AS475 Survival Models for Actuaries Formula
O. Peliminary SOA Exam P Formula

$\Gamma(\alpha; x) = \frac{1}{\Gamma(\alpha)} \int_0^x t^{\alpha-1} e^{-t} dt \qquad \Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt \qquad \Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1)$

Continuous	**pdf** $f(x)$	**mgf** $M_X(t)$	**Mean** $E[X]$	$Var(X)$
Uniform(a,b)	$f(x) = \begin{cases} (b-a)^{-1} & a < x < b \\ 0 & \text{otherwise} \end{cases}$	$\frac{e^{tb} - e^{ta}}{t(b-a)}$	$\frac{b+a}{2}$	$\frac{(b-a)^2}{12}$
Exponential(β)	$f(x) = \begin{cases} \frac{1}{\beta} e^{-\lambda x/\beta} & x \geq 0 \\ 0 & x < 0 \end{cases}$	$\frac{\lambda}{\lambda - t} \quad \beta = \frac{1}{\lambda} > 0$	β	β^2
Gamma(α,β)	$f(x) = \begin{cases} \frac{e^{-x/\beta} x^{\alpha-1}}{\beta^\alpha \Gamma(\alpha)} & x \geq 0 \\ 0 & x < 0 \end{cases}$	$\left(\frac{\lambda}{\lambda - t}\right)^\alpha$	$\alpha\beta$	$\alpha\beta^2$
Normal(μ,σ^2) $-\infty < x < \infty$	$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)}$	$exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)$	μ	σ^2
Pareto(α,θ)	$f(x) = \frac{\alpha\theta^\alpha}{(x+\theta)^{\alpha+1}}$ $F(x) = 1 - \left(\frac{\theta}{x+\theta}\right)^\alpha$	$M_X(t)$ not given $E[X^k] = \frac{\theta^k k! \Gamma(\alpha-k)}{\Gamma(\alpha)}$	$\frac{\theta}{\alpha-1}$	$\frac{\alpha\theta^2}{(\alpha-1)^2(\alpha-2)}$
Single Pareto(α,θ)	$f(x) = \frac{\alpha\theta^\alpha}{x^{\alpha+1}}$ $F(x) = 1 - \left(\frac{\theta}{x}\right)^\alpha$	$M_X(t)$ not given $E[X^k] = \frac{\alpha\theta^k}{\alpha-k} \quad k < \alpha$	$\frac{\alpha\theta}{\alpha-1}$	$\frac{\alpha\theta^2\left[(\alpha-1)^2-\alpha\right]}{(\alpha-1)^2(\alpha-2)}$

Discrete	**pmf** $p(x)$	**mgf** $M(t)$	**Mean** $E[X]$	$Var(X)$
Binomial(n,p) $0 \leq p \leq 1$	$\binom{n}{x} p^x (1-p)^{n-x}$ $x = 0, 1, \ldots, n$	$(pe^t + 1 - p)^n$	np	$np(1-p)$
Poisson(λ)	$e^{-\lambda}\lambda^x/x!$ $x = 0, 1, \ldots$	$exp[\lambda(e^{-t} - 1)]$	λ	$\lambda, \quad \lambda > 0$
Geometric(p)	$p^x(1-p)^{x-1}$ $x = 0, 1, \ldots$	$\frac{pe^t}{1-(1-p)e^t}$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Negative Binomial(r,p)	$\binom{n-1}{r-1} p^r (1-p)^{n-r}$ $n = r, r+1, \ldots$	$\left[\frac{pe^t}{1-(1-p)e^t}\right]^r$	$\frac{r}{p}$	$r\frac{1-p}{p^2}$
Hypergeometric(n,K,N)	$\frac{1}{\binom{N}{n}} \binom{K}{x}\binom{N-K}{n-x}$ $x = 0, 1, \ldots, \min(n,K)$	special function	$np^* = n\frac{K}{N}$	$np^*(1-p^*)\frac{N-n}{N-1}$

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; $t_2 = \infty$

$d = \begin{cases} 1 & \text{if failure} \\ 0 & \text{censored} \end{cases}$ $S(t)$ = survivor function, $S(t) = P(T > t)$ hazard function $h(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t}$

Hazard function =conditional failure rate $h(t)$ = instantaneous potential

Relationship of $S(t)$ and $h(t)$: If you know one, you can determine the other. $h(t) = \lambda$ iff $S(t) = e^{-\lambda t}$

$h(t) = -\left[\frac{dS(t)/dt}{S(t)}\right] \qquad S(t) = \exp\left[-\int_0^t h(u)du\right] \qquad \hat{S}(t)$ = 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 : $\overline{T} = \frac{1}{n}\sum_{i=1}^{n} t_i$

	Linear regression	Logistic regression	Survival analysis
Measure of effect:	regression coefficient β	odds ratio e^{β}	hazard ratio e^{β}

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 probablity rule $P(A \cap B) = P(A) \times P(B|A)$

Log-Rank Test for no difference in survival curves of Several Groups: $\mathbf{d}'\mathbf{V}^{-1}\mathbf{d} \backsim \chi^2_{G-1}$, $i = 1, 2, \cdots, G$

$$\begin{aligned}
\mathbf{d} &= (O_1 - E_1, O_2 - E_2, \cdots, O_{G-1} - E_{G-1})' \quad f = 1, 2, \cdots, k \text{ time intervals for the } G \text{ groups} \\
\mathbf{V} &= ((v_{ij})) \quad O_i - E_i = \sum_{f=1}^{k} (m_{if} - e_{if}) \qquad v_{ii} = Var(O_i - E_i) = \sum_{f=1}^{k} \frac{n_{if}(n_f - n_{if})m_f(n_f - m_f)}{n_f^2(n_f - 1)} \\
v_{ij} &= Cov(O_i - E_i, O_j - E_j) = \sum_{f=1}^{k} \frac{-n_{if}n_{jf}m_f(n_f - m_f)}{n_f^2(n_f - 1)} \qquad m_f = \sum_{i=1}^{G} m_{if} \quad e_{if} = \frac{n_{if}}{n_f} m_f \qquad n_f = \sum_{i=1}^{G} n_{if}
\end{aligned}$$

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

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

Approximate formula: $\sum_{i-1}^{G} \dfrac{(O_i - E_i)^2}{E_i} \backsim \chi^2_1$, $i = 1, 2.$

Alternative tests for 2 groups: Test statistic: $\dfrac{\left(\sum_f w(t_{(f)})(m_{if} - e_{if})\right)^2}{Var\left(\sum_f w(t_{(f)})(m_{if} - e_{if})\right)}$

where	LogRank	Wilcoxon	Tarone-Ware	Peto	Flamington-Harrington
$w(t_{(f)})$ =weights at the f^{th} failure time.	1	n_f	$\sqrt{n_f}$	$\tilde{s}(t_{(f)})$	$\hat{S}(t_{(f-1)})^p[1 - \hat{S}(t_{(f-1)})]^q$ $p = 0 \rightarrow$LogRank

KK3-KK6: Cox Models

	KK3. Cox PH	**KK5. Stratified** Cox PH	**KK6. Cox PH for Time-dependent Variables**
Model	$h_0(t)\exp(\sum_{i=1}^{p} \beta_i X_i)$	$h_{0g}(t)\exp(\sum_{i=1}^{p} \beta_i X_i)$ $g = 1, 2, \cdots, k$	$h_0(t)\exp(\sum_{i=1}^{p_1} \beta_i X_i + \sum_{j=1}^{p_2} \delta_j X_j)$
HR: $\dfrac{h(t, \mathbf{X}^*)}{h(t, \mathbf{X})}$	$\exp\left[\sum_{i=1}^{p} \beta_i (X_i^* - X_i)\right]$		$\exp[\sum_{i=1}^{p_1} \beta_i (X_i^* - X_i))$ $+ \sum_{j=1}^{p_2} \delta_j \left(X_j^* - X_j\right)]$
Meaning PH	$\dfrac{h(t, \mathbf{X}^*)}{h(t, \mathbf{X})} = \theta$		PH not satisfied
General model to assess		Interaction: $h_{0g}(t)\exp(\sum_{i=1}^{p} \beta_{ig} X_i)$ $g = 1, 2, \cdots, k$ strata defined from Z^* or $h_{0g}(t)\exp[\sum_{i=1}^{p} \beta_i X_i$ $+ \sum_{g=1}^{k-1} \sum_{i=1}^{p} \beta_{ig} X_i Z_g]$	PH assumption of Cox PH: $h_0(t)\exp(\sum_{i=1}^{p} \beta_i X_i + \sum_{i=1}^{p} \delta_i X_i g_i(t))$ where $g_i(t)$ is time-dependent fn heaviside $g_i(t) = \begin{cases} 1 & \text{if } t \text{ in interval } i \\ 0 & \text{otherwise} \end{cases}$
Likelihood ratio (LR) test	$-2\ln L_R - (-2\ln L_F)$ $LR \dot\sim \chi^2_{\#parameters \text{ in } F-R}$	$-2\ln L_R - (-2\ln L_F)$ $LR \dot\sim \chi^2_{p(k-1)}$	$-2\ln L_R - (-2\ln L_F)$ $LR \dot\sim \chi^2_{\#parameters \text{ in } F-R}$

95% Confidence Interval for Hazard Ratio, $HR = \exp(\ell)$ where $\ell = \beta_1 + \sum_{i=1}^{k} \delta_i W_i$:

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

Adjusted survival curve.

$$\begin{aligned} S(t, \mathbf{X}) &= \exp\left[-\int_0^t h(u)du\right] = \exp\left[-\int_0^t h_0(u)\exp(\sum_{i=1}^{p}\beta_i X_i)du\right] = \exp\left[-\exp(\sum_{i=1}^{p}\beta_i X_i)\int_0^t h_0(u)du\right] \\ &= \left[\exp\left(-\int_0^t h_0(u)du\right)\right]^{\exp(\sum_{i=1}^{p}\beta_i X_i)} = [S_0(t)]^{\exp(\sum_{i=1}^{p}\beta_i X_i)} \end{aligned}$$

KK4. Methods for checking PH assumptions

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

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

KPW11. 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 $F_n(x) = \frac{\text{number of observations} \le x}{n}$, 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 $\ge y_i$, the*

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

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*

$$F_n(x) = \frac{c_j - x}{c_j - c_{j-1}}F_n(c_{j-1}) + \frac{x - c_{j-1}}{c_j - c_{j-1}}F_n(c_j), \qquad c_{j-1} \le x \le c_j.$$

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*

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

KPW12. Estimation of Modified Data (See also KK2)

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

	from below *at d*	*(or* ***left*** *)*		***from above*** *at u*	*(or* ***right*** *)*
observation	***truncated***	***censored***	***observation***	***truncated***	***censored***
$x \le d$	*not recorded or missing*	d	$x < u$	x	x
$x > d$	x	x	$x \ge u$	*not recorded or missing*	u
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)

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

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

Kaplan-Meier estimate $S_n(x) = \begin{cases} 1, & 0 \le t < y_1 \\ \prod_{i=1}^{j-1} \left(\frac{r_i - s_i}{r_i} \right), & y_{j-1} \le x < y_j,\ j = 2, ...k, \\ \prod_{i=1}^{k} \left(\frac{r_i - s_i}{r_i} \right) \text{ or } 0, & t \ge y_k \end{cases}$

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

Definition 11 *(12.2) A* ***kernel density estimator*** *of a distribution function is* $\widehat{F}(x) = \sum_{j=1}^{k} p(y_j) K_{y_j}(x)$ *and the estimator of the density function is* $\widehat{f}(x) = \sum_{j=1}^{k} p(y_j) k_{y_j}(x)$,

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

	Uniform kernel	***Triangular kernel***	***Gamma kernel***
$k_y(x)$	$\begin{cases} 0, & x < y - b, \\ \frac{1}{2b}, & y - b \le x \le y + b, \\ 0, & x > y + b, \end{cases}$	$\begin{cases} 0, & x < y - b, \\ \frac{x - y + b}{b^2}, & y - b \le x \le y, \\ \frac{y + b - x}{b^2}, & y \le x \le y + b, \\ 0, & x > y + b, \end{cases}$	$\frac{x^{\alpha-1} e^{-x\alpha/y}}{(y/\alpha)^\alpha \Gamma(\alpha)}$ *shape* α *and scale parameter* y/α
$K_y(x)$	$\begin{cases} 0, & x < y - b, \\ \frac{x - y + b}{2b}, & y - b \le x \le y + b, \\ 1, & x > y + b. \end{cases}$	$\begin{cases} 0, & x < y - b, \\ \frac{(x - y + b)^2}{2b^2}, & y - b \le x \le y, \\ 1 - \frac{(y + b - x)^2}{2b^2}, & y \le x \le y + b, \\ 1, & x > y + b. \end{cases}$	*Gamma kernel has mean* $\alpha(y/\alpha) = y$ *& variance* $\alpha(y/\alpha)^2 = y^2/\alpha$

Exposure method	Exposure **definition**	q_j
Exact	exposure = exact total time under observation	$q_j = 1 - \exp(-d_j/e_j)$
Actuarial	exposure period extend to end of age interval	$q_j = d_j/e_j$

Life insurance Exposure method	Exposure **definition**
Insuring Ages	based on policy holder's age **at entry**
Anniversary based	based on when the policy reach **its anniversary**

Interval-based Exposure method	**UDD exposure** (risk set)	**midyear exposure** (risk set)	P_j=number who start n_j=new entrants
Exact	$P_j + (n_j - d_j - w_j)/2$	$P_j + (n_j - w_j)/2$	d_j=number who die
Actuarial	$P_j + (n_j - w_j)/2$	$P_j + (n_j - w_j)/2$	w_j=number leave

KPW13. Frequentist Estimation

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

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

Definition 15 *(13.3) The* ***smoothed empirical estimate*** *of a percentile is calculated as* $\widehat{\pi}_g = (1-h)x_{(j)} + hx_{(j+1)}$, *where* $j = \lfloor (n+1)g \rfloor$ *and* $h = (n+1)g - j$. *Here* $\lfloor \cdot \rfloor$ *indicates the* ***greatest integer*** *function and* $x_{(1)} \leq x_{(2)} \leq ... \leq x_{(n)}$ *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* θ *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* θ *in an interval containing the true value (for discrete variables, replace integrals by sums):*

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

(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* $\widehat{\theta}_n$ *converges to a normal distribution with mean* θ *and variance such that* $I(\theta)Var\left(\widehat{\theta}_n\right) \to 1$, *where the Fisher's information*

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

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

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* θ *and covariance matrix* $\mathbf{\Sigma}/n$, *where neither* θ *nor* $\mathbf{\Sigma}$ *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* $(\partial\mathbf{g})^T \mathbf{\Sigma} (\partial\mathbf{g})/n$, *where* $\partial\mathbf{g}$ *is the vector of* ***first derivatives***, *that is,* $\partial\mathbf{g} = (\partial g/\partial\theta_1, \cdots, \partial g/\partial\theta_k)^T$ *and it is to be evaluated at* θ, *the true parameters of the original random variable.*

KPW14. Frequentist Estimation for Discrete Distributions

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

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

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

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

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

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

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

The (a,b,1) class: $\overline{x}\left(1-e^{-\lambda}\right) = \frac{n-n_0}{n}\lambda$. (14.9) $\overline{x} = \frac{1-\widehat{p}_0^M}{1-p_0}\lambda$. (14.10)

Zero-modified Binomial: $\overline{x} = \frac{1-\widehat{p}_0^M}{1-p_0} mq$, (14.11) $l_1 = \sum_{k=1}^{\infty} n_k \ln p_k - (n-n_0)\ln(1-p_0)$, (14.12)

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

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

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* θ *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*** $f_{\mathbf{X}|\Theta}(\mathbf{x}|\theta)$ *is the probability distribution for the data given a particular value of the parameter.*

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

Definition 23 *(D15.3) The* ***marginal distribution*** *of* **X** *has pdf* $f_{\mathbf{X}}(\mathbf{x}) = \int f_{\mathbf{X}|\Theta}(\mathbf{x}|\theta)\pi(\theta)d\theta$.

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

Definition 25 *(D15.7) The* ***Predictive distribution*** $f_{Y|\mathbf{X}}(y|\mathbf{x})$ *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* $\pi_{\Theta|\mathbf{X}}(\theta|\mathbf{x}) = \frac{f_{\mathbf{X}|\Theta}(\mathbf{x}|\theta)\pi(\theta)}{\int f_{\mathbf{X}|\Theta}(\mathbf{x}|\theta)\pi(\theta)d\theta}$ *while the* ***predictive distribution*** *can be computed as* $f_{Y|\mathbf{X}}(y|\mathbf{x}) = \int f_{Y|\boldsymbol{\Theta}}(y|\theta)\pi_{\Theta|\mathbf{X}}(\theta|\mathbf{x})d\theta$,
where $f_{Y|\boldsymbol{\Theta}}(y|\theta)$*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* θ_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.*

	Square error	*absolute error*	*zero-one*
loss function $l_j(\hat{\theta}_j, \theta_j)$	$(\hat{\theta}_j - \theta_j)^2$	$\left\lvert\hat{\theta}_j - \theta_j\right\rvert$	$0 \quad if\ \hat{\theta}_j = \theta_j$ $1 \quad if\ \hat{\theta}_j \neq \theta_j$
Bayes estimate	*mean*	*median*	*mode of* $\pi_{\Theta\lvert\mathbf{X}}(\theta\lvert\mathbf{x})$

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

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

$$\int_a^b \pi_{\Theta_j|\mathbf{X}}(\theta_j|\mathbf{x})d\theta_j = 1 - \alpha \Longrightarrow \pi_{\Theta|\mathbf{X}}(a|\mathbf{x}) = \pi_{\Theta|\mathbf{X}}(b|\mathbf{x}).$$

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*

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

where c is the largest value for which the probability inequality holds.

Theorem 32 *(T15.16:* ***Bayesian Central Limit Theorem****) If* $\pi(\theta)$ *and* $f_{\mathbf{X}|\Theta}(\mathbf{x}|\theta)$ *are both* ***twice differentiable*** *in the elements of* θ *and* ***other commonly satisfied assumptions*** *hold, then the posterior distribution of* Θ *given* $X = \mathbf{x}$ *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* $X_1, X_2, \cdots, X_n$ *are i.i.d. with pf*

$$f_{X_j|\Theta}(x_j|\theta) = \frac{p(x_j)e^{r(\theta)x_j}}{q(\theta)} \quad \text{where } \Theta \text{ has pdf } \pi(\theta) = \frac{[q(\theta)]^{-k}\, e^{\mu k r(\theta)} r'(\theta)}{c(\mu, k)},$$

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

KPW16. Model Selection

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

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

Hypothesis tests

A) H_o: Data came from population with stated model
vs H_a: Data **did not** come from such population } →(1) **KS** (2) **AD** (3) **Chi-Square GoF** test

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

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

If $D \le CV$ don't reject H_o
$D > CV$ reject H_o,

where α	0.10	0.05	0.01
critical value	$1.22/\sqrt{n}$	$1.36/\sqrt{n}$	$1.63/\sqrt{n}$

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

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

If $A^2 \le CV$ don't reject H_o
$A^2 > CV$ reject H_o,

where α	0.10	0.05	0.01
critical value	1.933	2.492	3.857

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

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

If $\chi^2_{df} \le CV$ don't reject H_o
$\chi^2_{df} > CV$ reject H_o,
where $CV = \chi^2_{df,1-\alpha}$ is from a χ^2 table.

B) H_o: 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 $T = 2\ln(L_a/L_0) = 2(\ln L_a - \ln L_0)$ (*c.f. LR tests in Cox Models*)

If $T \le CV$ don't reject H_o where L_0 =Likelihood function maximized under H_o
$T > CV$ reject H_o, L_a =Likelihood function maximized under H_a.

$CV = \chi^2_{df,1-\alpha}$ is from a χ^2 table and $df = \#parameter_{H_a} - \#parameter_{H_0}$.

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 (i) Kolmogorov-Smirnov (ii) Anderson-Darling (iii) Chi-square goodness of fit statistic
- ■ Highest (iv) value of the likelihood function at its maximum (v) p-value for the Chi-square GoF statistic

KK7. Parametric Survival Models

	Weibull	**Exponential**	**Log-logistic**
$h_0(t)$	$pt^{p-1}\exp(\beta_0)$	$\exp(\beta_0)$	complicated form
$h(t,X)$	$\lambda p t^{p-1}$	λ	$\frac{\lambda p t^{p-1}}{1+\lambda t^p}$
	$p<1$ decreasing $p=1$ constant $p>1$ increasing	Weibull($p=1$)	$p\le 1$ decreasing $p>1$ increase then decrease
PH form	$\lambda = \exp(\beta_0 + \sum \beta_i X_i)$	$\lambda = \exp(\beta_0 + \sum \beta_i X_i)$	
PO form			$\lambda = \exp(\beta_0 + \sum \beta_i X_i)$
$S(t)$	$\exp(-\lambda t^p)$	$\exp(-\lambda)$	$\frac{1}{1+\lambda t^p}$
HR ($TRT = 1$ *vs* 0)	$\exp(\beta_1)$	$\exp(\beta_1)$	
$\ln[-\ln S(t)]$	$\ln(\lambda) + p\ln(t)$		
Failure odds $\frac{1-S(t)}{S(t)}$			λt^p
ln(failure odds)			$\ln(\lambda) + p\ln(t)$
$f(t) = h(t)S(t)$	$\lambda p t^{p-1}\exp(-\lambda t^p)$	$\lambda\exp(-\lambda)$	$\frac{\lambda p t^{p-1}}{(1+\lambda t^p)^2}$
AFT t	$t = [-\ln S(t)]^{1/p} \times \frac{1}{\lambda^{1/p}}$ $\frac{1}{\lambda^{1/p}} = \exp(\alpha_0 + \sum \alpha_i X_i)$	$t = [-\ln S(t)] \times \frac{1}{\lambda}$ $\frac{1}{\lambda} = \exp(\alpha_0 + \sum \alpha_i X_i)$	$t = \left[\frac{1}{S(t)} - 1\right]^{1/p} \times \frac{1}{\lambda^{1/p}}$ $\frac{1}{\lambda^{1/p}} = \exp(\alpha_0 + \sum \alpha_i X_i)$
α_i vs β_i	$\beta_i = -\alpha_i p$	$\beta_i = -\alpha_i$	$\beta_i = -\alpha_i p$
Accelaration γ	$\gamma = \exp(\alpha_0)$ AFT⇒PH then PH⇒AFT	$\gamma = \exp(\alpha_0)$	$\gamma = \exp(\alpha_0)$ $AFT \nLeftrightarrow PH$ $AFT \Leftrightarrow PO$

	General form	LogNormal	Gompertz
$h_0(t)$			$\exp(\gamma t)$
$h(t,X)$			$\lambda\exp(\gamma t)$ $\gamma<0$ exponentially decreasing $\gamma=0$ constant $\gamma>0$ exponentially increasing
PH form			$\lambda=\exp(\beta_0+\sum\beta_iX_i)$
t	$t=\exp(\alpha_0+\sum\alpha_iX_i+\sigma\epsilon)$	$t=\exp(\alpha_0+\sum\alpha_iX_i+\sigma\epsilon)$	$t=\exp(\alpha_0+\sum\alpha_iX_i+\epsilon)$
AFT		$\epsilon\backsim N(0,1)$	

Frailty Models: $h_j(t,X|\alpha_j)=\alpha_jh(t,X)\quad j=1,2,\cdots,n$ with $\mu_\alpha=1$ and variance$_\alpha=\theta$

model with Gamma frailty: α ˜ gamma ($\mu_\alpha=1$,variance$_\alpha=\theta$)

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

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

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** model

standard cox: $h(t,X)=h_0(t)\exp(\sum\beta_iX_i)$

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

Robust estimation for variance estimators: $\hat{\mathbf{R}}(\hat\beta)=\widehat{\mathbf{Var}}(\hat\beta)[\hat{\mathbf{R}}_{\mathbf{S}}'\hat{\mathbf{R}}_{\mathbf{S}}]\widehat{\mathbf{Var}}(\hat\beta)$ where $\widehat{\mathbf{Var}}(\hat\beta)$=information matrix and $\hat{\mathbf{R}}_{\mathbf{S}}$=matrix of score residuals.

(2) **Stratified Cox PH models for recurrent times**: time interval =strata

no interaction stratified cox: $h_g(t,X)=h_{0g}(t)\exp(\sum\beta_iX_i)$ 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 kth 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 occurence of k^{th} event.

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

b) Marginal $S_{km}(t)=Pr(T_{km}>t)$ T_k=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: $h_c(t,X)=h_{0c}(t)\exp(\sum_{i=1}^p\beta_{ic}X_i)\quad c=1,2,\cdots,C$.

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. $CIC(t_f)=\sum_{f'=1}^f\hat I_c(t_{f'})=\sum_{f'=1}^f\hat S(t_{f'-1})\hat h_c(t_{f'})$

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