Digital Actuarial Resources

Equation Study List

Sample Only – Purchase the Full Version

Society of Actuaries
Exam C: Construction and Evaluation of Actuarial Models

Introduction:

This compilation of equations is designed to aid the student in preparing for Exam C (Construction of Actuarial Models) offered through the Society of Actuaries starting in the spring of 2005. The comprehensive list contains equations for building and validating survival models used in the actuarial practice. This formula sheet includes topics from maximum likelihood estimation to credibility to simulation. Equations for creating cubic splines, a new topic on the exams, are also featured. The list is not guaranteed to cover all the material on the exam. Many nominal equations that are unlikely to show up on an exam are not included. The 350+ formulas are neatly organized into several categories. Refer to the table of contents for guidance.

www.digitalactuarialresources.com

Table of Contents

Section 3: Estimation of the Survival Function

Hazard Rate:

Hazard rate function = force of mortality = $h(x) = H'(x) = \dfrac{f(x)}{s(x)} = \dfrac{-s'(x)}{s(x)} = \mu_x$

Cumulative hazard rate function = $H(x) = \int_{-\infty}^{x} h(y)\ dy$

$H(x) = -\ln(s(x))$ AND $s(x) = e^{-H(x)}$

$Nelson - \dot{A}alen$ estimate:

$$\hat{H}(x) = \sum_{i=1}^{j-1} \frac{s_i}{r_i} \text{ for } t_{j-1} \le x < t_j$$

$\hat{H}(x)$ can be used to estimate $\hat{s}(x)$: $\hat{s}(x) = e^{-\hat{H}(x)}$

k = quantity of unique observations in the sample

x_i = the original ith observation in the data set

t_j = time of the jth unique event in the sample

s_j = total number of events at time t_j

d_i = left truncation value for observation i = number of people entering the study at time i

u_i = right-censored value for observation i

= number of people leaving the study for causes other than death at time i

P_j = population during time t_j

= $Count(x_i \mid x_i \ge t_j)$ + $Count(u_i \mid u_i \ge t_j)$ - $Count(d_i \mid d_i \ge t_j)$

= (all deaths occurring now or in the future) + (all members withdrawing now or in the future) – (all new entrants now or in the future)

r_j = risk set = $\sum_{i=j}^{k} s_i$

Distribution of entrants and withdrawals for large data sets:

α = proportion of additional insureds entering prior to t_j

β = proportion of surrenders occurring before t_j

$r_j = P_j + \alpha d_j - \beta u_j$

Common options for α and β:

(1.) $\alpha = 0.5$ and $\beta = 0.5$: Entrants and withdrawals occur uniformly over each time period. Half the withdrawals and half the new entrants occur before t_j, and the remainder of withdrawals and entrants for the time period occur after t_j.

(2.) $\alpha = 0.5$ and $\beta = 1$: Entrants occur uniformly during each time period, and all decrements occur at the start of each time period.

(3.) $\alpha = 1$ and $\beta = 0$: All entrants occur at the start of the time period, and all the surrenders occur at the end of the period following all the deaths.

Options for surrenders during a time period:

(1.) Surrenders are uniformly distributed during each period. Half of the surrenders occur before the midpoint of the period, followed by all the deaths at the midpoint, followed by the remaining half of the surrenders.

(2.) Surrenders happen only at the conclusion of a time period, along with all the deaths.

$$F_n(x) = \begin{cases} 0, x \le y_0 \\ \dfrac{n - r_j}{n}, t_{j-1} \le x < t_j \\ 1, x \ge y_k \end{cases}$$

Kaplan-Meier Product-Limit Estimator:

$$S_n(t) = \prod_{i=1}^{j-1} \frac{r_i - s_i}{r_i} \quad \text{for } t_{j-1} \le t < t_j$$

*Suppose x_k is right-censored and $k \ge i$. x_k is included in the calculation of r_i, but x_k is not included in the calculation of s_k.

Options for $S_n(t)$ estimates beyond the largest age of death recorded (x_n):

(1.) Let $S_n(t) = 0 \ \forall \ t > x_n$

(2.) Let $S_n(t) = S_n(x_n) \ \forall \ t > x_n$

(2.) Let $S_n(t) = S_n(x_n)^{(t/x_n)} = (\text{last survival rate})^{t/x_n}$

E(quantity of death events between times 0 and t) = $\int_0^t h(u) * r(u) \ du$

= integral over all forces of mortality and risk set sizes as the time goes from 0 to t

Table Format for Survival Estimation Problems:

j	t_j	s_j	r_j	$\hat{H}(t_j)$ OR $\hat{s}(t_j)$
↓	↓	↓	↓	Estimate from Nelson-Aalen or Kaplan-Meier

<u>Binomial Distribution:</u>

X = number of successes

X ~ Binomial(n, p)

n = number of trials

m = number of classes for each trial/observation = 2

p = probability of success, q = probability of failure = 1 – p

r = number of successes
n – r = number of failures
E(X) = np
Var(X) = Var(number of successes) = npq
Var(proportion of successes) = $\frac{npq}{n^2} = \frac{pq}{n}$

<u>Binomial Distribution for Survival:</u>
$S_n(x)$ = proportion of trials greater than x
$S_n(x)$ ~ Binomial Proportion
n = number of observations
p = S(x)
q = F(X)

Individual Data:
Suppose n_x = amount of insureds alive at time x

$$S_n(x) = \frac{n_x}{n}$$

$$Var(S_n(x)) = Var(F_n(x)) = \frac{S(x)(1-S(x))}{n} \approx V\hat{a}r(S_n(x)) = \frac{S_n(x)(1-S_n(x))}{n}$$

$$V\hat{a}r({}_t\hat{p}_x \mid S_n(x)) = V\hat{a}r({}_t\hat{q}_x \mid S_n(x)) = \frac{{}_t\hat{q}_x * {}_t\hat{p}_x}{n_{x+t}}$$

$$V\hat{a}r(S_n(x)) = \frac{n_x(n-n_x)}{n^3} = V\hat{a}r({}_x\hat{p}_0 \mid n_x)$$

$$V\hat{a}r({}_t\hat{q}_x \mid n_x) = V\hat{a}r({}_t\hat{p}_x \mid n_x) = \frac{n_{x+t} * (n_x - n_{x+t})}{n_x^{\,3}}$$

Grouped Data:
*Data exist in intervals (c_{j-1}, c_j)

$$p_j = \Pr(x \in (c_{j-1}, c_j))$$

$$n_j = Count(x_i \mid x_i \in (c_{j-1}, c_j)) \sim Binomial(n,\ p = n_j / n)$$

$$N_j = Count(x_i \mid x_i \le c_{j-1})) \sim Binomial(n,\ p = N_j / n)$$

$$V\hat{a}r(N_j) = \frac{N_j(n-N_j)}{n} = n * \frac{N_j}{n} * \frac{n-N_j}{n} = npq$$

$$V\hat{a}r(n_j) = \frac{n_j(n-n_j)}{n} = n * \frac{n_j}{n} * \frac{n-n_j}{n} = npq$$

$$C\hat{o}v(n_j, N_j) = -\frac{n_j * N_j}{n} = -n * \frac{n_j}{n} * \frac{N_j}{n}$$

Suppose $x \in (c_{j-1}, c_j)$, and p_j and $S_n(c_j)$ are binomial proportion variables. Then,

$$\hat{Var}(S_n(x)) = Var\left(\left(\frac{c_j - x}{c_j - c_{j-1}}\right) * p_j + S_n(c_j)\right)$$

$$= \left(\frac{c_j - x}{c_j - c_{j-1}}\right)^2 * \frac{p_j * q_j}{n} + \frac{S_n(c_j) * (1 - S_n(c_j))}{n} - 2 * \left(\frac{c_j - x}{c_j - c_{j-1}}\right) * \left(\frac{p_j * S_n(c_j)}{n}\right)$$

$$\hat{Var}(f_n(x)) = \hat{Var}\left(\frac{n_j}{n(c_j - c_{j-1})}\right) = \left(\frac{1}{n(c_j - c_{j-1})}\right)^2 * \hat{Var}(n_j) = \frac{\hat{Var}(n_j)}{n^2 * (c_j - c_{j-1})^2}$$

Suppose $x_1, x_2 \in (c_{j-1}, c_j)$ and p_j is a binomial proportion variable. Then,

$$\hat{Var}(\Pr(x_1 \le X \le x_2)) = \hat{Var}\left(\left(\frac{x_2 - x_1}{c_j - c_{j-1}}\right) * p_j\right) = \left(\frac{x_2 - x_1}{c_j - c_{j-1}}\right)^2 * \frac{p_j * q_j}{n}$$

Confidence Interval for $S_n(x_j)$:

$$\left(S_n(x_j) - z_{\alpha/2} * \sqrt{\frac{S_n(x_j) * (1 - S_n(x_j))}{n}},\ S_n(x_j) + z_{\alpha/2} * \sqrt{\frac{S_n(x_j) * (1 - S_n(x_j))}{n}}\right)$$

Greenwood's Approximation (for individual and grouped data):

*Applies to Kaplan-Meier estimate

$$\hat{Var}(S_n(y_j)) = S_n(y_j)^2 * \sum_{i=1}^{j} \frac{s_i}{r_i * (r_i - s_i)}$$

$$\hat{Var}({}_t\hat{p}_x) = \hat{Var}({}_t\hat{q}_x) = ({}_t\hat{p}_x)^2 * \sum_{i=1}^{j} \frac{s_i}{r_i * (r_i - s_i)}$$

*Start i at the next event after time x.
*End at time x + t, inclusive

Estimated variance for $\hat{H}(y_j)$:

*Applies to Nelson-Aalen estimate

$$\hat{Var}(\hat{H}(y_j)) = \sum_{i=1}^{j} \frac{s_i}{r_i^2}$$

<u>Multinomial Distribution:</u>

n = number of trials
m = number of classes for each trial/observation
p_i = probability of the next observation belonging to class i
q_i = probability of the next observation not belonging to class i = 1 – p_i

n_i = number of observations in class i
Var(number of observations in class i) = $Var(n_i) = n * p_i * q_i$

Var(proportion of observations in class i) = $Var(p_i) = \frac{n * p_i * q_i}{n^2} = \frac{p_i * q_i}{n}$

Cov(number in class i, number in class j) = $Cov(n_i, n_j) = -n * p_i * p_j$

Cov(proportion in class i, proportion in class j) = $Cov(p_i, p_j) = -\frac{p_i * p_j}{n}$

<u>Confidence Intervals:</u>
Linear Confidence Interval for $S_n(t)$:

Lower Limit: $S_n(t) - z_{\alpha/2} * \sqrt{V\hat{a}r(S_n(t))}$

Upper Limit: $S_n(t) + z_{\alpha/2} * \sqrt{V\hat{a}r(S_n(t))}$

Log-Transformed Confidence Interval for $S_n(t)$:

$$W = e^\wedge\left(\frac{z_{\alpha/2} * \sqrt{V\hat{a}r(S_n(t))}}{S_n(t) * \ln(S_n(t))}\right)$$

Lower Limit: $S_n(t)^{1/W}$

Upper Limit: $S_n(t)^W$

Linear Confidence Interval for $\hat{H}(t)$:

Lower Limit: $\hat{H}(t) - z_{\alpha/2} * \sqrt{V\hat{a}r(\hat{H}(t))}$

Upper Limit: $\hat{H}(t) + z_{\alpha/2} * \sqrt{V\hat{a}r(\hat{H}(t))}$

Log-Transformed Confidence Interval for $\hat{H}(t)$:

$$W = e^\wedge\left(\frac{z_{\alpha/2} * \sqrt{V\hat{a}r(\hat{H}(t))}}{\hat{H}(t)}\right)$$

Lower Limit: $\hat{H}(t) * W^{-1}$

Upper Limit: $\hat{H}(t) * W$

<u>Kaplan-Meier Approximations:</u>

$$\hat{q}_i = \frac{s_i}{r_i}$$