Exam C/Exam 4 Review

Johnew Zhang

September 14, 2012

Contents

1 Severity, Frequency, and Aggregate Loss

Let's define those terms in the section headers first.

Severity is the average size of a loss .

Frequency is the average number of claims per time period usually per year.

Aggregate loss is the total loss paid per time period, usually per year.

Pure Premium is the expected aggregate loss per policyholder per time period, usually per year.

1.1 Basic Probability

In the previous SOA EXAM P, it covers the basic knowledge of continuous, discrete and mixed random variable. Here I will provide some simple review and for those distributions you can find it through SOA website which I won't cover.

n^{th} **raw moment** $\mu'_n = E[X^n]$

n^{th} **central moment** $\mu_n = E[(X-\mu)^n]$

Skewness $\gamma_1 = \frac{\mu_3}{\sigma^3}$

Kurtoisis $\gamma_2 = \frac{\mu_4}{\sigma^4}$

Coefficient of variations $\frac{\sigma}{\mu}$

Variance $Var(X) = E[X^2] - E[X]^2$

Covariance $Cov(X,Y) = E[XY] - E[X]E[Y]$

Percentile $F(\pi_p) = p$. A very important criteria is $Pr(X \leq \pi_p) \geq p$ and $Pr(X < \pi_p) \leq p$

Conditional Mean Formula $E_X[X] = E_Y[E_X[X|Y]]$ or $E[_X[g(X)] = E_Y[E_X[g(X)|Y]]$

Moment generating function $M(t) = E[e^{tX}]$

Probability generating function $P(z) = E[z^X] = M(\ln z)$ where $p_n = \frac{p^{(n)}(0)}{n!}$

1.2 Parametric Distribution

A parametric distribution is one that is defined by a fixed number of parameters.

1.2.1 Scaling

If the distribution is parametrized in this fashion, so that the only parameter of cX having a different value from X is θ, and the value of θ for cX is c times the value of θ for X, then θ is called a scale parameter.

There are couple of rules below for parametric distribution scaling:

- Scaling for lognormal distribution: $X \sim LOGNOM(\mu, \sigma)$ will be transformed into $Y \sim LOGNOM(\mu + \ln c, \sigma)$
- One use of scaling is in handling inflation.

1.2.2 Transformations

$$f_Y(y) = f_X(g^{-1}(y))|\frac{dg^{-1}(y)}{dy}|$$

There are a few transformations that are used to create distributions:

1. $Y = X^{\tau}$ where $\tau > 0$ - transformed
2. $Y = X^{-1}$ - Inverse
3. $Y = X^{\tau}$ where $\tau < 0$ - inverse transformed
4. $Y = e^X$ - log

1.3 Mixture and Splices

1.3.1 Discrete Mixtures

A (finite) mixture distribution is a random variable X whose distribution function can be expressed as a weighted average of n distribution functions of random variable $X_i, i = 1, \cdots, n$. In other words,

$$F_X(x) = \sum_{i=1}^{n} w_i F_{X_i}(x)$$

$$f_X(x) = \sum_{i=1}^{n} w_i f_{X_i}(x)$$

1.3.2 Continuous Mixtures

Continuous mixtures means that the distribution function of the mixture is an integral of parametric distribution functions of random variables, and a parameter varies according to a distribution function.

1.3.3 Frailty models

A specific type of continuous mixtures is a frailty model. These models can be used to model loss sizes or survival times.

Suppose that hazard rate for each individual is $h(x|\Lambda) = \Lambda a(x)$, where $a(x)$ is some continuous function and the multiplier Λ varies by individual. thus the shape of the hazard rate function curve does not vary by individual. If you are given that A's hazard rate is twice B's at time 1, that implies Λ for A is twice Λ for B. That in turn implies that A's hazard rate is twice B's hazard rate at all times.

Let $H(x) = \int_0^x h(t)dt$ and $A(x) = \int_0^x a(t)dt$.

$$S(x|\Lambda) = e^{-H(x|\Lambda)} = e^{-(x)}$$

By the Law of Total Probability,

$$S(x) = Pr(X > x) = \int_0^\infty Pr(X > x|\lambda)f(\lambda)d\lambda = E_\Lambda[Pr(X > x|\Lambda)] = E[S(x|\Lambda) = M_\Lambda(-A(x))$$

1.3.4 Conditional Variance

$$Var_X(X) = Var_I(E_X[X|I]) + E_I[Var_X(X|I)]$$

1.3.5 Splices

Another way of creating distribution is by splicing them. This means using different probability distributions on different intervals in such a way that the total probability adds up to 1.

1.4 Policy Limit

To model insurance payments, define the limited loss variable

$$X \wedge u = \begin{cases} X & X < u \\ u & X \geq u \end{cases}$$

The expected value of $X \wedge u$ is called the limited expiated value.

$$E[X \wedge u] = \int_0^u xf(x)dx + u(1 - F(u)) = \int_0^u S(x)dx$$

$$E[X^k] = \int_0^\infty kx^{k-1}S(x)dx$$

$$E[(X \wedge u)^k] = \int_0^u k^{k-1}S(x)dx = \int_0^u x^k f(x)dx + u^k(1 - F(u))$$

$$E[Y \wedge u] = (1+r)E[X \wedge \frac{u}{1+r}]$$

where $Y = (1+r)X$.

1.5 Deductibles

1.5.1 Ordinary and franchise deductibles

- A policy with an ordinary deductible of d is one that pays the greater of 0 and $X-d$ for a loss of X.
- A policy with a franchise deductible of d is one that pays nothing if the loss is no greater than d, and pays the full amount of the loss if it is greater than d.

1.5.2 Payment per loss with deductible

Let X be the random variable for loss size. The random variable for the payment per loss with a deductible d is $Y^L = (X-d)_+$. The symbol $(X-d)_+$ means the positive part of $X-d$: in other words, $max(0, X-d)$.

$$E[(X-d)_+] = \int_d^\infty (x-d)f(x)dx = \int_d^\infty S(x)dx$$

The random variable $(X-d)_+$ is said to be shifted by d and censored. Censored means that you have some but incomplete, information about certain losers. IN this case, you are aware of losses below d, but don't know the amounts of such losses.

If you combine a policy with ordinary deductible d and a policy with policy limit d, the combination covers every loss entirely. In other words:

$$E[X] = E[X \wedge d] + E[(X-d)_+]$$

1.5.3 Payment per payment with deductible

The random variable for payment per payment on an insurance with an ordinary deductible is the payment per loss random variable conditioned on $X > d$, or $Y^p = (X-d)_+ | X > d$

$$F_{Y^p} = \frac{F_X(x+d) - F_X(d)}{1 - F_X(d)}$$

$$S_{Y^p}(x) = \frac{S_X(x+d)}{S_X(d)}$$

The expected value of Y^p is $E[(X-d)_+]/S(d)$. It is called the mean excess loss and is denoted by $e_X(d)$. In life contingency, it is called mean residual life or the complete life expectancy.

Special Case

Exponential An exponential distribution has no memory. This means that Y^phas the same distribution as X: it is exponential with mean θ.

Uniform If X has a uniform distribution on $(0, \theta]$, then $(X - d)_+|X > d$ has a uniform distribution on $(0, \theta - d]$.

Beta $e(d) = \frac{\theta-d}{1+b}$ $d < \theta$

Two-parameter Pareto $e(d) = \frac{\theta+d}{\alpha-1}$

Single-parameter Pareto $e(d) = \begin{cases} \frac{d}{\alpha-1} & d \geq \theta \\ \frac{\alpha(\theta-d)+d}{\alpha-1} & d \leq \theta \end{cases}$

1.6 Loss Elimination Ratio

The Loss Elimination Ratio is defined as the proportion of the expected loss which the insurer doesn't pay as a result of an ordinary deductible. In other words, for an ordinary deductible of d, it is

$$LER(d) = \frac{E[X \wedge d]}{E[X]}$$

1.7 Risk Measures and Tail Weight

A risk measure is a real-valued function of a random variable. We use the letter ρ for a risk measure; ρ is the risk measure of X. You can probably think of several real-valued functions of random variables:

- Moments $E[X], Var(X)$, etc.
- Percentiles.
- Premium principles.

$$\rho(X) = \mu_X + c\sigma_X$$

  with a suitable c may qualify as such a risk measure.

1.7.1 Coherent risk measures

Translation invariance

$$\rho(X + c) = \rho(X) + c$$

Positive homogeneity

$$\rho(cX) = c(x)$$

Subadditivity

$$\rho(X+Y) \leq \rho(X) + \rho(Y)$$

Monotonicity

$$\rho(X) \leq \rho(Y) \text{ if } Pr(X \leq Y) = 1$$

1.7.2 Value-at-Risk (VaR)

Definition. *The Value-at-Risk at security level p for a random variable X, denoted $VaR_p(X)$, is the 100pth percentile of X:*

$$VaR_p(X) = \pi_p = F_X^{-1}(p)$$

1.7.3 Tail-Value-at-Risk (TVaR)

Definition. *The tail-value-at-risk of a continuous random variable X at security level p, denoted $TVaR_p(X)$, is the expectation of the variable given that it is above its 100pth percentile:*

$$TVaR_p(X) = E[X|X > VaR_p(X)]$$

This measures is also called Conditional Tail Expectation (CTE), Tail Conditional Expectation (TCE), and Expected Shortfall (ES).

$$TVaR_p(X) = \frac{\int_{VaR_p(X)}^{\infty} xf(x)dx}{1-F(VaR_p(X))} = \frac{\int_p^1 VaR_y(X)dy}{1-p} = VaR_p(X) + e_X(VaR_p(X))$$
$$= VaR_p(X) + \frac{E[X] - E[X \wedge VaR_p(X)]}{1-p}$$

Distribution	$VaR_p(X)$	$TVaR_p(X)$
Exponential	$-\theta \ln(1-p)$	$\theta(1-\ln(1-p))$
Pareto	$\frac{\theta(1-\sqrt[\alpha]{1-p})}{\sqrt[\alpha]{1-p}}$	$E[X](1+\frac{\alpha(1-\sqrt[\alpha]{1-p})}{\sqrt[\alpha]{1-p}})$
Normal	$\mu + z_p\sigma$	$\mu + \sigma\frac{\phi(z_p)}{1-p}$
Lognormal	$e^{\mu+\sigma z_p}$	$E[X](\frac{\phi(\sigma-z_p)}{1-p})$

Note:

1. TVaR is coherent
2. $TVaR_0(X) = E[X]$
3. $TVaR_p(X) \geq VaR_p(X)$ with equality holding only if $VaR_p(X) = max(X)$

1.7.4 Tail Weight

Parametric distributions are often used to model loss size. Parametric distributions vary in the degree to which they allow for very large claims. Tail weight describes how much weight is placed on the tail of the distribution. The bigger the tail weight of the distribution, the more provision for high claims.

The following quantitative measures of tail weight are available:

1. The more positive raw or central moments exist, the less the tail weight.

2. To compare two distributions, the limits of the ratios of the survival functions, or equivalently the ratios of the density functions, can be examined as $x \to \infty$. The ratio going to infinity implies the function in the numerator has heavier tail weight.

3. An increasing hazard rate function means a lighter tail and a decreasing one means a heavier tail.

4. An increasing mean excess loss function means a heavier tail and vice versa.

1.8 Other Topics in Severity Coverage Modifications

A coverage may have both a policy limit and a deductible, and we then need to specify the order of the modifications. We distinguish the policy limit from the maximum covered loss:

Policy limit is the maximum amount that the coverage will pay . In the presence of a deductible or other modifications, perform the other modifications, then the policy limit.

Maximum covered loss is the stipulated amount considered in calculating the payment. Apply this limit first, and then deductible.

The payment per loss random variable in the presence of a maximum covered loss of u and an ordinary deductible of d is $Y^L = X \wedge u - X \wedge d$. The payment per payment is $Y^L | X > d$.

Coinsurance of α means that a portion, α, of each loss is reimbursed by insurance. The expected payment er loss if there is α coinsurance, d deductible, u maximum covered loss, and inflation rate r is

$$E[Y^L] = \alpha(1+r)(E(X \wedge \frac{u}{1+r}) - E(X \wedge \frac{d}{1+r}))$$

(we set r to zero for the non-inflation formula).

The following is for calculating the variance of payment per loss and payment per payment in the presence of an ordinary deductible.

$$(Y^L)^2 = (X \wedge u - X \wedge d)^2 = (X \wedge u)^2 - (X \wedge d)^2 + 2(X \wedge d)(X \wedge d - X \wedge u)$$

$$E[(Y^L)^2] = E[(X \wedge u)^2] - E[(X \wedge d)^2] - 2d(E[X \wedge u] - E[X \wedge d])$$

1.9 Discrete Distributions

1.9.1 The (a, b, 0) class

Discrete distributions are useful for modelling frequency. Three basic distributions are Poisson, negative binomial, and binomial.

1. A Poisson distribution with parameter $\lambda > 0$ is defined by

   $$p_n = e^{-\lambda}\frac{\lambda^n}{n!} \ \lambda > 0$$

   where the mean and variance are λ. A sum of n independent Poisson random variables $N_1, \cdots, N_n$ with parameters $\lambda_1, \cdots, \lambda_2$ has a Poisson distribution whose parameters is $\sum_{i=1}^{n} \lambda_i$.

2. A negative binomial distribution with parameters r and β is denied by

   $$p_n = \binom{n-1+r}{n}(\frac{1}{1+\beta})^r(\frac{\beta}{1+\beta})^n \ \beta > 0, r > 0$$

   where the mean is $r\beta$ and the variance is $r\beta(1+\beta)$.

3. Binomial distribution

   $$p_n = \binom{m}{n} q^n(1-q)^{m-n} \text{ m a positive integer,} 0 < q < 1$$

   where the mean is mq and variance is $mq(1-q)$.

Above three distributions form a set of $(a, b, 0)$ distribution with property

$$\frac{p_k}{p_{k-1}} = a + \frac{b}{k}$$

1.9.2 The (a, b, 1) class

The (a,b,1) class consists of distributions for which p_0 is arbitrary, but the $(a, b, 0)$ relationship holds above 1; in other words

$$\frac{p_k}{p_{k-1}} = a + \frac{b}{k} \text{ for } k = 2, 3, 4, \cdots$$

.

zero-truncated distributions Let p_n^T be the probabilities of the new distribution. We let $p_0^T = 0$ and make the rest be 1. In other words,

$$p_n^T = \frac{p_n}{1-p_0} \text{ for } n > 0$$

zero-modified distributions Let p_n^M be the probabilities of the new distributions. We give $p_0^M = c$. Then multiply p_n by $\frac{c}{1-p_0}$ so that they add up to 1.

$$p_n^M = (1-p_0^M)p_n^T \ n > 0$$

is the general relationship between zero-modified and zero-truncated distributions.

Note: for a zero-truncated geometric distribution, its mean is 1 more than the mean of an unmodified distribution, or $1+\beta$.

1.9.3 Summary

$$E[N] = cm$$

$$Var(N) = Var(E[N|I]) + E[Var(N|I)] = Var(0,m) + E[0,v] = c(1-c)m^2 + cv$$

where

- c is $1-p_0^M$
- m is the mean of the corresponding zero-truncated distribution.
- v is the variance of the corresponding zero-truncated distribution.

1.10 Frequency Distributions: Exposure& Coverage Modification

Exposure and Coverage Modifications

Model	Original Parameters	Exposure Modification	Coverage Modification
	Exposure $n_1, F(0)=1$	Exposure $n_2, F(0)=1$	Exposure $n_1, F(0)=v$
Poisson	λ	$(n_2/n_1)\lambda$	vλ
Binomial	m, q	$(n_2/n_1)m, q$	m, vq
Negative Binomial	r, β	$(n_2/n_1)r, \beta$	$r, v\beta$

These adjustments work for $(a,b,1)$ distributions as well as $(a,b,0)$ distributions. For $(a,b,1)$ distributions, p_0^M is adjusted as follows:

$$1-p_0^{M*} = (1-p_0^M)(\frac{1-p_0^*}{1-p_0})$$

1.11 Aggregate Loss Models

1.11.1 Introduction

Aggregate losses S can be expressed as

$$S = \sum_{i=1}^{N} X_i$$

where N is the number of claims and X_i is the size of each claim. X_i are iid and independent of N.

This model is called the collective risk model. S is a compound distribution: a distribution formed by summing up a rondo number of identical random variables. N is called the primary distribution and X is called the secondary distribution.

Alternatively, we have another model is called the individual risk model.

$$S = \sum_{i=1}^{n} X_i$$

where X_i's are independent, but not necessarily identically distributed random variables. Different insureds could have different distributions of aggregate losses. Typically, $Pr(X_i = 0) > 0$, since an insured may not submit any claims. This is unlike the collective risk model where X_i is a claim and therefore not equal to 0. THere is no random variable N. Instead, n is a fixed number, the size of the group.

$$Var(S) = E[N]Var(X) + Var(N)E[X]^2$$

1.11.2 Approximating Distribution

If severity is discrete, then the aggregate loss distribution is discrete, and a continuity correction is required. When we evaluate between a and b, then we do the mid point. For example, if we evaluate $Pr(X > a)$ and $Pr(X > b)$, then we do $Pr(X > (a + b)/2)$.

1.11.3 The Recursive Formula

We will use the following notations for the probability functions of the three distributions - frequency, severity, aggregate loss.

$$p_n = Pr(N = n) = f_N(n)$$
$$f_n = Pr(X = N) = f_X(n)$$
$$g_n = Pr(S = n) = f_S(n)$$

Then $F_S(x) = \sum_{n \le x} g_n$.

$$g_n = \sum_{k=0}^{\infty} p_k \sum_{i_1+\cdots+i_k=n} f_{i_1} f_{i_2} \cdots f_{i_k}$$

The product of the f_{i_t}'s is called the k-fold convolution of the f's, or f^{*k}.

For the $(a, b, 0)$ class, the formula is given:

$$g_k = \frac{1}{1 - af_0} \sum_{j=1}^{k} (a + \frac{bj}{k}) f_j g_{k-j} \quad k = 1, 2, 3, \cdots$$

For the $(a, b, 1)$ class, the formula is:

$$g_k = \frac{(p_1 - (a+b)p_0) f_k + \sum_{j=1}^{k} (a + bj/k) f_j g_{k-j}}{1 - af_0} \quad k = 1, 2, 3, \cdots$$

1.11.4 Aggregate Deductible

If frequency and severity are independent, then $E[S] = E[N]E[X]$. The expected value of aggregate losses above the deductible is called the net-stop-loss premium.

We will assume that for some h, $Pr(S = n)$ is nonzero only for n a multiple of h.

Since $E[(S - d)_+] = E[S] - E[S \wedge d]$, our first step is to calculate $E[S \wedge d]$.

Using the definition of $E[S \wedge d]$:

$$E[S \wedge d] = \sum_{j=0}^{u} hj g_{hj} + dPr(S \ge d)$$

where $u = \lceil d/h \rceil - 1$.

Calculating $E[S \wedge d]$ by integrating the survival function:

$$E[S \wedge d] = \sum_{j=0}^{u-1} hS(hj) + (d - hu)S(hu)$$

Proceeding backwards

1.11.5 Miscellaneous Topics

Coverage Modifications If you are calculating expected annual aggregate payments, you may use either of the following two formulas:

$$\text{Expected payment per loss} \times \text{Expected number of losses per year}$$

$$OR$$

$$\text{Expected payment per payment} \times \text{Expected number of payments per year}$$

Exact Calculation of Aggregate Loss Distribution

- Normal distribution. If X_i are normal with mean μ and variance σ^2, their som is normal.
- Exponential or gamma distribution. If X_i are exponential or gamma, their sum has a gamma distribution.

2 Empirical Models

2.1 Review of Mathematical Statistics

We define $\hat{\theta}$ to be the estimator, $\hat{\theta}_n$ to be the estimator based on n observations, and θ to be the parameter being estimated.

Bias is the excess of the expected value of the estimator over its true value.

$$bias_{\hat{\theta}}(\theta) = E[\hat{\theta}|\theta] - \theta$$

An estimator is unbiased if $bias_{\hat{\theta}}(\theta) = 0$. Even if an estimator is biased, it may be asymptotically unbiased,

$$\lim n \to \infty bias_{\hat{\theta}}(\theta) = 0$$

An estimator is **consistent** if it is, with probability 1, arbitrarily close to the true value if the sample is large enough. In other words, $\forall \delta > 0, \lim_{n\to\infty} Pr(|\hat{\theta}_n - \theta| < \delta) = 1$. This is sometimes called weak consistency. A sufficient but not necessary condition for consistency is that the estimator be asymptotically unbiased and that its variance goes to zero asymptotically as the sample size goes to infinity.

Mean square error is the average square difference between the estimator and the true value of the parameter, or

$$MSE_{\hat{\theta}}(\theta) = E[(\hat{\theta} - \theta)^2|\theta]$$

The lower the MSE, the better the estimator.

An estimator is called a uniformly minimum variance unbiased estimator (UMVUE) if it is unbiased and if there is no other unbiased estimator with a smaller variance for any true value θ.

An important relationship:

$$MSE_{\hat{\theta}}(\theta) = Var(\hat{\theta}) + (bias_{\hat{\theta}}(\theta))^2$$

2.1.1 Hypothesis Testing

To decide on a hypothesis, we set up two hypothesis: a null hypothesis, one that we will believe unless proved otherwise, and an alternative hypothesis. A boundary point is called the critical value such that, let the point be c, then we reject H_0 if $X < c$. The set of values for which we reject H_0 is called the critical region. The lower we make c, the more likely we will accept H_0; vise versa.

Rejecting H_0 when it is true is called a Type I error. The probability of a type I error, assuming H_0 is true is the significance level of the test. We use α to represent the significance level. The precise probability of getting the observed statistic given that the null hypothesis is true is called the p-value.

If we reject H_1 when it is true, it is a Type II error. The power of a test is the probability of rejecting H_0 when it is false. A uniformly most powerful test gives us the most power for a fixed significance level.

The last thing to mention is confidence intervals.

$$1 - \alpha = Pr(-z_{\alpha/2} \leq \frac{\hat{\theta} - \theta}{\sqrt{v(\theta)}} \leq z_{\alpha/2})$$

where $1 - \alpha$ is the confidence level and z_α is the $100(1 - \alpha)$th percentile.

2.2 The Empirical Distribution for Complete Data

Complete data for a study means that every relevant observation is available and the exact value of every observation is known.

2.2.1 Individual Data

$F_n(x)$ is the empirical cumulative distribution function, $f_n(x)$ is the empirical probability or probability density function, and so on. Since the empirical distribution for individual data is discrete, $f_n(x)$ would be the probability of x, and would equal k/n, where k is the number of x_i in the sample equal to x. The empirical cumulative hazard function is $H_n(x) = -\ln S_n(x)$. Grouped data Suppose we have grouped data that has a set of intervals and the number of losses in each interval. Then to generate the cumulative distribution function for all points, we connect the dots. We interpolate linearly between endpoints of intervals. The resulting distribution function is denoted by $F_n(x)$ and is called the ogive.

The derivative of the ogive is denoted by $f_n(x)$. It is the density function corresponding to the ogive, and is called the histogram.

$$f_n(x) = \frac{n_j}{n(c_j - c_{j-1})}$$

where x is in the interval $[c_{j-1}, c_j)$, there are n_j points in the interval, and n points altogether.

2.3 Variance of Empirical Estimators with Complete Data

2.3.1 Individual data

If the empirical distribution is being used as the model with individual data, then $S_n(x)$ is the proportion of observation above x. Since the probability of an observation being above x is $S(x)$, $S_n(x)$ is a binomial proportion random variable with parameters $m = n$ and $q = S(x)$; its variance is therefore

$$Var(S_n(x)) = \frac{S(x)(1 - S(x))}{n}$$

Since we don't know $S(x)$, we estimate the variance using $S_n(x)$:

$$\hat{Var}(S_n(x)) = \frac{S_n(x)(1 - S_n(x))}{n}$$

If n_x is the observed number of survivors past time x, then $S_n(x) = n_x/n$. Hence

$$\hat{Var}(S_n(x)) = \frac{n_x(n - n_x)}{n^3}$$

2.3.2 Grouped data

Let Z be the number of observations in the interval $(c_{j-1}, c_j]$, then

$$\hat{Var}(S_n(x)) = \frac{\hat{Var}(Y)(c_j - c_{j-1})^2 + \hat{Var}(Z)(x - c_{j-1})^2 + 2\hat{Cov}(Y, Z)(c_j - c_{j-1})(x - c_{j-1})}{n^2(c_j - c_{j-1})^2}$$

and

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

where

$$\hat{Var}(Y) = \frac{Y(n - Y)}{n}$$

$$\hat{Var}(Z) = \frac{Z(n - Z)}{n}$$

$$\hat{Cov}(Y, Z) = -\frac{YZ}{n}$$

2.4 Kaplan-Merier and Nelson-Åalen Estimators

Two scenarios such that data may be incomplete:

1. No information at all is provided for certain ranges of data. The data are not provided for a range, the data is said to be truncated. (deductible is truncated below)
2. The exact data point is not provided; instead, a range is provided. When a range of values rather than an exact value is provided, the data is said to be censored. (policy limit is censored from above)

2.4.1 Kaplan-Meier Product Limit Estimator

$$S_n(t) = \prod_{i=1}^{j-1}(1 - \frac{s_i}{r_i}),\ y_{j-1} \leq t < y_i$$

where r_i is called the risk set at time y_i.

2.4.2 Nelson-Åalen Estimator

$$\hat{H}(t) = \sum_{i=1}^{j-1}\frac{s_i}{r_i},\ y_{i-1} \leq t < y_i$$

2.4.3 Exponential Extrapolation

$$\hat{S}(t) = \hat{S}(t_0)^{t/t_0}, t > t_0$$

2.5 Estimation of Related Quantities

2.5.1 Complete data

When using the empirical distribution as the model, do not divide by n-1 when calculating the variance. When the distribution is estimated from grouped data, moments are calculated using the ogive or histogram. The mean can be calculated as the average of the averages-sum up the averages of the groups, weighted by the probability of being in the group. Higher moments may require integration.

2.5.2 Incomplete data

When data are incomplete due to censoring or truncation, the product limit estimator or the Nelson-Åalen estimator is used. You can estimate S(x) to obtain:

$$E[X] = \int_0^\infty S(x)dx$$

$$E[X \wedge d] = \int_0^d S(x)dx$$

In fact, S(x) is estimated. Hence

$$\int_0^\infty \hat{S}(x)dx = \sum_{j=0}^{\infty} \hat{S}(y_j t)(y_{j+1} - y_j)$$

2.5.3 Deductibles and limits

For a deductible of d and a maximum covered claim of u, the average payment per loss is

$$E[X \wedge u] - E[X \wedge d]$$

and the average payment per payment is

$$\frac{E[X \wedge u] - E[X \wedge d]}{1 - F(d)}$$

2.6 Variance of Kaplan-Meier and Nelson-Åalen Estimators

The Kaplan-Meier estimator is an unbiased estimator of the survival function. Greenwood's approximation of the variance is:

$$V\hat{a}r(\hat{S}(t)) = \hat{S}(t)^2 \sum_{y_i \leq t} \frac{s_j}{r_j(r_j - s_j)}$$

A useful fact to remember is that if there is complete data-no censoring or truncation-the Greenwood approximation is identical to the empirical approximation of the variance.

The approximation variance of the Nelson-Åalen estimator is

$$V\hat{a}r(\hat{H}(t)) = \sum_{y_j \leq t} \frac{s_j}{r_j^2}$$

2.7 Kernel Smoothing

The kernel-smoothed distribution is also an equally weighted mixture of n distributions. The kernel-smoothed distribution consisting of selecting a distribution used for each sample point x_i. Let $K_{x_i}(x)$ be the cumulative distribution function of the distribution used for the point x_i, evaluated at x. Let $k_{x_i}(x)$ be the probability density function. Then the kernel-smoothed distribution function is

$$\hat{F}(x) = \sum_{i=1}^{n} (\frac{1}{n}) K_{x_i}(x)$$

and the kernel-smoothed density function is

$$\hat{f}(x) = \sum_{i=1}^{n} (\frac{1}{n}) k_{x_i}(x)$$

2.8 Approximations for Large Data Sets

d_j is the number of left truncated observations in the interval $[c_j, c_{j+1})$.

u_j is the number of right censored observations in the interval $(c_j, c_{j+1}]$

x_j is the number of events in the interval $(c_j, c_{j+1}]$

r_j is the risk set to use for calculating the conditional mortality rate in the interval $(c_j, c_{j+1}]$

q'_j is the decrement rate in the interval $(c_j, c_{j+1}]$, and is computed as $q'_j = x_j / r_j$

The sample size n can be expressed as $\sum_{j=0}^{k-1} d_j$ or as $\sum_{j=0}^{k-1} (u_j + x_j)$. Let P_j be the population at time c_j. Hence

$$P_j = \sum_{i=0}^{j-1} (d_i - u_i - x_i)$$

3 Parametric Models

4 Credibility

4.1 Limited Fluctuation Credibility: Poisson Frequency

4.1.1 The Algorithm

1. You make assumption for the mean and variance of claim size and claim frequency. Often claim frequency is assumed to be Poisson.

2. You establish credibility standards based on two parameters: the probability of being in a certain interval, which is something like a confidence level, and the size of the interval you want to be in, which expressed as a percentage of the mean.

3. You determine how many exposures, or claims, or aggregate claim amounts, you would need to satisfy this standard and grant full credibility.

4. If full credibility cannot be granted, you determine what percentage of credibility can be granted.

4.1.2 The formula

Let e_F be the exposure needed for full credibility, μ be the expected aggregate claims per exposure, σ be the standard deviation per exposure, y_p be the coefficient from the standard normal distribution for the confidence interval which you desire, k be the maximum fluctuation you will accept. Sometimes, confidence level p is called the probability parameter and k is called the range parameter. Hence

$$y_p \frac{\sqrt{e_F \sigma^2}}{e_F \mu} = k$$

$$e_F = n_0 (\frac{\sigma}{\mu})^2 = n_0 CV^2$$

There are three things we can calculate credibility for:

1. Number of claims. This means we want to number of claims to be within k of expected p of the time.
2. Claim sizes. This means we want the size of each claim to be within k of expected p of the time
3. Aggregate losses or pure premium. This means we want aggregate losses (or pure premium, which is aggregate losses per exposure) to be within k of expected p of that time.

4.2 Bayesian Estimation and Credibility-Discrete Prior

5 Simulation

5.1 Simulation-Inversion Method

5.2

References

[Weishaus] Abraham Weishaus. Study Manual for Exam C/Exam 4, Construction and Evaluation of Actuarial Models, Thirteenth Edition