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

<table>
<thead>
<tr>
<th>Continuous</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^{\lambda x} - e^{\lambda a}}{t(b - a)} )</td>
<td>( \frac{b + a}{2} )</td>
<td>( \frac{(b - a)^2}{12} )</td>
</tr>
<tr>
<td>Exponential((\beta))</td>
<td>( f(x) = \begin{cases} \frac{1}{\beta} e^{-\lambda x/\beta} &amp; x \geq 0 \ 0 &amp; x &lt; 0 \end{cases} )</td>
<td>( \frac{\lambda}{\lambda - t} )</td>
<td>( \beta )</td>
<td>( \beta^2 )</td>
</tr>
<tr>
<td>Gamma((\alpha, \beta))</td>
<td>( f(x) = \begin{cases} \frac{e^{-x/\beta} x^{\alpha-1}}{\beta^\alpha \Gamma(\alpha)} &amp; x \geq 0 \ 0 &amp; x &lt; 0 \end{cases} )</td>
<td>( \left( \frac{\lambda}{\lambda - t} \right)^\alpha )</td>
<td>( \alpha\beta )</td>
<td>( \alpha\beta^2 )</td>
</tr>
<tr>
<td>Normal((\mu, \sigma^2))</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{\alpha \theta^\alpha}{x^{\alpha+1}} )</td>
<td>( M_X(t) ) not given</td>
<td>( \frac{\theta^k \Gamma(\alpha - k)}{\Gamma(\alpha)} )</td>
<td>( \frac{\alpha \theta^2}{(\alpha - 1)^2(\alpha - 2)} )</td>
</tr>
<tr>
<td>Right</td>
<td>( F(x) = 1 - \left( \frac{\theta}{x + \theta} \right)^\alpha )</td>
<td>( E[X^k] = \frac{\theta^k \Gamma(\alpha - k)}{\Gamma(\alpha)} )</td>
<td>( \frac{\theta}{\alpha - 1} )</td>
<td>( \frac{\alpha \theta^2}{(\alpha - 1)^2(\alpha - 2)} )</td>
</tr>
<tr>
<td>Single Pareto((\alpha, \theta))</td>
<td>( f(x) = \frac{\alpha \theta^\alpha}{x^{\alpha+1}} )</td>
<td>( M_X(t) ) not given</td>
<td>( \frac{\theta^k \Gamma(\alpha - k)}{\Gamma(\alpha)} )</td>
<td>( \frac{\alpha \theta^2}{(\alpha - 1)^2(\alpha - 2)} )</td>
</tr>
<tr>
<td>Left</td>
<td>( F(x) = 1 - \left( \frac{\theta}{x + \theta} \right)^\alpha )</td>
<td>( E[X^k] = \frac{\theta^k \Gamma(\alpha - k)}{\Gamma(\alpha)} )</td>
<td>( \frac{\theta}{\alpha - 1} )</td>
<td>( \frac{\alpha \theta^2}{(\alpha - 1)^2(\alpha - 2)} )</td>
</tr>
</tbody>
</table>

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

**KK1 Introduction to Survival Analysis**

- **Time** = event time
- **Event** = failure
- **Left-censored** = true survival time ≤ observed survival time
- **Right-censored** = true survival time ≥ observed survival time
- **Interval-censored** = true survival time is within a known interval
time

Left censoring \( t_1 = 0 \); Right censoring \( t_1 = \) lower bound; \( t_2 = \) infinity

\[ d = \begin{cases} 1 & \text{if failure} \\ 0 & \text{censored} \end{cases} \]

\( S(t) = \text{survivor function} \)

\( h(t) = \frac{dS(t)/dt}{S(t)} \)

\( S(t) = \exp \left[ - \int_0^t h(u) du \right] \)

\( \hat{S}(t) = \text{observed survivor function} \)

**Goals of Survival Analysis:**

1. To estimate & interpret survivor &/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: \( T = \frac{1}{n} \sum_{i=1}^{n} t_i \)

**Survival Models for Actuaries Formula**

**O. Preliminary SOA Exam P Formula**

\( \text{Descriptive measures of survival experience: } \text{Average survival time} : T = \frac{1}{n} \sum_{i=1}^{n} t_i \)
<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></td>
<td>regression coefficient $\beta$</td>
<td>odds ratio $e^\beta$</td>
<td>hazard ratio $e^\beta$</td>
</tr>
</tbody>
</table>

**Censoring 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). $S(t_{(f)}) = S(t_{(f-1)})P(T > t_{(f)}|T \geq t_{(f)}) = \prod_{i=1}^{f} P(T > t_{(i)}|T \geq t_{(i)})$

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: $d'V^{-1}d \sim \chi^2_{G-1}, \forall i = 1, 2, \cdots, G$

**Censoring Assumptions:**

- b) Random (vs. non-random) censoring
- c) Non-informative (vs. informative) censoring

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

**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

- $w(t_f) =$ weights at the $f^{th}$ failure time.
- $m_{if} = \text{observed counts for the } i^{th} \text{ group at time } f.$
- $e_{if} = \left(\frac{n_{if}}{n_{if} + n_{jf}}\right) (m_{if} + m_{jf}) = \text{expected counts (proportion in risk set) \times (# failures over both groups)}$

### KK3-KK6: Cox Models

**KK3. Cox PH**

**KK5. Stratified Cox PH**

**KK6. Cox PH for Time-dependent Variables**

<table>
<thead>
<tr>
<th>Model</th>
<th>$h_0(t) \exp(\sum_{i=1}^{p} \beta_i X_i)$</th>
<th>$h_0(t) \exp(\sum_{i=1}^{p} \beta_i X_i)$</th>
<th>$h_0(t) \exp(\sum_{i=1}^{p} \beta_i X_i + \sum_{j=1}^{p_2} \delta_j X_{ij})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>HR: $\frac{h(t, X^* )}{h(t, X)}$</td>
<td>$\exp(\sum_{i=1}^{p} \beta_i (X_i^* - X_i))$</td>
<td>$\exp(\sum_{i=1}^{p} \beta_i (X_i^* - X_i)) + \sum_{j=1}^{p_2} \delta_j (X_{ij}^* - X_{ij})$</td>
<td></td>
</tr>
<tr>
<td>Meaning PH</td>
<td>$\frac{h(t, X^* )}{h(t, X)} = \theta$</td>
<td></td>
<td>PH not satisfied</td>
</tr>
</tbody>
</table>

**General model to assess**

**Interaction:**

- $h_0(t) \exp(\sum_{i=1}^{p} \beta_i X_i)$
- $g = 1, 2, \cdots, k$ strata defined from $Z^*$
- or $h_0(t) \exp(\sum_{i=1}^{p} \beta_i X_i)$
- $+ \sum_{g=1}^{k-1} \sum_{i=1}^{p} \beta_{ig} X_{ig} Z_{g}$

**Likelihood ratio (LR) test**

- $LR \sim \chi^2_{\#parameters \text{ in } F - R}$
- $-2 \ln L_R - (-2 \ln L_F)$
- $LR \sim \chi^2_{p(k-1)}$
- $-2 \ln L_R - (-2 \ln L_F)$
95% Confidence Interval for Hazard Ratio, $HR = \exp(\ell)$ where $\ell = \beta_1 + \sum_{i=1}^{k} \delta_i W_i$:

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

Adjusted survival curve.

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

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></td>
<td>b) Obs vs predicted S(t)</td>
<td>$h_0(t) \exp(\sum_{i=1}^{p} \beta_i X_i + \sum_{i=1}^{p} \delta_i X_i g_i(t) )$ Test for $H_0$: $\delta_1 = \delta_2 = \cdots = \delta_p = 0$ using LR with $\chi^2_p$</td>
</tr>
<tr>
<td>2) Time dependent covariate interaction terms: $X \times g(t)$</td>
<td></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 (D11.1) 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 (D11.2) 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 fixed and finite.

Definition 3 (D11.3) The empirical distribution is obtained by assigning probability 1/n to each data point.

Definition 4 (D11.4) 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.

Definition 5 (11.5) 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 $H(x) = −\ln S(x)$. 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) Where the risk set $r_i = \sum_{j=i}^{k} s_j =$ number of observations $\geq y_i$, the

Nelson-Åalen estimate of cumulative hazard rate function $\hat{H}(x) = \begin{cases} 0, & x < y_1 \\ \sum_{i=1}^{j-1} \frac{s_i}{r_i}, & y_{j-1} \leq x < y_j, \ j = 2, \cdots, k, \\ \sum_{i=1}^{k} \frac{s_i}{r_i}, & x \geq y_k \end{cases}$

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 as below

$$F_n(x) = \frac{c_j - x}{c_j - c_{j-1}} F_n(c_{j-1}) + \frac{x - c_{j-1}}{c_j - c_{j-1}} F_n(c_j), \quad c_{j-1} \leq x \leq c_j.$$
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** as below

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

**Example**

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

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

(12.1)

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

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

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

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

(12.3)

Definition 10 (12.1) Observations can be truncated or censored from above (right) or below (left).

<table>
<thead>
<tr>
<th>Observation from below at ( d ) (or left)</th>
<th>Truncated</th>
<th>Censored</th>
<th>Observation from above at ( u ) (or right)</th>
<th>Truncated</th>
<th>Censored</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x \leq d )</td>
<td>not recorded or missing</td>
<td>( d )</td>
<td>( x &lt; u )</td>
<td>( x )</td>
<td>( u )</td>
</tr>
<tr>
<td>( x &gt; d )</td>
<td>( x )</td>
<td>( x )</td>
<td>( x \geq u )</td>
<td>not recorded or missing</td>
<td>( u )</td>
</tr>
</tbody>
</table>

**Example**

deductible

policy limit

r_j = (number of \( d_i s < y_j \)) - (number of \( x_i s < y_j \)) - (number of \( u_i s < y_j \))

(12.1)

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

(12.2)

s_j = \# of time the uncensored event \( y_j \) occurs in the sample.

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

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

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

(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 \( \hat{f}(x) = \sum_{j=1}^{k} p(y_j) k_{y_j}(x) \).

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_u(x) )</td>
<td>( K_u(x) )</td>
<td></td>
</tr>
<tr>
<td>( \left{ \begin{array}{ll} 0, &amp; x &lt; y - b, \ \frac{1}{2b}, &amp; y - b \leq x \leq y + b, \ 0, &amp; x &gt; y + b, \end{array} \right. )</td>
<td>( \left{ \begin{array}{ll} 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{array} \right. )</td>
<td>( x^{\alpha-1}e^{-x/\theta}/\theta )</td>
</tr>
<tr>
<td>( \frac{1}{2b} )</td>
<td>( \frac{(x - y + b)^2}{2b^2} )</td>
<td>( (y/\alpha)^{\alpha} \Gamma(\alpha) )</td>
</tr>
<tr>
<td>( x &gt; y - b, )</td>
<td>( y - b \leq x \leq y, )</td>
<td>shape ( \alpha ) and scale parameter ( y/\alpha )</td>
</tr>
<tr>
<td>( 1, )</td>
<td>( 1 - \frac{(y + b - x)^2}{2b^2}, )</td>
<td>Gamma kernel has mean ( \alpha (y/\alpha) = y ) &amp; variance ( \alpha (y/\alpha)^2 = y^2/\alpha )</td>
</tr>
<tr>
<td>( x &gt; y + b, )</td>
<td>( y \leq x \leq y + b, )</td>
<td></td>
</tr>
<tr>
<td>( x &gt; y - b, )</td>
<td>( x &gt; y + b, )</td>
<td></td>
</tr>
</tbody>
</table>

**Exposure method**

<table>
<thead>
<tr>
<th>Exact</th>
<th>Exposure definition</th>
<th>( q_j )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actuarial</td>
<td>exposure = exact total time under observation</td>
<td>( q_j = 1 - \exp\left(-d_j/c_j\right) )</td>
</tr>
</tbody>
</table>

**Life insurance Exposure method**

<table>
<thead>
<tr>
<th>Insuring Ages</th>
<th>Exposure definition</th>
<th>( q_j )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anniversary based</td>
<td>based on policy holder’s age at entry</td>
<td>( q_j = d_j/c_j )</td>
</tr>
</tbody>
</table>
Definition 13 (13.1) A method-of-moments estimate of \( \theta \) is any solution of the \( p \) equations \( \hat{\mu}_k^2(\theta) = \bar{\mu}_k^2 \), \( k = 1, 2, ..., p \).

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

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

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

\[
\begin{align*}
& \text{(i)} \ln f(x; \theta) \text{ is three times differentiable with respect to } \theta. \\
& \text{(ii)} \int \frac{\partial}{\partial \theta} f(x; \theta) dx = 0. \text{ This formula implies that the derivatives may be taken outside the integral} \\
& \text{and so we are just differentiating the constant 1.} \\
& \text{(iii)} \int \frac{\partial^2}{\partial \theta^2} f(x; \theta) dx = 0. \text{ This formula is the same concept for the second derivative.} \\
& \text{(iv)} \int f(x; \theta) \frac{\partial^2}{\partial \theta^2} \ln f(x; \theta) dx < 0. \text{ This inequality establishes that the indicated integral exists} \\
& \text{and that the location where the derivative is zero is a maximum.} \\
& \text{(v)} \text{There exists a function } H(x) \text{ such that } \int H(x)f(x; \theta) dx < \infty \text{ with } \left| \frac{\partial^3}{\partial \theta^3} \ln f(x; \theta) \right| < H(x). \text{ This} \\
& \text{inequality makes sure that the population is not overpopulated with regard to extreme values.}
\end{align*}
\]

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 such that \( I(\theta) \text{Var}((\hat{\theta}_n)) \to 1 \), where the Fisher’s information

\[
\begin{align*}
I(\theta) &= -nE \left[ \frac{\partial^2}{\partial \theta^2} \ln f(X; \theta) \right] = -n \int f(x; \theta) \frac{\partial^2}{\partial \theta^2} \ln f(x; \theta) dx \\
&= nE \left[ \left( \frac{\partial}{\partial \theta} \ln f(X; \theta) \right)^2 \right] \\
&= n \int f(x; \theta) \left( \frac{\partial}{\partial \theta} \ln f(x; \theta) \right)^2 dx.
\end{align*}
\]

That is, \( \lim_{n \to \infty} \Pr \left( \frac{\hat{\theta}_n - \theta}{\sqrt{I(\theta)^{-1/2}}} < z \right) = \Phi(z) \).

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 \( \theta \) 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, \( \partial g = (\partial g/\partial \theta_1, \ldots, \partial g/\partial \theta_k)^T \) and it is to be evaluated at \( \theta \), the true parameters of the original random variable.
Negative Binomial: The moment equation are $r\beta = \frac{\sum_{k=0}^{\infty} kn_k}{n} = \tau$.  

$$r(1 + \beta) = \frac{\sum_{k=0}^{\infty} k^2 n_k}{n} - \left( \frac{\sum_{k=0}^{\infty} kn_k}{n} \right)^2 = s^2. \quad (14.2) \quad \frac{\partial l}{\partial \beta} = \sum_{k=0}^{\infty} n_k \left( \frac{k}{\beta} - \frac{r + k}{1 + \beta} \right). \quad (14.3)$$

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

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

Setting these (14.4) to zero yields $\hat{r} = \hat{r} \hat{\beta} = \frac{\sum_{k=0}^{\infty} kn_k}{n} = \tau \quad (14.5)$ and $n \ln(1+\hat{\beta}) = \sum_{k=0}^{\infty} n_k \left( \frac{k-1}{m-0} \frac{1}{r + m} \right). \quad (14.6)$

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

Binomial: $\hat{q} = \frac{1}{m} \sum_{k=0}^{\infty} \frac{kn_k}{n k}$, \quad (14.8)

The (a,b,1) class: $\pi(1 - e^{-\lambda}) = \frac{n-n_0}{n} \lambda$. \quad (14.9) \quad $\pi = \frac{1-P_0}{1-P_0} \lambda$. \quad (14.10)

Zero-modified Binomial: $\pi = \frac{1-P_0}{1-P_0} m q$, \quad (14.11) \quad $l_1 = \sum_{k=1}^{\infty} n_k \ln p_k - (n-n_0) \ln(1-P_0)$, \quad (14.12)

Hence, $l_1 = \sum_{k=1}^{\infty} n_k \ln \left[ \left( \frac{k+r-1}{k} \right) \left( \frac{1+\beta}{1+\beta} \right)^r \left( \frac{1}{\beta} \right)^k \right] - (n-n_0) \ln \left[ 1 - \left( \frac{1}{1+\beta} \right)^r \right]. \quad (14.13)$

$$g_k = \frac{\lambda^k}{k!} \sum_{j=1}^{k} f_j g_{k-j}, \quad k = 1, 2, 3, \ldots, \quad (14.14) \quad \text{where } f_j = \beta^{j-1}/(1+\beta)^j, \quad j = 1, 2, 3, \ldots.$$ 

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** $l_j(\theta_j, \hat{\theta}_j)$ describes the penalty paid by the investigator when $\hat{\theta}_j = \text{estimator}$ while $\theta_j = \text{true value of the } j^{th} \text{ parameter}$. 

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.

<table>
<thead>
<tr>
<th>loss function $l_j(\theta_j, \hat{\theta}_j)$</th>
<th>Square error</th>
<th>absolute error</th>
<th>zero-one</th>
</tr>
</thead>
<tbody>
<tr>
<td>$</td>
<td>\hat{\theta}_j - \theta_j</td>
<td>$</td>
<td>mean</td>
</tr>
</tbody>
</table>

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

$$\int_a^b \pi_{\theta_j|x}(\theta_j|x) \, d\theta_j = 1 - \alpha \implies \pi_{\theta_j|x}(a|x) = \pi_{\theta_j|x}(b|x).$$

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

$$\Pr(\theta_j \in C) \geq 1 - \alpha \text{ and } C = \{\theta_j : \pi_{\theta_j|x}(\theta_j|x) \geq c\} \text{ for some } c$$

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).

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 $\Theta = \theta$, the random variables $X_1, X_2, \ldots, X_n$ are i.i.d. with pf $f_{X_j|\Theta}(x_j|\theta) = \frac{p(x_j)e^{\Gamma(x_j)\theta}}{q(\theta)}$ where $\Theta$ has pdf $\pi(\theta) = \frac{[q(\theta)]^{-k}e^{\mu k r(\theta)}r(\theta)}{c(\mu,k)}$.

$k$ and $\mu$ are parameters of the distribution and $c(\mu, k)$ is the normalizing constants. Then the posterior pf $\pi_{\Theta|x}(\Theta|x)$ is of the same form as $\pi(\Theta)$.

**KPW16. Model Selection**

Models: $F^*(x) = \begin{cases} 0 & x < t, \\ F(x) - F(t) & x \geq t. \end{cases}$  

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

Models to data Graphical comparison: Check discrepancies (1) Empirical & model plot ($F_n(x) & F^*(x)$ vs $x$ plot) (2) Deviation plot ($D(x) = F_n(x) - F^*(x)$ vs $x$ plot) (3) Probability $p - p$ plot: check for straight $45^\circ$ line

**Hypothesis tests**

$H_0$: Data came from population with stated model vs $H_a$: Data did not come from such population \;

$\rightarrow$ (1) **KS** (2) **AD** (3) **Chi-Square GoF** test

(1) **Kolmogorov-Smirnov (KS) Test**: Statistic $D = \max_{t \leq x \leq u} |F_n(x) - F^*(x)|$ where


\( t \) =left truncation point (\( t = 0 \) if no truncation) \quad u =right censoring point (\( u = \infty \) if no censoring).

If \( D \leq CV \) don’t reject \( H_0 \)

\[
\begin{array}{cccc}
\text{where } \alpha & 0.10 & 0.05 & 0.01 \\
\text{critical value} & 1.22/\sqrt{n} & 1.36/\sqrt{n} & 1.63/\sqrt{n} \\
\end{array}
\]

If \( D > CV \) reject \( H_0 \)

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

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

\[
A^2 > CV \Rightarrow \text{reject } H_0,
\]

\[
\text{critical value } 1.933, 2.492, 3.857
\]

(2) **Anderson-Darling (AD) Test**: Statistic \( A^2 = n \int_{t=0}^{u} \left[ F_n(x) - F^*(x) \right]^2 \frac{F^*(x)}{1 - F^*(x)} f^*(x) \, dx \)

B) **Score-based approach**: Some scores worth considering:

\[
\begin{align*}
\text{Selection of Models:} \\
H_0 & \iff \text{don’t reject } H_0 \quad \text{where } CV = \chi^2_{df,1}\alpha \quad \text{is from a } \chi^2 \text{ table and } df = \# \text{parameter}_{H_0} - \# \text{parameter}_{H_A}. \\
\text{LR} & \iff \text{Test: } \text{Statistic } LR = \sum_{g=1}^{k} \left( \hat{p}_g - p_{ng} \right)^2 = \sum_{g=1}^{k} \frac{(E_g - O_g)^2}{E_g} \quad \text{where } \hat{p}_g = F^*(c_g) - F^*(c_{g-1}) \quad p_{ng} = F_n(c_g) - F_n(c_{g-1}), \\
& \iff \text{don’t reject } H_0 \\
\chi^2_{df} & \iff \text{reject } H_0, \quad \text{where } CV = \chi^2_{df,1}\alpha \\
\end{align*}
\]

\( B ) \text{ H}_0 \): Data came from population with distribution model A vs \( H_A \): Data came from population with distribution model B (where A is special case of B).

**Likelihood ratio (LR)** Test: Statistic \( T = 2 \ln \left( L_A / L_0 \right) = 2 \ln L_A - \ln L_0 \) (c.f. LR tests in Cox Models)

If \( T \leq CV \) don’t reject \( H_0 \) \quad \text{where } \ln L_0 = \text{Likelihood function maximized under } H_0.

If \( T > CV \) reject \( H_0 \), \quad \text{where } L_A = \text{Likelihood function maximized under } H_A.

\( CV = \chi^2_{df,1}\alpha \) is from a \( \chi^2 \) table and \( df = \# \text{parameter}_{H_0} - \# \text{parameter}_{H_A} \).

**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:

- Lowest value of \( \text{a) Kolmogorov-Smirnov} \quad \text{b) Anderson-Darling} \quad \text{c) Chi-square goodness of fit statistic} \)
- Highest value of \( \text{d) the likelihood function at its maximum} \quad \text{e) } p\text{-value for the Chi-square GoF statistic} \)

**KK7. Parametric Survival Models**

<table>
<thead>
<tr>
<th>h(t)</th>
<th>( pt^{p-1} \exp(\beta_0) )</th>
<th>( \exp(\beta_0) )</th>
<th>( \lambda t^{p-1} \exp(-\lambda t^p) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( h(t, X) )</td>
<td>( \lambda t^{p-1} )</td>
<td>( \lambda )</td>
<td>( 1 + \lambda t^p )</td>
</tr>
<tr>
<td>( p \leq 1 ) decreasing</td>
<td></td>
<td></td>
<td>( p \leq 1 ) decreasing</td>
</tr>
<tr>
<td>( p = 1 ) constant</td>
<td></td>
<td></td>
<td>( p &gt; 1 ) increase</td>
</tr>
<tr>
<td>( p &gt; 1 ) increasing</td>
<td>Weibull(( p = 1 ))</td>
<td></td>
<td>then decrease</td>
</tr>
</tbody>
</table>

\( \text{PH form} \quad \lambda = \exp(\beta_0 + \sum \beta_j X_j) \)

\( \text{PO form} \quad \lambda = \exp(\beta_0 + \sum \beta_j X_j) \)

<table>
<thead>
<tr>
<th>S(t)</th>
<th>( \exp(-\lambda t^p) )</th>
<th>( \exp(-\lambda) )</th>
<th>( \lambda \exp(-\lambda) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{HR (TRT = 1 vs 0)} )</td>
<td>( \exp(\beta_1) )</td>
<td>( \exp(\beta_1) )</td>
<td>( 1 + \lambda t^p )</td>
</tr>
<tr>
<td>( \ln [-\ln S(t)] )</td>
<td>( \ln(\lambda) + p \ln(t) )</td>
<td>( \exp(\beta_1) )</td>
<td></td>
</tr>
</tbody>
</table>

\( \text{Failure odds} \)

\( \frac{1 - S(t)}{S(t)} \)

| \( \ln(\text{failure odds}) \) | \( \lambda t^p \) | \( \ln(\lambda) + p \ln(t) \) |

| \( f(t) = h(t)S(t) \) | \( \lambda t^{p-1} \exp(-\lambda t^p) \) | \( \lambda \exp(-\lambda) \) | \( \lambda t^{p-1} \exp(-\lambda t^p) \) |

| \( \text{AFT} \) | \( t = [-\ln S(t)]^{1/p} \times \frac{1}{\lambda t^p} \) | \( t = [-\ln S(t)] \times \frac{1}{\lambda t^p} \) | \( t = \left[ \frac{1}{S(t)} - 1 \right]^{1/p} \times \frac{1}{\lambda t^p} \)|

\( \frac{1}{\lambda t^p} = \exp(\alpha_0 + \sum \alpha_i X_i) \quad \frac{1}{\lambda} = \exp(\alpha_0 + \sum \alpha_i X_i) \quad \frac{1}{\lambda t^p} = \exp(\alpha_0 + \sum \alpha_i X_i) \)

\( \alpha_i \) vs \( \beta_i \)

\( \beta_i = -\alpha_i p \quad \beta_i = -\alpha_i \quad \beta_i = -\alpha_i p \)

\( \gamma \) = \( \exp(\alpha_0) \)  

\( \gamma = \exp(\alpha_0) \)

AFT \( \iff \) PH  

AFT \( \iff \) 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 ) exponentially decreasing</td>
<td>( \gamma = 0 ) constant</td>
</tr>
<tr>
<td></td>
<td>( \gamma &gt; 0 ) exponentially increasing</td>
<td></td>
</tr>
</tbody>
</table>

**PH form**

\[
t = \exp(\alpha_0 + \sum \alpha_i X_i + \sigma) \quad t = \exp(\alpha_0 + \sum \alpha_i X_i + \sigma) \quad t = \exp(\alpha_0 + \sum \alpha_i X_i + \epsilon)
\]

**AFT**

\[
\epsilon \sim N(0, 1)
\]

**Frailty Models:**

\[
h_j(t, X|\alpha_j) = \alpha_j h(t, X) \quad j = 1, 2, \cdots, n \text{ with } \mu_\alpha = 1 \text{ and variance } \alpha = \theta
\]

- model with Gamma frailty: \( \alpha \sim \text{gamma} (\mu_\alpha = 1, \text{variance } \alpha = \theta) \)
- Weibull with gamma frailty \( \text{HR(TRT=2 vs 1)} = \begin{cases} \exp(\beta_1) & \alpha_1 = \alpha_2 \\ \frac{\alpha_1}{\alpha_2} \exp(\beta_1) & \alpha_1 \neq \alpha_2 \end{cases} \)

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
2. **Stratified Cox PH**
3. **Parametric with frailty model**

**1. Counting Process (CP) with Cox PH**

- standard cox: \( h(t, X) = h_0(t) \exp(\sum \beta_i X_i) \)
- likelihood function is different than nonrecurrent event (subjects remain in risk set until last follow-up interval )
- Robust estimation for variance estimators: \( \hat{R}(\beta) = \hat{\text{Var}}(\beta) | \hat{R}_S \hat{\text{Var}}(\beta) \) where \( \hat{\text{Var}}(\beta) \) = information matrix and \( \hat{R}_S = \text{matrix of score residuals.} \)

**2. Stratified Cox PH models for recurrent times:**

- **no interaction** stratified cox: \( h_g(t, X) = h_0g(t) \exp(\sum \beta_i X_i) \)
- **interaction** stratified cox: \( h_g(t, X) = h_0g(t) \exp(\sum \beta_i g X_i) \)
- 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 \( T_k = \text{survival time up to occurrence of } k^{th} \text{ event} \)

- Stratum \( S_{kc}(t) = Pr(T_{kc} > t) \quad T_{kc} = \text{time from } (k-1)^{st} \text{ to } k^{th} \text{ event} \) restricts data to subjects with \((k-1) \text{ events} \)
- Marginal \( S_{km}(t) = Pr(T_{km} > t) \quad T_{km} = \text{time from study entry to } k^{th} \text{ event} \) 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 \( T_c = \text{time to failure from event } c \)
- cause-specific model: \( h_c(t, X) = h_{0c}(t) \exp(\sum \beta_i c X_i) \quad c = 1, 2, \cdots, C. \)
- **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. \( \text{CIC}(t_f) = \sum_{j'=1}^f \hat{I}_c(t_{f'}) = \sum_{j'=1}^f \hat{S}(t_{f'-1}) \hat{h}_c(t_{f'}) \)
- Conditional Probability Curves (CPC): \( \text{CPC}_c = P(T_c > t) \) where \( T_c = \text{time until event } c \) occurs while \( T = \text{time any competing risk event occurs} \)

\( \text{CPC}_c = \text{CIC}_c/(1 - \text{CIC}_c) \)

- (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