(a) As $n \to \infty$, the probability that the likelihood equation $|L'(\theta) = 0|$ has a solution goes to 1.

(b) As $n \to \infty$, the distribution of the mle $\hat{\theta}_n$ converges to a normal distribution with mean $\theta$ and variance $\text{of a percentile is calculated as}$ $\hat{\pi}_g = (1-h)x_{(j)} + hx_{(j+1)}$, where $j = \lfloor(n + 1)g\rfloor$ and $h = (n + 1)g - j$. Here $\lfloor\cdot\rfloor$ indicates the greatest integer function and $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$ are the order statistics from the sample.
such that $I(\theta)\text{Var}(\hat{\theta}_n) \to 1$, where the Fisher’s information

$$I(\theta) = -nE\left[\frac{\partial^2}{\partial \theta^2} \ln f(X; \theta)\right] = -n \int f(x; \theta) \frac{\partial^2}{\partial \theta^2} \ln f(x; \theta) dx$$

$$= nE \left[ \left( \frac{\partial}{\partial \theta} \ln f(X; \theta) \right)^2 \right] = n \int f(x; \theta) \left( \frac{\partial}{\partial \theta} \ln f(x; \theta) \right)^2 dx.$$

That is, $\lim_{n \to \infty} \text{Pr} \left( \frac{\hat{\theta}_n - \theta}{\sqrt{I(\theta)}} \right) \to N(0, 1)$.

**Theorem 18 (T13.6- Delta Method)** Let $X_n = (X_{1n}, ..., X_{kn})^T$ be a multivariate random variable of dimension $k$ based on a sample of size $n$. Assume that $X$ is asymptotically normal with mean $\mu$ and covariance matrix $\Sigma/n$, where neither $\theta$ nor $\Sigma$ depend on $n$. Let $g$ be a function of $k$ variables that is totally differentiable. Let $G_n = g(X_{1n}, ..., X_{kn})$. Then $G_n$ is asymptotically normal with mean $g(\theta)$ and variance $(\partial g)^T \Sigma (\partial g)/n$, where $\partial g$ is the vector of first derivatives, that is, $\partial g = (\partial g/\partial \theta_1, ..., \partial g/\partial \theta_k)^T$ and it is to be evaluated at $\theta$, the true parameters of the original random variable.

**KPW14. Frequentist Estimation for Discrete Distributions**

**Negative Binomial**: The moment equation is $r \beta = \sum_{k=0}^{\infty} k \frac{ kn_k }{ n } = \bar{x}$. (14.1)

and $r \beta (1 + \beta) = \sum_{k=0}^{\infty} k^2 \frac{ kn_k }{ n } - \left( \sum_{k=0}^{\infty} k \frac{ kn_k }{ n } \right)^2 = s^2$. (14.2)

$$\frac{\partial l}{\partial \beta} = \sum_{k=0}^{\infty} n_k \left( k \beta - \frac{ r + k }{ 1 + \beta } \right).$$ (14.3)

and $\frac{\partial l}{\partial r} = -n \ln(1 + \beta) + \sum_{k=0}^{\infty} n_k \ln (1 + k m) = -n \ln(1 + \beta) + \sum_{k=0}^{\infty} n_k \frac{ k-1 }{ m-1 } \ln (r + m)$

$$= -n \ln(1 + \beta) + \sum_{k=0}^{\infty} n_k \sum_{m=0}^{k-1} \frac{ 1 }{ r + m }.$$ (14.4)

Setting these equations to zero yields $\hat{\mu} = \hat{r} \hat{\beta} = \sum_{k=0}^{\infty} k \frac{ kn_k }{ n } = \bar{x}$ (14.5)

and $n \ln(1 + \beta) = \sum_{k=0}^{\infty} n_k \left( \sum_{m=0}^{k-1} \frac{ 1 }{ r + m } \right).$ (14.6)

$H(\bar{r}) = n \ln \left( 1 + \frac{ \bar{x} }{ \bar{r} } \right) - \sum_{k=0}^{\infty} n_k \left( \sum_{m=0}^{k-1} \frac{ 1 }{ r + m } \right) = 0$ (14.7)

**Binomial**: $\bar{q} = \frac{ 1 }{ m } \sum_{k=0}^{\infty} k \frac{ kn_k }{ n }.$ (14.8)

The (a,b,1) class: $\bar{x} \left( 1 - e^{-\lambda} \right) = \frac{ n - n_0 }{ n } \lambda$. (14.9) $\bar{x} = \frac{ 1 - \hat{\rho}_0^M }{ 1 - \hat{\rho}_0 } \lambda$. (14.10)

**Zero-modified Binomial**: $\bar{x} = \frac{ 1 - \hat{\rho}_0^M }{ 1 - \hat{\rho}_0 } mq$. (14.11)

$l_1 = \sum_{k=1}^{\infty} n_k \ln p_k - (n - n_0) \ln(1 - p_0)$, (14.12)

$$Hence, \quad l_1 = \sum_{k=1}^{\infty} n_k \ln \left[ \binom{ k + r - 1 }{ k } \left( \frac{ 1 }{ 1 + \beta } \right)^r \left( \frac{ \beta }{ 1 + \beta } \right)^k \right]$$

$$- (n - n_0) \ln \left[ 1 - \left( \frac{ 1 }{ 1 + \beta } \right)^r \right]. \quad (14.13)$$
\[ g_k = \frac{\lambda}{k} \sum_{j=1}^{k} j f_j g_{k-j}, \quad k = 1, 2, 3, \ldots \]  
(14.14) where \( f_j = \beta^{j-1}/(1 + \beta)^j \), \( j = 1, 2, 3, \ldots \)

**KPW15. Bayesian Estimation**

**Definition 19** \((D15.1)\) **Prior distribution** \( \pi(\theta) \) is a probability distribution over the space of parameter values. It represents our opinion about the relative chances various \( \theta \) values are the true parameter value.

**Definition 20** \((D15.2)\) **Improper prior distribution** is one for which the probabilities (or pdf) are non-negative but their sum (or integral) is infinite.

**Definition 21** \((D15.3)\) The **model distribution** \( f_{X|\Theta}(x|\theta) \) is the probability distribution for the data given a particular value of the parameter.

**Definition 22** \((D15.4)\) The **joint distribution** \( f_{X,\Theta}(x,\theta) \) has pdf \( f_{X,\Theta}(x,\theta) = f_{X|\Theta}(x|\theta)\pi(\theta) \).

**Definition 23** \((D15.5)\) The **marginal distribution** of \( X \) has pdf \( f_X(x) = \int f_{X|\Theta}(x|\theta)\pi(\theta)d\theta \).

**Definition 24** \((D15.6)\) The **Posterior distribution** \( \pi_{\Theta|X}(\theta|x) \) is the conditional probability distribution of parameter values given the observed data.

**Definition 25** \((D15.7)\) The **Predictive distribution** \( f_{Y|X}(y|x) \) is the conditional probability distribution of a new observation \( y \) given the observed data \( x \).

**Theorem 26** \((T15.8)\) The **posterior distribution** can be computed as \( \pi_{\Theta|X}(\theta|x) = \frac{f_{X|\Theta}(x|\theta)\pi(\theta)}{\int f_{X|\Theta}(x|\theta)\pi(\theta)d\theta} \)

while the **predictive distribution** can be computed as \( f_{Y|X}(y|x) = \int f_{Y|\Theta}(y|\theta)\pi_{\Theta|X}(\theta|x)d\theta \),

where \( f_{Y|\Theta}(y|\theta) \) is the pdf of the new observation given the parameter value.

**Inference and Prediction**

**Definition 27** \((D15.9)\) A **loss function** \( l_j(\hat{\theta}_j, \theta_j) \) describes the penalty paid by the investigator when \( \hat{\theta}_j = \) estimator while \( \theta_j = \) true value of the \( j \)th parameter.

**Definition 28** \((D15.10-12)\) The **Bayes estimate** for a given loss function is the one that minimizes the expected loss **given the posterior distribution** of the parameter in question.

| loss function \( l_j(\hat{\theta}_j, \theta_j) \)       | \( (\hat{\theta}_j - \theta_j)^2 \) | \( |\hat{\theta}_j - \theta_j| \) | \( 0 \) if \( \hat{\theta}_j = \theta_j \) | \( 1 \) if \( \hat{\theta}_j \neq \theta_j \) |
|-----------------------------------------------------|-----------------------------------|---------------------------------|-----------------|-----------------|
| Bayes estimate                                     | mean                              | median                          | mode of \( \pi_{\Theta|X}(\theta|x) \) |

**Definition 29** \((D5.13)\) The points \( a < b \) defines a \( 100(1 - \alpha)\% \) **credibility interval** for \( \theta_j \) provided that \( \Pr(a < \Theta_j < b|x) \geq 1 - \alpha \).

**Theorem 30** \((T15.14)\) If the posterior random variable \( \theta_j|x \) is continuous and unimodal, then the \( 100(1 - \alpha)\% \) credibility interval with the smallest width \( b-a \) is the unique solution to

\[
\int_a^b \pi_{\Theta_j|X}(\theta_j|x)d\theta_j = 1 - \alpha
\]

\[
\pi_{\Theta|X}(a|x) = \pi_{\Theta|X}(b|x).
\]

The interval is a special case of a highest posterior density (HPD) credibility set.

**Definition 31** \((D15.15)\) For any posterior distribution, the \( (1 - \alpha)100\% \) **HPD credibility set** is the set of parameter values \( C \) such that \( \Pr(\theta_j \in C) \geq 1 - \alpha \)

and \( C = \{ \theta_j : \pi_{\Theta_j|X}(\theta_j|x) \geq c \} \) for some \( c \) where \( c \) is the largest value for which the probability inequality holds.

**Theorem 32** \((T15.16)\) **Bayesian Central Limit Theorem** If \( \pi(\theta) \) and \( f_{X|\Theta}(x|\theta) \) are both twice differentiable in the elements of \( \Theta \) and other commonly satisfied assumptions hold, then the posterior distribution of \( \Theta \) given \( X = x \) is asymptotically normal. (see Theorem T13.5 for commonly satisfied assumptions).
Conjugate prior distributions

**Definition 33** (D15.17) A prior distribution is said to be a *conjugate prior distribution* for a given model if the resulting posterior distribution is from the same family as the prior (but perhaps with different parameters).

**Theorem 34** (T15.18) Suppose for Θ = θ, the random variables X_1, X_2, ⋯, X_n are i.i.d. with pdf \( f_{X_j|θ}(x_j|θ) = \frac{q(θ)}{c(μ, k)} e^{-k r(θ)} r(θ) \)

where Θ has pdf \( π(θ) = \frac{[q(θ)]^{-k} e^{μk r(θ)} r(θ)}{c(μ, k)} \)

where \( k \) and \( μ \) are parameters of the distribution and \( c(μ, k) \) is the normalizing constants. Then the posterior pdf \( π_Θ|X(θ|x) \) is of the same form as \( π(θ) \).

**KPW16. Model Selection**

**Models and Data Representations**

\[
F^*(x) = \begin{cases} 0 & x < t, \\ \frac{F(x) - F(t)}{1 - F(t)} & x \geq t. \end{cases}
\]

**Graphical comparison of models to data:** Check discrepancies

1) Empirical and model plot \((F_n(x) \text{ vs } x \text{ plot})\)
2) Deviation plot \((D(x) = F_n(x) - F^*(x) \text{ vs } x \text{ plot})\)
3) Probability \( p - p \) plot: check for straight 45° line

**Hypothesis tests**

A) \( H_0: \text{ Data came from population with stated model} \)
B) \( H_a: \text{ Data did not come from such population} \)

\( D \leq CV \) don’t reject \( H_0 \)
\( D > CV \) reject \( H_0 \), where

\[
\begin{array}{ccc|c|c|c}
\alpha & 0.10 & 0.05 & 0.01 \\
\hline
\text{critical value} & 1.22/\sqrt{n} & 1.36/\sqrt{n} & 1.63/\sqrt{n} \\
\end{array}
\]

(1) Kolmogorov-Smirnov (KS) Test: Statistic \( D = \max_{t_j < x < u_j} |F_n(x) - F^*(x)| \)

where

\( t = \text{left truncation point (} t = 0 \text{ if no truncation)} \)
\( u = \text{right censoring point (} u = \infty \text{ if no censoring)} \)

(2) Anderson-Darling (AD) Test: Statistic \( A^2 = n \int_0^u \frac{[F_n(x) - F^*(x)]^2}{F^*(x) [1 - F^*(x)]} f^*(x) \, dx \)

\[
A^2 = -nF^*(u) + n \sum_{j=0}^k [1 - F_n(y_j)]^2 \left\{ \ln [1 - F^*(y_j)] - \ln [1 - F^*(y_{j+1})] \right\} + n \sum_{j=1}^k F_u(y_j) [\ln F^*(y_{j+1}) - \ln F^*(y_j)]
\]

If \( A^2 \leq CV \) don’t reject \( H_0 \)
\( A^2 > CV \) reject \( H_0 \), where

\[
\begin{array}{ccc|c|c|c}
\alpha & 0.10 & 0.05 & 0.01 \\
\hline
\text{critical value} & 1.933 & 2.492 & 3.857 \\
\end{array}
\]

(3) Chi-Square goodness of fit (GoF) Test: Statistic \( \chi^2_f = \sum_{g=1}^k \frac{n(\hat{p}_g - p_{ng})^2}{\hat{p}_g} = \sum_{g=1}^k \frac{(E_g - O_g)^2}{E_g} \)

where

\( t = c_0 < c_1 < \cdots < c_k < u \leq \infty \), \( \hat{p}_g = F^*(c_g) - F^*(c_{g-1}) \), \( p_{ng} = F_n(c_g) - F_n(c_{g-1}) \),
\( E_g = \hat{n}_g \hat{p}_g \), \( O_g = \hat{n}_p p_{ng}, \) \( df = k - 1 - \# \text{ parameter} \).

If \( \chi^2_f \leq CV \) don’t reject \( H_0 \) where \( CV = \chi^2_{df,1-α} \) is from a \( \chi^2 \) table.

B) \( H_a: \text{ Data came from population with distribution model A} \)

**Selection of Models**

1) Use a simple model if possible
2) Restrict universe of potential models

A) Judgement-based approach

B) Score-based approach

Some scores worth considering:

- \( CV \) = Likelihood function maximized under \( H_0 \)
- \( L_a \) = Likelihood function maximized under \( H_0 \).
- \( \chi^2_{df,1-α} \) is from a \( \chi^2 \) table and \( df = \# \text{ parameter}_{H_a} - \# \text{ parameter}_{H_0} \).
a) Lowest value of Kolmogorov-Smirnov statistic
b) Lowest value of Anderson-Darling statistic
c) Lowest value of Chi-square goodness of fit statistic
d) Highest p-value for the Chi-square goodness of fit statistic
e) Highest value of the likelihood function at its maximum.

**KK7. Parametric Survival Models**

<table>
<thead>
<tr>
<th></th>
<th>Weibull</th>
<th>Exponential</th>
<th>Log-logistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>( h_0(t) )</td>
<td>( pt^{p-1} \exp(\beta_0) )</td>
<td>( \exp(\beta_0) )</td>
<td>complicated form</td>
</tr>
<tr>
<td>( h(t, X) )</td>
<td>( \lambda t^{p-1} )</td>
<td>( \lambda )</td>
<td>( \lambda t^{p-1} )</td>
</tr>
<tr>
<td></td>
<td>( p &lt; 1 ) \quad \text{decreasing}</td>
<td>( p = 1 ) \quad \text{constant}</td>
<td>( p \leq 1 ) \quad \text{decreasing}</td>
</tr>
<tr>
<td></td>
<td>( p &gt; 1 ) \quad \text{increasing}</td>
<td>( \text{Weibull}(p = 1) )</td>
<td>( p &gt; 1 ) \quad \text{increasing}</td>
</tr>
<tr>
<td><strong>PH form</strong></td>
<td>( \lambda = \exp(\beta_0 + \sum \beta_i X_i) )</td>
<td>( \lambda = \exp(\beta_0 + \sum \beta_i X_i) )</td>
<td>( \lambda = \exp(\beta_0 + \sum \beta_i X_i) )</td>
</tr>
<tr>
<td><strong>PO form</strong></td>
<td>( S(t) ) \quad \exp(-\lambda t^p) )</td>
<td>( \exp(-\lambda) )</td>
<td>( \frac{1}{1 + \lambda t^p} )</td>
</tr>
<tr>
<td></td>
<td>( HR(\text{TRT} = 1 \text{ vs } 0) )</td>
<td>( \exp(\beta_1) )</td>
<td>( \frac{1}{1 + \lambda t^p} )</td>
</tr>
<tr>
<td>ln\left[\frac{1 - \ln S(t)}{\ln(\lambda) + p \ln(t)}\right]</td>
<td>( S(t) ) \quad \exp(-\lambda t^p) )</td>
<td>( \exp(-\lambda) )</td>
<td>( \frac{1}{1 + \lambda t^p} )</td>
</tr>
</tbody>
</table>

\[
\begin{align*}
\alpha_i \text{ vs } \beta_i & \\
\beta_i = -\alpha_i p & \\
\beta_i = -\alpha_i & \\
\beta_i = -\alpha_i p & \\
\gamma & = \exp(\alpha_0) \\
\gamma & = \exp(\alpha_0) \\
\gamma & = \exp(\alpha_0) \\
\end{align*}
\]

\[
\gamma = \exp(\alpha_0) \quad AFT \Rightarrow PH \quad AFT \Leftrightarrow PO
\]

<table>
<thead>
<tr>
<th>General form</th>
<th>LogNormal</th>
<th>Gompertz</th>
</tr>
</thead>
<tbody>
<tr>
<td>( h_0(t) )</td>
<td>( \exp(\gamma t) )</td>
<td>( \lambda \exp(\gamma t) )</td>
</tr>
<tr>
<td>( h(t, X) )</td>
<td>( \gamma &lt; 0 ) \quad \text{exponentially decreasing}</td>
<td>( \gamma = 0 ) \quad \text{constant}</td>
</tr>
<tr>
<td></td>
<td>( \gamma &gt; 0 ) \quad \text{exponentially increasing}</td>
<td>( \gamma = 0 ) \quad \text{constant}</td>
</tr>
<tr>
<td><strong>PH form</strong></td>
<td>( t = \exp(\alpha_0 + \sum \alpha_i X_i + \sigma \epsilon) )</td>
<td>( t = \exp(\alpha_0 + \sum \alpha_i X_i + \sigma \epsilon) )</td>
</tr>
<tr>
<td>( AFT )</td>
<td>( t = \exp(\alpha_0 + \sum \alpha_i X_i + \sigma \epsilon) )</td>
<td>( t = \exp(\alpha_0 + \sum \alpha_i X_i + \sigma \epsilon) )</td>
</tr>
</tbody>
</table>

\[
\begin{align*}
\epsilon & \sim N(0, 1) \\
\end{align*}
\]

**Frailty Models:** \( h_j(t, X|\alpha_j) = \alpha_j h(t, X) \) \( j = 1, 2, \ldots, n \) with \( \mu_\alpha = 1 \) and variance \( \sigma_\alpha = \theta \)

model with Gamma frailty: \( \alpha \sim \text{gamma} \) (\( \mu_\alpha = 1 \), variance \( \sigma_\alpha = \theta \))

Weibull with gamma frailty \( HR(\text{TRT}=2 \text{ vs } 1) = \begin{cases} \exp(\beta_1) & \alpha_1 = \alpha_2 \\ \frac{\alpha_1}{\alpha_2} \exp(\beta_1) & \alpha_1 \neq \alpha_2 \end{cases} \)

unconditional hazard with gamma frailty: \( h_U(t, X) = \frac{h(t)}{1 - \theta \ln S(t)} \)

**KK8. Recurrent Event Survival Analysis**

Events can occur more than 1 times during study
1. Counting Process (CP) with Cox PH model
2. Stratified Cox PH models
3. Parametric with frailty model
4. Counting Process (CP) with Cox PH model
standard cox: \( h(t, X) = h_0(t) \exp(\sum \beta_i X_i) \)

likelihood function is different than nonrecurrent event (subjects remains in the risk set until last interval for follow-up)

Robust estimation for variance estimators: \( \hat{\mathbf{R}}(\hat{\beta}) = \hat{\text{Var}}(\hat{\beta}) | \hat{\mathbf{R}}_S \hat{\mathbf{R}}_S \hat{\text{Var}}(\hat{\beta}) \) where \( \hat{\text{Var}}(\hat{\beta}) \) = information matrix and \( \hat{\mathbf{R}}_S \) = matrix of score residuals.

(2) Stratified Cox PH models for recurrent times

time interval = strata

no interaction stratified cox: \( h_g(t, X) = h_{0g}(t) \exp(\sum \beta_i X_i) \) or interaction stratified cox: \( h_g(t, X) = h_{0g}(t) \exp(\sum \beta_{ig} X_i) \)

Robust estimation for variance estimators

(a) Stratified Counting Process approach: time interval = time from \((k-1)^{st}\) to \(k^{th}\) event

(b) Gap Time approach: time interval = additional time between 2 recurrent events

(c) Marginal Time approach: time interval = total time to \(k^{th}\) event

(3) Parametric with shared frailty model

Survival curves with recurrent events: on one ordered event at a time.

\( S_k(t) = Pr(T_k > t) \) where \( T_k \) = survival time up to occurrence of \(k^{th}\) event.

a) Stratified \( S_{kc}(t) = Pr(T_{kc} > t) \) where \( T_{kc} \) = time from \((k-1)^{st}\) to \(k^{th}\) event: restricts data to subjects with \((k-1)\) events.

b) Marginal \( S_{km}(t) = Pr(T_{km} > t) \) where \( T_{km} \) = time from study entry to \(k^{th}\) event: ignores previous events.

**KK9. Competing Risk Survival Analysis**

Only one event of different type can occur to a subject during study: Events compete with each other.

Usually one event is death. Example: Accidental, Illness vs natural death.

(1) Separate models for each event type (2) Lunn-McNeil (LM) approach

(1) Separate models for each event type

Use Cox PH model for each hazard separately while treating other competing risks as censored.

cause-specific hazard function: \( h_c(t) = \lim_{\Delta t \to 0} P(t \leq T_c \leq t + \Delta t) / \Delta t \) where \( T_c \) = time to failure from event \( c \), \( c = 1, 2, \cdots, C \).

cause-specific model: \( h_c(t, X) = h_{0c}(t) \exp(\sum_{i=1}^p \beta_{ic} X_i) \) \( c = 1, 2, \cdots, C \).

**Independence Assumptions:** Independent censoring. Competing risks are independent.

**Cumulative Incidence Curves (CIC):** KM curves may not be informative.

alternative to KM curves for competing risks. \( CIC(t_f) = \sum_{\ell=1}^f \hat{I}_c(t_f) = \sum_{\ell=1}^f \hat{S}(t_f - 1) \hat{h}_c(t_f) \)

Conditional Probability Curves (CPC): \( CPC_c = P(T_c \leq t | T \geq t) \) where \( T_c \) = time until event \( c \) occurs while \( T \) = time until any competing risk event occurs

\( CPC_c = CIC_c / (1 - CIC_c) \)

(a) Pepe & Mori (1993) test for 2 CPC curves (b) Lunn (1998) test for \( g \) CPC curves

(2) Lunn-McNeil (LM) approach

uses an augmented data layout