Actuarial Statistics

Part III Course, Lent 2010

Revision Notes

Daniel Guetta
guetta@cantab.net

Based on lectures by Susan Pitts, Lent 2010

Loss Distributions

A loss is the value of actual damage caused by the insured-against event. We treat this loss as a positive random variable.

Common Distributions

Here is a table of common distributions. Note that

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} \, \mathrm{d}x \qquad (\alpha > 0)$$

(for integers, $\Gamma(n) = (n-1)!$).

	Density f	Tail $\bar{F}(x) = \mathbb{P}(X > x)$	$\mathbb{E}(X)$	$\mathrm{var}(X)$	$\mathbb{E}(X^r)$	$M(t) = \mathbb{E}\left(e^{tX}\right)$
Exponential ($\lambda > 0$)	$\lambda e^{-\lambda x}$ $(x > 0)$	$e^{-\lambda x}$ $(x > 0)$	$1/\lambda$	$1/\lambda^2$	$\frac{\Gamma(r+1)}{\lambda^r}$	$\frac{\lambda}{\lambda - t}$ $(t < \lambda)$
Gamma ($\alpha > 0, \lambda > 0$)	$\frac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)}$ $(x > 0)$		α/λ	α/λ^2	$\frac{\Gamma(\alpha+r)}{\lambda^r \Gamma(\alpha)}$	$\left(\frac{\lambda}{\lambda - t}\right)^\alpha$ $(t < \lambda)$
	Note: (a) $X \sim \Gamma(1,\lambda) \Leftrightarrow X \sim \exp(\lambda)$ **(b)** $X \sim \Gamma(\alpha,\lambda) \Leftrightarrow 2\lambda X \sim \chi^2_{2\alpha}$					
Weibull ($a > 0,\ b > 0$)	$abx^{b-1} \exp\left(-ax^b\right)$ $(x > 0)$	$\exp\left(-ax^b\right)$ $(x > 0)$	$\mathbb{E} = a^{-1/b}\Gamma(1+b^{-1})$ $\mathrm{var} = a^{-2/b}\left\{\Gamma(1+2b^{-1}) - \Gamma^2(1+b^{-1})\right\}$			
	Notes: (a) $X \sim W(a,1) \Leftrightarrow X \sim \exp(a)$ **(b)** $X \sim W(a,b) \Leftrightarrow aX^b \sim \exp(1)$					
Normal ($\mu, \sigma^2 > 0$)	$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)$		μ	σ^2		$\exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right)$
Lognormal ($\mu, \sigma^2 > 0$)	$\frac{1}{x}\phi(\log x)$ $(x > 0)$	$1 - \Phi\left(\frac{\log x - \mu}{\sigma}\right)$ $(x > 0)$	$\mathbb{E} = \exp\left(\mu + \frac{1}{2}\sigma^2\right)$ $\mathrm{var} = \exp\left(2\mu + 2\sigma^2\right) - \exp\left(2\mu + \sigma^2\right)$			
Pareto ($\alpha > 0, \lambda > 0$)	$\frac{\alpha\lambda^\alpha}{\left(\lambda + x\right)^{\alpha+1}}$ $(x > 0)$	$\left(\frac{\lambda}{\lambda + x}\right)^\alpha$ $(x > 0)$	k^{th} moment only exists if $\alpha > k$. $\mathbb{E} = \frac{\lambda}{\alpha - 1}$ $(\alpha > 1)$ $\mathrm{var} = \frac{\alpha\lambda^2}{\left(\alpha-1\right)^2\left(\alpha-2\right)}$ $(\alpha > 2)$			
	Notes: **(a)** often given in translated form with density $f(x) = \frac{\alpha\lambda^\alpha}{x^{\alpha+1}}$ for $x > \lambda$ **(b)** often used for modelling tails of loss distributions (=large claims) **(c)** arises as exponential distribution where parameter is mixed over gamma distribution.					
Pareto (three param)	$f(x) = \frac{\Gamma(\alpha+k)\lambda^\alpha x^{k-1}}{\Gamma(\alpha)\Gamma(k)(\lambda+x)^{\alpha+k}}$ $(x > 0)$					

	Notes: when $k = 1$, we recover the two parameter form.					
Burr	$\dfrac{\alpha\gamma\lambda^{\alpha}x^{\gamma-1}}{(\lambda + x^{\gamma})^{\alpha+1}}$ $(x > 0)$	$\dfrac{\lambda^{\alpha}}{\left(\lambda + x^{\gamma}\right)^{\alpha}}$ $(x > 0)$				
	Notes: when $\gamma = 1$, we recover the two-parameter Pareto.					

I can't really be bothered to make a similar table for discrete distributions. But they're available everywhere. Common ones are the Bernoulli, binomial, geometric, negative binomial and Poisson.

Notation

- Let X be the loss, a positive random variable with **distribution function (DF)** F so that $F(x) = \mathbb{P}(X \le x)$, and density f.
- The **moment generating function (MGF)** of X is
$$M(t) = \mathbb{E}(e^{tX})$$
It certainly exists for $t \le 0$ if X is positive, but might not exist for some, or all $t > 0$.
- The **r^{th} moment of X**, $\mathbb{E}(X^r)$ may be found by direction integration, $\mathbb{E}(X^r) = \int x^r f(x)\,\mathrm{d}x$ or using
$$\mathbb{E}(X^r) = M^{(r)}(0) = \left.\frac{\mathrm{d}^r M}{\mathrm{d}t^r}\right|_{t=0}$$
- Let $\mu = \mathbb{E}(X)$. Assume μ is finite, and define the **r^{th} central moment of X** as
$$\mu_r = \mathbb{E}\left[\left(X - \mu\right)^r\right]$$
(The notation μ_r is non-standard). In this notation, $\mu_2 = \operatorname{var}(X)$.
- Let $\kappa(t) = \ln M(t)$ be the cumulant generating function of X. For two independent random variables X and Y, $\kappa_{X+Y}(\theta) = \kappa_X(\theta) + \kappa_Y(\theta)$.
- The r^{th} cumulant of X is
$$\kappa_r = \kappa^{(r)}(0) = \left.\frac{\mathrm{d}^r \kappa}{\mathrm{d}t^r}\right|_{t=0}$$
(so again, the cumulants of independent random variables are additive). We find

$$\kappa_1 = \mu$$
$$\kappa_2 = \mu_2 = \mathrm{var}(X)$$
$$\kappa_3 = \mu_3 = \text{skewness}$$
$$\kappa_4 = \mu_4 - 3\mu_2^2$$

The standardised 3rd cumulant is $\kappa_3 / \mu_2^{3/2}$ is the **skewness** or **coefficient of skewness**. If f is symmetric, the skewness is 0:

Positively skewed   Negatively skewed

Loss distributions are typically positively skewed, with heavy tails.

- The **probability generating function** of a random variable X is given by

$$\mathcal{G}_X(z) = \mathbb{E}(z^X) = M_X\left(\log z\right)$$

Mixed Distributions

EXAMPLE: Each policy holder in a portfolio has loses that are exponentially distributed, but each with a different expectation. We model this assuming a distribution of the parameter

$$X \mid \lambda \sim \exp(\lambda) \qquad f_X(x) = \lambda e^{-\lambda x} \qquad x > 0$$
$$\lambda \sim \Gamma(\alpha, \theta) \qquad f_\lambda(\lambda) = \tfrac{\theta^\alpha \lambda^{\alpha-1} e^{-\theta\lambda}}{\Gamma(\alpha)} \qquad \lambda > 0$$

We say that X has a **mixed distribution**. $\Gamma(\alpha,\theta)$ is the **mixing distribution** and we say λ is **mixed over** $\Gamma(\alpha,\theta)$.

We then have that

$$\begin{aligned}\mathbb{P}\left(X > x\right) &= \int_0^\infty \mathbb{P}(X > x \mid \lambda) f_\lambda(\lambda)\,\mathrm{d}\lambda \\ &= \int_0^\infty \tfrac{e^{-\lambda x}\left(\theta^\alpha \lambda^{\alpha-1} e^{-\theta\lambda}\right)}{\Gamma(\alpha)}\,\mathrm{d}\lambda \\ &= \tfrac{\theta^\alpha}{(\theta+x)^\alpha} \int_0^\infty \frac{(\theta+x)^\alpha \lambda^{\alpha-1} e^{-(\theta+x)\lambda}}{\Gamma(\alpha)}\,\mathrm{d}\lambda \\ &= \frac{\theta^\alpha}{(\theta+x)^\alpha}\end{aligned}$$

(Where, in the last line, we used the fact that the quantity in the integral is a Γ density).

We then have

$$f_X(x) = -\frac{\mathrm{d}}{\mathrm{d}x}\left\{1 - F_X(x)\right\} = -\frac{\mathrm{d}}{\mathrm{d}x}\mathbb{P}(X > x) = \frac{\alpha\theta^\alpha}{(x+\theta)^{\alpha+1}}$$

This is a two-parameter Pareto.

Morals of the example:

1. Identify a density to set an integral to 1.
2. If you have $\mathbb{P}(X > x)$ rather than $\mathbb{P}(X < x)$, use the fact that $\frac{\mathrm{d}}{\mathrm{d}x}F(x) = -\frac{\mathrm{d}}{\mathrm{d}x}\left\{1 - F(x)\right\}$ to calculate f_X. □

Fitting loss distributions to data

Standard statistical methods are used (like maximum likelihood estimation). Bayesian methods can also be used.

<u>EXAMPLE (*truncated data*)</u>: Let X be a random variable with density f_X and DF F_X. Assume the distribution is such that $\mathbb{P}(X > d) > 0$ and let $Y = X \mid X > d$.

We then have

$$\begin{aligned} F_Y(x) = \mathbb{P}\left(Y \le x\right) &= \mathbb{P}\left(X \le x \mid X > d\right) \\ &= \begin{cases} 0 & x \le d \\ \frac{\mathbb{P}(d < X \le x)}{\mathbb{P}(X > d)} & x > d \end{cases} \\ &= \begin{cases} 0 & x \le d \\ \dfrac{F_x(x) - F_x(d)}{1 - F_x(d)} & x > d \end{cases} \end{aligned}$$

So

$$f_Y(x) = \frac{\mathrm{d}}{\mathrm{d}x}F_Y(x) = \begin{cases} 0 & x \le d \\ \dfrac{f_X(x)}{1 - F_x(d)} & x > d \end{cases}$$

Now, if we observe $Y_1, \cdots, Y_n$, we have

$$\begin{aligned} \ell_n(\theta) &= \sum_{i=1}^n \ln f_Y(y;\theta) \\ &= \left\{\sum_{i=1}^n \ln f_X(y_i;\theta)\right\} - n\ln\left[1 - F_x(d;\theta)\right] \end{aligned}$$

□

We sometimes use plots in the exploratory stages of fitting

Definition (mean residual life): The *mean residual life* at x of a random variable X is

$$e(x) = \mathbb{E}\Big(X - x \mid X > x\Big)$$

It is the case that

$$e(x) = \int_0^\infty y \frac{f_X(x+y)}{1 - F_X(x)} \, \mathrm{d}y = \frac{\int_x^\infty 1 - F_X(w) \, \mathrm{d}w}{1 - F_X(x)}$$

Proof: In this case, $Y = X - x \mid X > x$. Using the method in the last example:

$$\begin{aligned} \mathbb{P}(Y \le y) &= \mathbb{P}(X \le y + x \mid X > x) \\ &= \mathbb{P}\Big(x < X < y + x\Big) / \mathbb{P}(X > x) \\ &= \Big\{F_X(x+y) - F_X(x)\Big\} / \Big\{1 - F_X(x)\Big\} \end{aligned}$$

So

$$\begin{aligned} f_Y(y) &= f_x(x+y) / \Big\{1 - F_X(x)\Big\} \\ \mathbb{E}\Big(Y\Big) &= \int_0^\infty y \frac{f_X(x+y)}{1 - F_X(x)} \, \mathrm{d}y = e(x) \end{aligned}$$

This can then be re-written in a more convenient form

$$\begin{aligned} e(x) &= \frac{\int_0^\infty y f_X(x+y) \, \mathrm{d}y}{1 - F_X(x)} \\ &= \frac{\int_x^\infty (w - x) f_X(w) \, \mathrm{d}w}{1 - F_X(x)} \\ &= \frac{\int_x^\infty (w - x) \frac{\mathrm{d}}{\mathrm{d}w}\Big[-(1 - F(w))\Big] \mathrm{d}w}{1 - F_X(x)} \\ &= \frac{\Big[-(w - x)(1 - F(w))\Big]_{w=x}^\infty + \int_x^\infty (1 - F(w)) \, \mathrm{d}w}{1 - F_X(x)} \\ &= \frac{\int_x^\infty 1 - F_X(w) \, \mathrm{d}w}{1 - F_X(x)} \end{aligned}$$

Where we have assumed, in the last line, that the tail is small enough to ensure $w(1 - F(w)) \to 0$ as $w \to \infty$. As required. ■

Definition (empirical mean residual life): The *empirical mean residual life* of a sample $X_1, \cdots, X_n$ is $e_n(x)$. It is the mean residual life of a distribution that puts mass $1/n$ at each point $X_1, \cdots, X_n$ in our sample. It is given by

$$\begin{aligned} e_n(x) &= \left\{ \frac{1}{\#(X_i > x)} \sum_{X_i > x} X_i \right\} - x \\ &= \left\{ \begin{matrix} \text{Mean of all } X_i \\ \text{greater than } x \end{matrix} \right\} - x \end{aligned}$$

Proof: From the definition above, we have

$$e(x) = \frac{1}{1 - F_X(x)} \int_0^\infty y f_X(x+y) \, \mathrm{d}y$$

Substitute $u = x + y$ into the integral:

$$\begin{aligned} e(x) &= \frac{1}{1 - F_X(x)} \int_x^\infty (u - x) f_X(u) \, \mathrm{d}u \\ &= \frac{1}{1 - F_X(x)} \left[\int_x^\infty u f_X(u) \, \mathrm{d}u - \int_x^\infty x f_X(u) \, \mathrm{d}u \right] \end{aligned}$$

Now, since our distribution puts mass $1/n$ at each point $X_1, \cdots, X_n$, $f_X(x)$ will be $1/n$ when x is one of these points, and 0 otherwise. The integral above therefore becomes

$$e(x) = \frac{1}{1 - F_X(x)} \frac{1}{n} \left[\left(\sum_{X_i > x} X_i \right) - x \#(X_i > x) \right]$$

We also note that $1 - F_X(x) = \mathbb{P}(X > x)$, which is simply equal to $1/n$ multiplied by the number of X_i which are indeed greater than x:

$$e(x) = \frac{1}{\frac{1}{n} \#(X_i > x)} \frac{1}{n} \left[\left(\sum_{X_i > x} X_i \right) - x \#(X_i > x) \right]$$

Re-arranging, we obtain

$$e(x) = \frac{1}{\#(X_i > x)} \sum_{X_i > x} X_i - x$$

As required. ■

We usually plot the empirical mean residual life against x, and then compare it to $e(x)$ for some known distributions.

Compound Distributions

Definition (compound distribution): Let $X_1, X_2, \cdots$ be IID random variables and let N be a random variable taking values in $\{0, 1, 2, \cdots\}$ independently of the X_i. Then the **random sum**

$$S = X_1 + \cdots + X_N$$

is said to be a **compound distribution**.

Moments & Distributions

For the random sum S defined above, it is the case that:

$$\begin{aligned} \mathbb{E}(S) &= \mathbb{E}\left(X_1\right)\mathbb{E}\left(N\right) \\ \operatorname{var}\left(S\right) &= \mathbb{E}\left(N\right)\operatorname{var}\left(X_1\right) + \operatorname{var}\left(N\right)\left[\mathbb{E}\left(X_1\right)\right]^2 \end{aligned} \quad (2.1)$$

Proof: First, the mean

$$\mathbb{E}(S) = \mathbb{E}\left(\mathbb{E}(S \mid N)\right) = \mathbb{E}\left(N\mathbb{E}(X_1)\right) = \mathbb{E}(X_1)\mathbb{E}(N)$$

then, the variance. We first derive the **conditional variance formula**:

$$\begin{aligned} \operatorname{var}(S) &= \mathbb{E}(S^2) - \mathbb{E}^2(S) \\ &= \mathbb{E}\left(\mathbb{E}(S^2 \mid N)\right) - \left\{\mathbb{E}\left(\mathbb{E}(S \mid N)\right)\right\}^2 \\ &= \mathbb{E}\left(\operatorname{var}(S \mid N) + \mathbb{E}^2(S \mid N)\right) - \left\{\mathbb{E}\left(\mathbb{E}(S \mid N)\right)\right\}^2 \\ &= \mathbb{E}\left(\operatorname{var}(S \mid N)\right) \\ &\qquad + \left\{\mathbb{E}\left(\left\{\mathbb{E}(S \mid N)\right\}^2\right) - \left\{\mathbb{E}\left(\mathbb{E}(S \mid N)\right)\right\}^2\right\} \\ &= \mathbb{E}\left(\operatorname{var}(S \mid N)\right) + \operatorname{var}\left(\mathbb{E}(S \mid N)\right) \end{aligned}$$

we then note that, since all the X are independent:

$$\operatorname{var}(S \mid N) = \operatorname{var}(X_1 + \cdots + X_N \mid N) = N \operatorname{var}(X_1)$$

and so

$$\begin{aligned} \operatorname{var}(S) &= \mathbb{E}\left(N \operatorname{var}(X_1)\right) + \operatorname{var}\left(N\mathbb{E}(X_1)\right) \\ &= \operatorname{var}(X_1)\mathbb{E}(N) + \left[\mathbb{E}(X_1)\right]^2 \operatorname{var}(N) \end{aligned}$$

as required. ■

We can find the distribution of S either using convolutions of MGFs (ie: transforms).

- **Convolutions**

> **Definition (n-fold convolution of F)**: Let X have the distribution function F. The *n-fold convolution of F*, denoted F^{*n}, is the distribution of $X_1 + \cdots + X_n$.
>
> We define
> $$F^{*0}(x) = \begin{cases} 0 & x < 0 \\ 1 & x \geq 0 \end{cases}$$
> and we have that
> $$F^{*k}(x) = \int F^{*(k-1)}(x-t) f(t)\ \mathrm{d}t$$

(The general form given above can be intuited from the first few convolutions:

$$\begin{aligned} F^{*1}(x) &= F(x) \\ F^{*2}(x) &= \mathbb{P}\left(X_1 + X_2 \leq x\right) = \int \mathbb{P}\left(X_1 + X_2 \leq x \mid X_1 = t\right) f(t)\ \mathrm{d}t \\ &= \int \mathbb{P}\left(X_2 \leq x - t\right) f(t)\ \mathrm{d}t \\ &= \int F(x-t) f(t)\ \mathrm{d}t \end{aligned}$$

)

The distribution function of S is then

$$\begin{aligned} F_s(x) &= \mathbb{P}(S \leq x) \\ &= \sum_{n=0}^{\infty} \mathbb{P}\left(S \leq x \mid N = n\right) \mathbb{P}\left(N = n\right) \\ &= \sum_{n=0}^{\infty} F^{*n}(x) \mathbb{P}\left(N = n\right) \end{aligned}$$

This expression is hard to use because of the infinite sum and the recursive integral.

Note, however, that if X is non-negative, then $F(0) = 0$ and

$$\mathbb{P}\left(S = 0\right) = \mathbb{P}\left(N = 0\right)$$

thus, if the claim sizes are nonnegative, S has an atom at 0 of size $\mathbb{P}\left(N = 0\right)$.

- **Moment generating functions**

> Let $\mathcal{G}_N(z) = \mathbb{E}\left(z^N\right)$ be the **probability generating function** (PGF) of N, and let $M_X(u) = \mathbb{E}\left(e^{uX_1}\right)$ be the MGF of X_1. Then the MGF of S is
>
> $$M_S(u) = \mathcal{G}_N\left[M_X(u)\right] \tag{2.2}$$

Proof: We have

$$\begin{aligned} M_S(u) = \mathbb{E}(e^{uS}) &= \mathbb{E}\left(\mathbb{E}(e^{uS} \mid N)\right) \\ &= \mathbb{E}\left(\mathbb{E}\left(e^{uX_1}\cdots e^{uX_n} \mid N\right)\right) \\ &= \mathbb{E}\left(\left\{\mathbb{E}\left(e^{uX_1}\right)\right\}^N\right) \\ &= \mathbb{E}\left(\left\{M_X(u)\right\}^N\right) \\ &= \mathcal{G}_N\left(M_X(u)\right) \end{aligned}$$

As required. ■

Note as well that since $\mathcal{G}_X(z) = M_X\left(\log z\right)$, we also have

$$\kappa_S(\theta) = \log M_S(\theta) = \log \mathcal{G}_N\left(M_X(\theta)\right) = \log M_N\left(\log M_X(\theta)\right) = \kappa_N(\kappa_X(\theta))$$

Depending on the situation, the cumulant generating function might be easier to use than the moment generating function.

In some cases, this allows us to work out the distribution of S directly.

<u>EXAMPLE</u>: Suppose N is geometric so that $\mathbb{P}(N = n) = q^n p$ where $p \in [0,1]$ and $q = 1 - p$. Then $\mathcal{G}_N(z) = p / \left(1 - qz\right)$ and $\mathbb{E}(N) = q / p$.

Suppose $X_1 \sim \exp\left(1/\mu\right)$ so that $f_X(x) = \mu^{-1}e^{-x/\mu}$ then $M_X(u) = \left(1 - \mu u\right)^{-1}$.

From the (2.2), we have

$$
\begin{aligned}
M_S(u) = \mathcal{G}_N\left(M_X(u)\right) &= \frac{p}{1-q\frac{1}{1-\mu u}} \\
&= \frac{p\left(1-\mu u\right)}{1-\mu u - q} \\
&= \frac{p\left(1-\mu u\right)}{p-\mu u} \\
&= p\left(\frac{p-\mu u}{p-\mu u}\right) + \frac{p-p^2}{p-\mu u} \\
&= p \underbrace{1}_{\substack{\text{MGF of } Y \text{ if } Y \\ \text{is s.t. } \mathbb{P}(Y=0)=1}} + \left(1-p\right) \underbrace{\frac{1}{1-\frac{\mu}{p}u}}_{\substack{\text{MGF of } W \text{ if} \\ W\sim\exp\left(p/\mu\right)}} \quad (*)
\end{aligned}
$$

We note that the following three statements are equivalent:

$$
\begin{aligned}
F_Z &= pF_X + \left(1-p\right)F_Y \\
f_Z &= pf_X + \left(1-p\right)f_Y \\
M_Z &= pM_X + \left(1-p\right)M_Y
\end{aligned}
$$

and so (*) implies that

$$F_S(x) = pF_Y(x) + qF_W(x)$$

in other words, the distribution of S is a discrete mixture of the exponential distribution and the distribution with an atom at 0. Note that

$$\mathbb{E}(S) = \left(p\cdot 0\right) + \left(q\cdot\tfrac{\mu}{p}\right) = \tfrac{q}{p}\mu = \mathbb{E}(N)\mathbb{E}(X_1)$$

in accordance with equation (2.1). Also

$$
\begin{aligned}
F_S(x) &= \begin{cases} p + q\int_0^x \frac{p}{\mu}e^{-pt/\mu}\ \mathrm{d}t & x \geq 0 \\ 0 & x \leq 0 \end{cases} \\
&= 1 - qe^{-px/\mu} \qquad (x \geq 0)
\end{aligned}
$$

□

Common choices of N in insurance

Common examples of distributions that are used for N are geometric, negative binomial, binomial, mixed Poisson, etc…, leading to **compound** Poisson, compound geometric, etc… Some examples:

1. For a group life insurance policy covering m lives, the distribution of N (= # deaths in 1 year, of 1 year policy) is **binomial** if we assume that

each life is subject to the same mortality rate, and that the deaths are independent.

2. Suppose that $N \mid \lambda \sim \text{Po}(\lambda)$ and $\lambda \sim$ some distribution with density f_λ, then N has a **mixed Poisson distribution**, and

$$\begin{aligned}\mathbb{P}(N=n) &= \int \mathbb{P}(N=n \mid \lambda) f_\lambda(\lambda)\,\mathrm{d}\lambda \\ &= \int \frac{e^{-\lambda}\lambda^n}{n!} f_\lambda(\lambda)\,\mathrm{d}\lambda\end{aligned}$$

if $\lambda \sim \Gamma(k,\delta)$ then $M_\lambda(u) = \left(\frac{\delta}{\delta-u}\right)^k$, and N has probability generating function

$$\begin{aligned}\mathbb{E}(z^N) &= \mathbb{E}\left(\mathbb{E}(z^N \mid \lambda)\right) \\ &= \mathbb{E}\left[e^{\lambda(z-1)}\right] = M_\lambda(z-1) \\ &= \left(\frac{\delta}{\delta-(z-1)}\right)^k \\ &= \left(\frac{\delta/(\delta+1)}{1-\frac{1}{\delta+1}z}\right)^k \\ &= \left(\frac{p}{1-qz}\right)^k \qquad \left[p=\frac{\delta}{\delta+1} \qquad q=1-p \qquad p\in[0,1]\right]\end{aligned}$$

This is the PGF of a negative binomial with parameters k and $p=\delta/(\delta+1)$, so

$$\mathbb{P}(N=n) = \binom{n+k-1}{n} q^n p^k \qquad n=0,1,2,\cdots$$

note that for a negative binomial, $\text{var} = kq/p^2 > kq/p = \mathbb{E}$. In practice, this often gives a better fit to data than a Poisson distribution.

Important properties of independent compound Poisson distributions

Suppose $S_1,\cdots,S_n$ (n fixed) are independent compound Poisson random variables with Poisson parameters $\lambda_1,\cdots,\lambda_n$, and the claim sizes for each of the sums have distribution functions $F_1,\cdots,F_n$. Let M_i be the MGF belonging to the claim size for compound variable i. Let $S = S_1+\cdots+S_n$ and $\lambda = \lambda_1+\cdots+\lambda_n$.

$$\begin{aligned}
M_S(y) = \mathbb{E}\left(e^{uS}\right) &= \mathbb{E}\left(e^{u(S_1+\cdots+S_n}\right) \\
&= \prod_{i=1}^{n} \mathbb{E}\left(e^{uS_i}\right) \qquad \leftarrow (\text{MGF of } S_i) \\
&= \prod_{i=1}^{n} \mathcal{G}_N\left[M_i(u)\right] \\
&= \prod_{i=1}^{n} \exp\left\{\lambda_i\left(M_i(u)-1\right)\right\} \\
&= \exp\left[\left(\sum_{i=1}^{n} \lambda_i M_i(u)\right) - \lambda\right] \\
&= \exp\left[\lambda\left(\left\{\sum_{i=1}^{n} \frac{\lambda_i}{\lambda} M_i(u)\right\} - 1\right)\right]
\end{aligned}$$

thus, we see that the sum has itself a compound Poisson distribution with Poisson parameter λ. Note also that the multipliers in the sum all sum to 1 ($\sum_i \lambda_i / \lambda = 1$), so we have a discrete mixture, and the equivalent claim size distribution function of the sum is $F = \sum_{i=1}^{n} \frac{\lambda_i}{\lambda} F_i$.

Numerical Methods ~ Panjer Recursion

Assume X_1 takes values in $\left\{1,2,3,\cdots\right\}$ and let $f_k = \mathbb{P}\left(X_1 = k\right)$. Let also $p_n = \mathbb{P}(N = n)$, and assume that

$$p_n = \left(a + \frac{b}{n}\right) p_{n-1} \qquad n = 1,2,\cdots \tag{2.3}$$

This is satisfied by Poisson ($a = 0, b = \lambda$), binomial ($a = -{}^{p}\!/_{q}, b = {}^{(n+1)p}\!/_{q}$) and negative binomial ($a = q, b = (k-1)q$).

Assume $\{f_k\}$, a, b and p_0 are known. We have that $S = X_1 + \cdots + X_N$, which can take values in $\left\{0,1,2,\cdots\right\}$, since the claim sizes take integer values.

Let $g_k = \mathbb{P}(S = k)$. We have that

1. $g_0 = p_0$
2. $g_r = \sum_{j=1}^{r} \left(a + \frac{bj}{r}\right) f_j g_{r-j}$ (2.4)

Proof: It is obvious that $g_0 = p_0$, because since the claim sizes cannot be 0, $\mathbb{P}(S = 0) = \mathbb{P}(N = 0) = p_0$.

Now, mutiply (2.3) by z^n and sum, to get

$$\sum_{n=1}^{\infty} p_n z^n = \sum_{n=1}^{\infty} z^n \left(a + \frac{b}{n}\right) p_{n-1}$$
$$\sum_{n=0}^{\infty} p_n z^n - p_0 = \sum_{n=1}^{\infty} azz^{n-1} p_{n-1} + b \sum_{n=1}^{\infty} \frac{z^n}{n} p_{n-1}$$
$$G_N(z) - p_0 = az \sum_{n=0}^{\infty} z^n p_n + b \sum_{n=1}^{\infty} \frac{z^n}{n} p_{n-1}$$
$$\left(1 - az\right) G_N(z) = p_0 + b \sum_{n=1}^{\infty} \frac{z^n}{n} p_{n-1}$$

Differentiating with respect to z

$$-aG_N(z) + \left(1 - az\right) G_N'(z) = bG_N(z)$$
$$G_N'(z) = \frac{a+b}{1-az} G_N(z) \qquad (2.5)$$

Now, let

$$G_S(z) = \sum_{n=0}^{\infty} g_n z^n$$

We have $M_S(u) = G_N\left(M_X(u)\right)$, and we also know that $G(z) = M\left(\log z\right)$, so

$$G_S(z) = M_S\left(\log z\right) = G_N\left(M_X(\log z)\right) = G_N\left(G_X(z)\right)$$

Differentiating, we get

$$\begin{aligned} G_S'(z) &= G_N'\left(G_X(z)\right) G_X'(z) \\ &= \frac{a+b}{1 - aG_X(z)} G_N\left(G_X(z)\right) G_X'(z) \\ &= \frac{a+b}{1 - aG_X(z)} G_S(z) G_X'(z) \end{aligned}$$

So

$$\left(1 - aG_X(z)\right) G_S'(z) = (a+b) G_S(z) G_X'(z)$$

We now feed in the fact that [note: the second sum goes from 1 instead of 0 because $f_0 = 0$]

$$G_S(z) = \sum_{n=0}^{\infty} g_n z^n \qquad G_X(z) = \sum_{n=1}^{\infty} f_k z^k$$

And get

$$\left(1 - a\sum_{\alpha=1}^{\infty} f_\alpha z^\alpha\right)\left(\sum_{\beta=1}^{\infty} \beta g_\beta z^{\beta-1}\right) = (a+b)\left(\sum_{\alpha=0}^{\infty} g_\alpha z^\alpha\right)\left(\sum_{\beta=1}^{\infty} \beta f_\beta z^{\beta-1}\right)$$

Now, equate coefficients of z^{r-1}

$$rg_r - a\sum_{\alpha+\beta=r} \beta f_\alpha g_\beta = (a+b)\sum_{\alpha+\beta=r} \beta g_\alpha f_\beta$$

$$rg_r - a\sum_{\alpha=1}^{r-1}(r-\alpha) f_\alpha g_{r-\alpha} = (a+b)\sum_{\beta=1}^{r} \beta f_\beta g_{r-\beta}$$

And so

$$\begin{aligned} rg_r &= \sum_{\beta=1}^{r}(a\beta + b\beta) f_\beta g_{r-\beta} + \sum_{\alpha=1}^{r-1}(ar - a\alpha) f_\alpha g_{r-\alpha} \\ &= \sum_{\beta=1}^{r-1}(ar + b\beta) f_\beta g_{r-\beta} + (ar+br) f_r g_0 \\ &= \sum_{\beta=1}^{r}(ar + b\beta) f_\beta g_{r-\beta} \end{aligned}$$

Which means that

$$g_r = \sum_{j=1}^{r}\left(a + \frac{bj}{r}\right) f_j g_{r-j}$$

As required. ■

To use this method with continuous claim distributions for X, we must approximate X by a discrete distribution. One way to do this is

$$f_k = \mathbb{P}\Big(X \in \big([k-\tfrac{1}{2}]h, [k+\tfrac{1}{2}]h\big)\Big)$$

for small h and $k = 0, 1, 2, \ldots$

Approximations to Compound Distributions

Some simple approximations to the distribution of S can be obtained using only a few moments of N and X_1:

- **Normal approximation**: Assume $\mathbb{E}(S^2) < \infty$ and let $\mu_S = \mathbb{E}(S)$ and $\sigma_S^2 = \text{var}(S)$. We can then approximate the distribution of S as $N\left(\mu_S, \sigma_S^2\right)$. This is a quick an easy approximation, with two major drawbacks:
  - S is always positive, whereas a normal distribution can take negative values.
  - S is often skewed, whereas a normal distribution is symmetric.
- **Translated gamma approximation**: Assume $\mathbb{E}(S^2) < \infty$, and let the coefficient of skewness of S be $\beta_S = \mathbb{E}\left((S-\mu_S)^3\right) / \sigma_S^3$ (note that β_S is non-standard notation). We can then approximate the distribution of S as that of $Y + k$, where k is a constant and $Y \sim \Gamma(\alpha, \delta)$, where k, α

and δ are chosen such that $k + Y$ has mean, variance and coefficient of skewness equal to that of S. This distribution can also be negative, but less often than the normal.

Many other approximations exist, some based only on a few moments (eg: normal power, Edgeworth expansions) and some based on the Laplace transform of the moment generating function (eg: Esscher transforms, saddlepoint approximations). See Daykin et. al. for details.

Reinsurance

An insurance company may be able to take out insurance, against last claims, for example. The **direct insurer** cedes part of the risk to a **re-insurer**, and pays a premium to do that.

Proportional reinsurance

A common example is **quota share**. The direct insurer pays a fixed portion $\alpha \in [0,1]$ of each claim (irrespective of its size) and the re-insurer pays the rest.

For a claim X, the direct insurer pays $Y = \alpha X$ and the insurer pays $Z = \left(1-\alpha\right)X$. The total claim amount paid by the direct insurer in a fixed period is

$$\tilde{S} = \sum_{i=1}^{N} aX_i = \alpha S$$

Non-proportional reinsurance

A common example is **excess loss (XoL)**. For a claim X, the direct insurer pays $Y = \min\left(X, M\right)$ and the re-insurer pays $Z = \max\left(0, X - M\right) = \left(X - M\right)_+$. M is called the **retention limit**.

- Clearly, $\mathbb{E}(Y) \leq \mathbb{E}(X)$. Furthermore, if X has density f_X, then

$$\begin{aligned} \mathbb{E}(Y) &= \int_0^M x f_X(x)\ \mathrm{d}x + M\mathbb{P}(X > M) \\ &= \left(\int_0^\infty x f_X(x)\ \mathrm{d}x - \int_M^\infty x f_X(x)\ \mathrm{d}x\right) + M\int_M^\infty f_X(x)\ \mathrm{d}x \\ &= \int_0^\infty x f_X(x)\ \mathrm{d}x - \int_M^\infty \left(x - M\right) f_X(x)\ \mathrm{d}x \end{aligned}$$

  and so

$$\mathbb{E}(X) - \mathbb{E}(Y) = \int_M^\infty \left(x - M\right) f_X(x)\ \mathrm{d}x = \int_0^\infty \mu f_X\left(\mu + M\right)\ \mathrm{d}\mu$$

- We now consider the effect on the total claim amount. Let

$$S_I = \sum_{i=1}^{N} Y_i$$

  If $N \sim \mathrm{Po}(\lambda)$ then S_I is a compound Poisson, and since $\mathbb{E}(Y_i) < \mathbb{E}(X_i)$

$$\begin{aligned} \lambda\mathbb{E}\left(Y_1\right) &\leq \lambda\mathbb{E}\left(X_1\right) \\ \mathbb{E}\left(S_I\right) &\leq \mathbb{E}\left(S\right) \end{aligned}$$

- Similarly, let

$$S_R = \sum_{i=1}^{N} Z_i$$

This is also compound Poisson $\left(\lambda\right)$. However, $\mathbb{P}\left(Z_i = 0\right) = $, $\mathbb{P}\left(X_1 \leq M\right) = F_X\left(M\right)$, so if $F_X\left(M\right) > 0$, then there is a positive probability that $Z_i = 0$, and in practice, the re-insurer only "sees" the non-zero Z_i. Suppose there are $\tilde{N}$ of these, which we'll call $w_1, \cdots, w_{\tilde{N}}$, then $S_R = \sum_{j=1}^{\tilde{N}} w_j$.

Now, note that

$$\omega_1 \sim Z_1 \mid Z_1 > 0 \sim X_1 - M \mid X_1 > M$$

and so

$$f_W(w) = \frac{f_X\left(w + M\right)}{1 - F_X(M)} \tag{3.1}$$

Finally, we note that $\tilde{N} = \sum_{i=1}^{N} \mathcal{I}_{X_i > M}$. It is a random sum and therefore has a compound distribution. We note that

$$\tilde{N} \mid N = n \sim \mathrm{Bin}\left(n, \mathbb{P}(X > M)\right)$$

and that, writing $\mathbb{P}(X > m) = p$, $\tilde{N}$ has probability generating function

$$\begin{aligned} \mathcal{G}_{\tilde{N}}(z) = \mathbb{E}\left(z^{\tilde{N}}\right) &= \mathbb{E}\left[\mathbb{E}\left(z^{\tilde{N}} \mid N\right)\right] \\ &= \mathbb{E}\left[(pz + q)^N\right] \\ &= \mathcal{G}_N\left(pz + q\right) \end{aligned}$$

<u>EXAMPLE</u>: if $N \sim \mathrm{Po}(\lambda)$, then

$$\begin{aligned} \mathcal{G}_{\tilde{N}}(z) &= \mathcal{G}_N\left(pz + q\right) \\ &= \exp\left\{\lambda(pz + q - 1)\right\} \\ &= \exp\left\{p\lambda(z - 1)\right\} \end{aligned}$$

and so $\tilde{N} \sim \mathrm{Po}(p\lambda)$. □

- In practice, **limited excessive loss** reinsurance is more common, in which

$$Z = \begin{cases} 0 & X \leq M \\ X - M & X \in (M, A + M] \\ A & X > A + M \end{cases}$$

In many insurance policies, the insured has to pay the first part of any claim up to an amount of **deductible** (or **excess**), say L. The insurer therefore pays $\left(X-L\right)_+$, and the calculation is similar as for excess loss insurance.

Example

Here's an example comparing quota share reinsurance and XoL reinsurance, in which X is exponentially distributed with mean and standard deviation 10:

	Insurer (Y)		Reinsurer (Z)	
	Mean	SD	Mean	SD
No reinsurance	10	10	0	0
Quota share ($\alpha = 3/4$)	7.5	7.5	2.5	2.5
XoL (M chosen such that the direct insurer's mean payment is the same as in quota share. This gives $M \approx 13.86$)	7.5	4.94	2.5	6.61

Clearly, the reinsurer takes up more of the risk in XoL. XoL reinsurance is therefore more expensive.

The figures in the last row are found as follows:

- <u>Finding M</u>

  From above, we have that

$$\mathbb{E}(X) - \mathbb{E}(Y) = \int_0^\infty \mu f_X\left(\mu + M\right) \, \mathrm{d}\mu$$

  in this case, $\mathbb{E}(X) = \frac{1}{\lambda}$, and $f_X = \lambda e^{-\lambda x}$, so

$$\begin{aligned}
\mathbb{E}(Y) &= \tfrac{1}{\lambda} - \lambda \int_0^\infty \mu e^{-\lambda(\mu+M)} \, \mathrm{d}\mu \\
&= \tfrac{1}{\lambda} - \lambda \left\{ \left[-\tfrac{1}{\lambda} \mu e^{-\lambda(\mu+M)} \right]_0^\infty + \tfrac{1}{\lambda} \int_0^\infty e^{-\lambda(\mu+M)} \, \mathrm{d}\mu \right\} \\
&= \tfrac{1}{\lambda} - \int_0^\infty e^{-\lambda(\mu+M)} \, \mathrm{d}\mu \\
&= \tfrac{1}{\lambda}\left(1 - e^{-\lambda M}\right)
\end{aligned}$$

  We require this to be equal to 7.5 (the mean in quota share), so

$$\begin{aligned}
\tfrac{1}{\lambda}\left(1 - e^{-\lambda M}\right) &= 7.5 \\
M &= -\tfrac{1}{\lambda} \ln\left(1 - 7.5\lambda\right)
\end{aligned}$$

  in this case, $\mathbb{E}(X) = 10$, and so $\lambda = 0.1$. So:

$$M = 13.86$$

as advertised.

- Finding SD(Y)

Consider

$$\begin{aligned}\mathbb{E}(Y^2) &= \int_0^M x^2 f_X(x)\ \mathrm{d}x + M^2\mathbb{P}(X > M)\\ &= \left(\int_0^\infty x^2 f_X(x)\ \mathrm{d}x - \int_M^\infty x^2 f_X(x)\ \mathrm{d}x\right) + M^2\int_M^\infty f_X(x)\ \mathrm{d}x\\ &= \int_0^\infty x^2 f_X(x)\ \mathrm{d}x - \int_M^\infty \left(x^2 - M^2\right) f_X(x)\ \mathrm{d}x\end{aligned}$$

And so

$$\mathbb{E}(X^2) - \mathbb{E}(Y^2) = \int_M^\infty \left(x^2 - M^2\right) f_X(x)\ \mathrm{d}x$$

In our case, $\mathbb{E}(X^2) = \mathrm{var}(X) + \mathbb{E}(X)^2 = 2 / \lambda^2$ and f_X is as above, so

$$\begin{aligned}\mathbb{E}(Y^2) &= \tfrac{2}{\lambda^2} - \lambda\int_M^\infty \left(x^2 - M^2\right)e^{-\lambda x}\ \mathrm{d}x\\ &= \tfrac{2}{\lambda^2} - \lambda\left\{\left[-\tfrac{1}{\lambda}(x^2 - M^2)e^{-\lambda x}\right]_M^\infty + \tfrac{1}{\lambda}\int_M^\infty 2xe^{-\lambda x}\ \mathrm{d}x\right\}\\ &= \tfrac{2}{\lambda^2} - \int_M^\infty 2xe^{-\lambda x}\ \mathrm{d}x\\ &= \tfrac{2}{\lambda^2} - \left\{\left[-\tfrac{1}{\lambda}2xe^{-\lambda x}\right]_M^\infty + \tfrac{1}{\lambda}\int_M^\infty 2e^{-\lambda x}\ \mathrm{d}x\right\}\\ &= \tfrac{2}{\lambda^2} - \left\{\tfrac{1}{\lambda}2Me^{-\lambda M} + 2\tfrac{1}{\lambda^2}e^{-\lambda M}\right\}\\ &= 2\left[\tfrac{1}{\lambda^2} - \tfrac{1}{\lambda}Me^{-\lambda M} - \tfrac{1}{\lambda^2}e^{-\lambda M}\right]\end{aligned}$$

Furthermore,

$$\mathrm{SD}(Y) = \sqrt{\mathbb{E}(Y^2) - \mathbb{E}(Y)^2}$$

In our case, we know $M = 13.86$, $\lambda = 0.1$ and $\mathbb{E}(Y) = 7.5$. Feeding in numbers to all the above, we obtain

$$\mathrm{SD}(Y) = 4.94$$

As advertised.

- Finding $\mathbb{E}(Z)$

$$\begin{aligned}\mathbb{E}(Z) &= \int_M^\infty (x - M) f_X(x)\ \mathrm{d}x\\ &= \lambda\int_M^\infty (x - M)e^{-\lambda x}\ \mathrm{d}x\\ &= \lambda\left\{\left[-\tfrac{1}{\lambda}(x - M)e^{-\lambda x}\right]_M^\infty + \tfrac{1}{\lambda}\int_M^\infty e^{-\lambda x}\ \mathrm{d}x\right\}\\ &= \tfrac{1}{\lambda}e^{-\lambda M}\end{aligned}$$

From above, we know

$$e^{-\lambda M} = 1 - 7.5\lambda$$

and so

$$\mathbb{E}(Z) = \tfrac{1}{\lambda} - 7.5 = 2.5$$

As advertised.

- <u>Finding SD(Z)</u>

  Consider

$$\begin{aligned}
\mathbb{E}(Z^2) &= \int_M^\infty (x-M)^2 f_X(x)\ \mathrm{d}x \\
&= \lambda \int_M^\infty (x-M)^2 e^{-\lambda x}\ \mathrm{d}x \\
&= \lambda \left\{ \left[-\tfrac{1}{\lambda}(x-M)^2 e^{-\lambda x} \right]_M^\infty + 2\tfrac{1}{\lambda}\int_M^\infty (x-M)e^{-\lambda x}\ \mathrm{d}x \right\} \\
&= 2\int_M^\infty (x-M)e^{-\lambda x}\ \mathrm{d}x \\
&= 2\int_0^\infty \mu e^{-\lambda(\mu+M)}\ \mathrm{d}\mu
\end{aligned}$$

  Thankfully, this is an integral we've already worked out when finding M above, and we get

$$\mathbb{E}(Z^2) = \tfrac{2}{\lambda^2} e^{-\lambda M}$$

  Finally, we know $\mathbb{E}(Z) = 2.5$, and so feeding numbers in,

$$\mathrm{SD}(Z) = \sqrt{\mathbb{E}(Z^2) - \mathbb{E}(Z)^2} = 6.61$$

  Unsurprisingly, as advertised. □

Ruin Probabilities

Suppose $X_1, X_2, \cdots$ are IID with distribution function F, and $\mathbb{E}\left(X_1\right) < \infty$. Let $N(t)$ be the number of claims arriving in $(0, t]$, independent of the X_i. Let $S(t) = \sum_{i=1}^{N(t)} X_i$ be the total claim amount in $(0, t]$ (with $S(t) = 0$ if $N(t) = 0$).

In the **classical risk model**, the X_i are positive random variables, and $\left\{N(t), t \geq 0\right\}$ is a Poisson process with rate $\lambda > 0$ (which means that (a) $N(t) \sim \text{Po}(\lambda t)$ and (b) the times between consecutive arrivals are IID exponential variables with mean $1/\lambda$).

$\left\{S(t), t \geq 0\right\}$ is then a compound Poisson process (in other words, for every t, $S(t)$ has a compound Poisson distribution). Using (2.1) and (2.2), we have

$$\begin{aligned}
\mathbb{E}\left[S(t)\right] &= \lambda\mu t \\
\text{var}\left[S(t)\right] &= \lambda t \,\text{var}\left(X_1\right) + \lambda t \mu^2 \\
&= \lambda t\left(\text{var}\, X_1 + \mu^2\right) \\
&= \lambda t \mathbb{E}\left(X_1^2\right) \\
M_{S(t)}(u) &= \exp\left\{\lambda t\left(M_X(u) - 1\right)\right\} \\
\mathbb{P}\left(S(t) \leq x\right) &= \sum_{n=0}^{\infty} \frac{e^{-\lambda t}\left(\lambda t\right)^n}{n!} F^{*n}(x)
\end{aligned}$$

We further suppose that premium income is received continuously at a constant rate $c > 0$. Suppose that at $t = 0$, the insurance company has capital $u \geq 0$.

The **surplus** or **risk reserve** at time t is then

$$U(t) = u + ct - \sum_{i=1}^{N(t)} X_i \tag{4.1}$$

We call $\left\{U(t) : t \geq 0\right\}$ the **risk reserve process** or **surplus process**.

This classical risk model involves a number of simplifications. For example:

- The claims are all paid out immediately
- No interest is earned on the surpluss
- λ remains constant (unlikely if there are seasonal variations in accident rates, for example)
- c is continuous in time, and constant

Miscellaneous definitions

- **Safety loading**: From (4.1), note that the expected profit in unit time (0, t] is

$$\frac{\mathbb{E}\left[U(t)\right]-u}{t}=\frac{ct-\lambda\mu t}{t}=c-\lambda\mu$$

The **net profit condition** is then

$$c>\lambda\mu \qquad (4.2)$$

We put $c=\left(1+\rho\right)\lambda\mu$, so that $\rho>0$ if the net profit condition is satisfied. ρ is called the **(relative) safety loading** (or **premium loading factor**). However, random fluctuations in $U(t)$ mean that the company could still face ruin.

- **Probability of ruin**: if $U(t) < 0$ for some $t > 0$, then **ruin** is said to occur. The **probability of ruin** given initial capital $u \geq 0$ is

$$\begin{aligned}\psi(u) &= \mathbb{P}\left(U(t)<0 \text{ for some } t\geq 0\right)\\ &= \mathbb{P}\left(\text{Ruin ever occuring}\right)\end{aligned}$$

this is also know as the **probability of ultimate ruin** or the **infinite time horizon ruin probability**.

Other quantities of interest include the time to ruin, and the deficit at ruin, $\left|U(\psi(u))\right|$.

- **Finite-time ruin probability**: given an initial capital $u \geq 0$, the **finite-time ruin probability is**

$$\psi(u,T)=\mathbb{P}\left(U(t)<0 \text{ for some } t \text{ in } [0,T]\right)$$

- **Constraint on ruin probabilities**: if $0\leq u_1\leq u_2$ and $0\leq T_1\leq T_2\leq\infty$, then

$$\begin{gathered}\psi(u_1)\leq\psi(u_2)\\ \psi(u_1,T)\leq\psi(u_2,T)\\ \psi(u,T_1)\leq\psi(u,T_2)\leq\psi(u) \qquad \forall u\geq 0\end{gathered}$$

Furthermore,

$$\psi(u,T)\xrightarrow{T\to\infty}\psi(u)$$

- **Discrete-time ruin probabilities**: the classical risk model assumes we check for ruin continuously in time. However in practice, it may be only possible to observe $U(nh)$, where $n = 0,1,2,\cdots$, and we might miss a time at which ruin occurs. In this course, however, we always use a continuous time model.

$\psi(u)$ is hard to calculate explicitly. We therefore develop bounds, and latter approximations, for this quantity.

The Lundberg Inequality

We will first need to state a condition on the moment generating function. Recall that if X is positive, the MGF $M_X(r)$ exists in $\left(-\infty,\gamma\right)$, where $\gamma \in [0,\infty]$, and if $\gamma < \infty$, the MGF may or may not exist at $r = \gamma$. Now consider the following condition:

Condition C: Assume there exists r_∞, $0 < r_\infty \le \infty$, such that $M_X(r) \uparrow \infty$ as $r \uparrow r_\infty$.

Remark: To get an intuitive understanding of this condition, consider that, some fixed $r > 0$ and any $x > 0$. Then:

$$e^{rx}\mathbb{I}_{\{X_1 > x\}} \le e^{rX_1}$$

now take expectations

$$e^{rx}\mathbb{P}\left(X_1 > x\right) \le M_X(r)$$

now, condition C implies that there *must* be some finite r for which M is finite. Let the r we chose above be such an r, and let $M_X(r) = k$. We then have

$$\mathbb{P}\left(X_1 > x\right) \le ke^{-rx}$$

Intuitively, this stipulates that the tails of X be small enough. So any X that satisfies condition C must have $1 - F_X(x)$ decreasing at least exponentially fast.

<u>EXAMPLE</u>: (1) if X_1 has density $f_x(x) = \theta / x^{1+\theta}$ for $x \geq 1$, then

$$\mathbb{P}(X_1 > x) = 1 / x^\theta$$

which clearly does not decrease at least exponentially. Thus, this distribution does not satisfy condition C. (This is to be expected; the distribution is Pareto, which has very heavy tails).

(2) if $X_1 \sim \exp(1/\mu)$, then

$$M_X(r) = \frac{1}{1-\mu r} \qquad r < 1/\mu$$

This satisfies C with $r_\infty = 1/\mu$. □

We're now ready to derive our bound:

Theorem 4.2 (Lundberg inequality): Under positive safety loading and condition C in the classical risk model, we have

$$\psi(u) \le e^{-Ru} \qquad \forall u \ge 0$$

Where R called the **adjustment coefficient** or **Lundberg exponent** is the unique positive solution of

$$M_X(R) - 1 = \frac{cR}{\lambda} \tag{4.3}$$

This equation can also be written as

$$M_X(R) - 1 = \left(1+\rho\right)\mu R$$

Proof: The structure of the proof is as follows:

1. Prove that (4.3) has a unique solution.
2. Define a new function $\psi_n(u)$ and show that $\psi_n(u) \le e^{-Ru} \Leftrightarrow \psi(u) \le e^{-Ru}$.
3. Show that $\psi_n(u) \le e^{-Ru}$.

Step 1

We first show (4.3) has a unique solution. Let

$$g(r) = M_X(r) - 1 - \frac{cr}{\lambda}$$

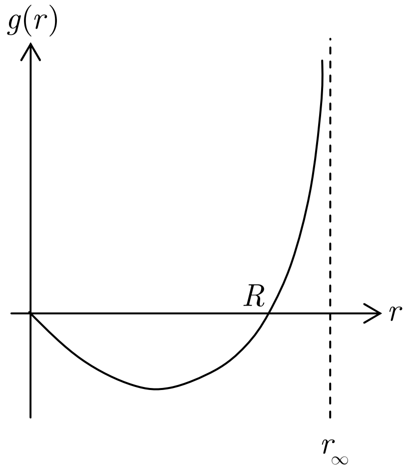
We want a solution for $g(r) = 0$.

Case 1; $r_\infty < \infty$: We know that

- $g(0) = 0$, because $M_X(0) = 1$.

- $g(r)$ is continuous for $r < r_\infty$, because M is continuous over that region (a property of Laplace transforms).
- $g'(r) = M_X'(r) - \frac{c}{\lambda}$, and
  $$g'(0) = M_X'(0) - \tfrac{c}{\lambda} = \mu - \tfrac{c}{\lambda} < 0$$
  because since we have positive safety loading, $c > \lambda\mu$.
- $g''(r) = M_X''(r) > 0$, because M is convex (a property of Laplace transforms).
- Condition C holds, and so g tends to infinity as r approaches r_∞.

Together, the above statements imply that the graph of g looks something like this:



There is clearly a unique strictly positive solution for $g(r) = 0$.

Case 2; $r_\infty = \infty$: in that case, the argument above no longer works, because its unclear whether M or r will tend to infinity faster. If M tends to infinity faster, then the graph will look as above and all is good. If r tends to infinity faster, then things will be very different.

We observe that since $X_1 > 0$, there exists some $\eta > 0$ such that $\mathbb{P}(X_1 > \eta) = p > 0$. Then, for $r > 0$,

$$\begin{aligned} M_X(r) &= \mathbb{E}\left(e^{rX_1}\right) \\ &= \mathbb{E}\left(e^{rX_1} \mid X_1 > \eta\right)p \\ &\qquad + \mathbb{E}\left(e^{rX_1} \mid X_1 \le \eta\right)\left(1-p\right) \\ &\ge p\mathbb{E}\left(e^{rX_1} \mid X_1 > \eta\right) \\ &\ge pe^{r\eta} \end{aligned}$$

This implies that

$$g(r) \ge pe^{r\eta} - 1 - \frac{cr}{\lambda} \xrightarrow{r\to\infty} \infty$$

(Because the exponential term "beats" the linear term). Thus, all is well and we have a unique solution to $g(r) = 0$ in all cases.

<u>STEP 1</u>

We note that ruin can only occur at the time of a claim. Let

$$\psi_n(u) = \mathbb{P}(\text{Ruin occurs before } n\text{th claim})$$

and note that $\psi_n(u) \uparrow \psi(u)$ as $n \to \infty$, and so

$$\psi(u) \le e^{-Ru} \Leftrightarrow \psi_n(u) \le e^{-Ru} \ \forall n$$

<u>STEP 3</u>

We now prove that $\psi_n(u) \le e^{-Ru} \quad \forall n$ by induction on n. For convenience, assume X_1 has density f_X (the proof goes through in the general case).

Basic case ($n = 1$): Ruin cannot occur *before* the first claim. Thus

$$\begin{aligned} \psi_1(u) &= \mathbb{P}(\text{Ruin occurs on or before 1st claim}) \\ &= \mathbb{P}(\text{Ruin occurs } at \ 1st \text{ claim}) \\ &= \int_0^\infty \mathbb{P}(\text{Ruin occurs at } 1st \text{ claim} \\ &\qquad\qquad | 1st \text{ claim occurs at } t)\lambda e^{-\lambda t}\, \mathrm{d}t \end{aligned}$$

We note that if the first claim occurs at time t, the total money held at that time is $u + ct$. Thus, the amount of the first claim must exceed that amount for ruin to occur then:

$$= \int_{t=0}^{\infty} \lambda e^{-\lambda t} \int_{x=u+ct}^{\infty} f_X(x)\ \mathrm{d}x\ \mathrm{d}t$$

We also note that for $x \geq u + ct$, $e^{-R\left(u+ct-x\right)} \geq 1$, so

$$\leq \int_{t=0}^{\infty} \lambda e^{-\lambda t} \int_{x=u+ct}^{\infty} e^{-R(u+ct-x)} f_X(x)\ \mathrm{d}x\ \mathrm{d}t$$

Note that the second integrand is always positive, so

$$\begin{aligned}
&\leq \int_{t=0}^{\infty} \lambda e^{-\lambda t} \int_{x=0}^{\infty} e^{-R(u+ct-x)} f_X(x)\ \mathrm{d}x\ \mathrm{d}t \\
&= e^{-Ru} \int_{t=0}^{\infty} \lambda e^{-(\lambda+Rc)t} \int_{x=0}^{\infty} e^{Rx} f_X(x)\ \mathrm{d}x\ \mathrm{d}t \\
&= e^{-Ru} \int_{t=0}^{\infty} \lambda e^{-(\lambda+Rc)t} \mathbb{E}\left(e^{RX}\right)\ \mathrm{d}t \\
&= e^{-Ru} \int_{t=0}^{\infty} \lambda e^{-(\lambda+Rc)t} M_X(R)\ \mathrm{d}t
\end{aligned}$$

Recall the definition of R is $M_X(R) = 1 + \frac{cR}{\lambda}$, so

$$\begin{aligned}
&= e^{-Ru} \int_{t=0}^{\infty} \left(\lambda + cR\right) e^{-(\lambda+Rc)t}\ \mathrm{d}t \\
&= e^{-Ru}
\end{aligned}$$

So we do indeed have

$$\psi_1(u) \leq e^{-Ru}$$

Inductive step: Assume $\psi_n(u) \leq e^{-Ru}$, and condition on the time and amount of the first claim:

$$\psi_{n+1}(u) = \int_0^{\infty} \lambda e^{-\lambda t} \mathbb{P}(\text{Ruin on or before } (n+1)\text{th claim} \mid \text{1st claim at time } t)\ \mathrm{d}t$$

We split this probability into two:

- First assuming that ruin occurs at the first claim, so that the amount of the first claim is greater than $u + ct$
- then assume that it doesn't, so that the amount of the first claim is less than $u + ct$, and we "restart from scratch" after the first claim with wealth $u + ct - x_1$

$$\psi_{n+1}(u) = \int_0^{\infty} \lambda e^{-\lambda t} \left\{ \int_{x=u+ct}^{\infty} f_X(x)\ \mathrm{d}x + \int_{x=0}^{u+ct} \psi_n(u+ct-x) f_X(x)\ \mathrm{d}x \right\}\ \mathrm{d}t$$

As before, we note that $x \geq u + ct$, $e^{-R\left(u+ct-x\right)} \geq 1$. Furthermore, by our inductive hypothesis, $\psi_n(u+ct-x) \leq e^{-R(u+ct-x)}$, and so

$$\psi_{n+1}(u) \le \int_{t=0}^{\infty} \lambda e^{-\lambda t} \int_{x=0}^{\infty} e^{-R(u+ct-x)} f_X(x) \, \mathrm{d}x \, \mathrm{d}t$$

this is identical to the expression obtained in the basic step. Thus, $\psi_n(u) \le e^{-Ru}$. This proves our theorem. ■

The Adjustment Coefficient

This section contains a number of miscellaneous points regarding the adjustment coefficient R:

- R is used as a measure of risk. A large R means a smaller bound on $\psi(u)$, and so we "like" large R.
- In certain cases, R can be found explicitly. For example, if $X_1 \sim \exp\left(1/\mu\right)$, then $R = \frac{1}{\mu} - \frac{\lambda}{c} = \frac{\rho}{\mu(1+\rho)}$. However, R often needs to be found numerically (eg: by Newton-Raphson iteration).
- We can find an upper bound for R:

$$\begin{aligned} \frac{cR}{\lambda} = M_X(R) - 1 &= \int_0^{\infty} e^{Rx} f_X(x) \, \mathrm{d}x - 1 \\ &\ge \int_0^{\infty} \left(1 + Rx + \tfrac{1}{2} R^2 x^2\right) f_X(x) \, \mathrm{d}x - 1 \\ &= R\mu + \tfrac{1}{2} R^2 \mathbb{E}\left(X_1^2\right) \end{aligned}$$

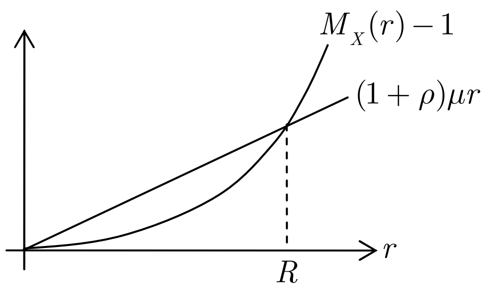
  this implies that

$$\frac{cR}{\lambda} \ge R\mu + \tfrac{1}{2} R^2 \mathbb{E}\left(X_1^2\right)$$

  Finding the critical points of this quadratic inequality and subbing in a few values confirms that

$$R \le \frac{2\mu}{\mathbb{E}\left(X_1^2\right)} \rho$$

- R satisfies $M_X(R) - 1 = \left(1+\rho\right)\mu R$. We know $M_X(R) - 1$ is convex with positive gradient μ at the origin, so R is:



Clearly, increasing ρ increases the gradient of the line and therefore increases R. This makes sense; the more safety loading, the less risk.

- We can express the defining equation for R in a different form using integration by parts:

$$M_X(R) - 1 = \frac{cR}{\lambda}$$
$$\int_0^\infty e^{Rx} f_X(x)\ \mathrm{d}x - 1 = \frac{cR}{\lambda}$$
$$\int_0^\infty e^{Rx} \frac{\mathrm{d}}{\mathrm{d}x}\left[-(1 - F_X(x))\right]\ \mathrm{d}x - 1 = \frac{cR}{\lambda}$$
$$\int_0^\infty e^{Rx}\left(1 - F_X(x)\right)\ \mathrm{d}x = \frac{c}{\lambda}$$

- In practice, λ and the distribution of X_1 are not known; they need to be estimated from data, using statistical techniques.

<u>EXAMPLE</u>: We now consider the effect of XoL reinsurance with retention limit M on R in a classical risk model with positive safety loading ($c > \lambda\mu$). Recall that $\rho = (c - \lambda\mu) / \lambda\mu$ and $c = \left(1 + \rho\right)\lambda\mu$

- Claims arrive at a poisson rate λ. The direct insurer pays $Y = \min\left\{X, M\right\}$ and the re-insurer pays $Z = \max\left\{0, X - M\right\}$.
- The re-insurer can expect to pay out $\lambda\mathbb{E}(Z)$ per unit time, so we expect the re-insurer to charge $\left(1 + \xi\right)\lambda\mathbb{E}(Z)$ per unit time, were ξ is the **premium loading factor** for the re-insurer.

  Taking this into account, the direct insurer's "premium income" per unit time is

$$c^* = \left(1 + \rho\right)\lambda\mu - \left(1 + \xi\right)\lambda\mathbb{E}(Z)$$

  If $M = 0$ and all the risk is passed to the re-insurer, $\mathbb{E}(Z) = \mathbb{E}(X) = \mu$ and $c^\star = (\rho - \xi)\lambda\mu$. To ensure the direct insurer does not make a steady profit without taking any risk, we insist that

$$\rho \leq \xi$$

  We also assume the re-insured process is safety-loaded, so that

$$c^\star > \lambda\mathbb{E}(Y_1)$$

- Condition C is satisfied with $r_\infty = \infty$. Intuitively, this is because condition C is a "short tail" condition, and re-insurance ensures the tail is short. This can be formally verified using the monotone convergence theorem.
- The direct insurer's adjustment coefficient $R^\star$ therefore satisfies

$$M_Y(R^\star) - 1 = \frac{c^\star R^\star}{\lambda}$$

ie:

$$\int_0^M e^{R^\star x} f_X(x)\ \mathrm{d}x + e^{MR^\star}\left(1 - F_X(M)\right) - 1 = \frac{c^\star R^\star}{\lambda}$$

for given f_X, ξ, ρ and λ, we can solve this equation numerically. □

The Cramér-Lundberg Approximation

We now obtain an approximation for $\psi(u)$ when u is large. Let

$$\phi(u) = 1 - \psi(u) = \mathbb{P}(\text{Never ruined})$$

this is also known as the **survival probability**.

Lemma 4.3: In a classical risk model with positive safety loading, we have

$$\phi(t) = \phi(0) + \frac{\lambda}{c}\int_0^t \phi(t-x)\left(1 - F_X(x)\right)\ \mathrm{d}x \quad (4.4)$$

where $\phi(0) = 1 - \frac{\lambda\mu}{c} = \frac{\rho}{1+\rho}$

Proof: We consider a case in which X_1 has a density, though the proof generalises.

We begin by conditioning on the time T_1 and size X_1 of the first claim:

$$\begin{aligned}\phi(u) &= \mathbb{P}\left(U(t) \geq 0 \quad \forall t\right) \\ &= \int_{s=0}^{\infty}\int_{x=0}^{\infty} \mathbb{P}\left(U(t) \geq 0 \ \ \forall t \mid X_1 = x_1, T_1 = s\right) \\ &\qquad f_X(x)\ \mathrm{d}x\ \lambda \mathrm{e}^{-\lambda s}\ \mathrm{d}s\end{aligned}$$

We note, however, that

- If the first claim is greater than $u + cs$, we're immediately ruined.

- Once the first claim has occurred, we effectively "re-start the clock" with capital $u + cs - x$.

The above therefore becomes:

$$\phi(u) = \int_{s=0}^{\infty} \int_{x=0}^{u+cs} \phi(u + cs - x) f_X(x) \, \mathrm{d}x \; \lambda e^{-\lambda s} \, \mathrm{d}s$$

We substitute $z = u + cs$ into the outer integral to get:

$$\begin{aligned}
\phi(u) &= \int_{z=u}^{\infty} \frac{\lambda}{c} e^{-\lambda\left(\frac{z-u}{c}\right)} \int_{x=0}^{z} \phi(z - x) f_X(x) \, \mathrm{d}x \, \mathrm{d}z \\
&= \frac{\lambda}{c} e^{\frac{\lambda u}{c}} \int_{z=u}^{\infty} e^{-\frac{\lambda z}{c}} \int_{0}^{z} \phi(z - x) f_X(x) \, \mathrm{d}x \, \mathrm{d}z
\end{aligned}$$

We now differentiate ϕ with respect to u (sigh)

$$\phi'(u) = \frac{\lambda}{c}\phi(u) - \frac{\lambda}{c} e^{\lambda u/c} e^{-\lambda u/c} \int_{x=0}^{u} \phi(u - x) f_X(x) \, \mathrm{d}x$$

$$\boxed{\phi'(u) = \frac{\lambda}{c}\phi(u) - \frac{\lambda}{c} \int_{x=0}^{u} \phi(u - x) f_X(x) \, \mathrm{d}x} \qquad (4.5)$$

This is an integro-differential equation for ϕ. We now integrate this from 0 to t

$$\begin{aligned}
\phi(t) = \phi(0) + \frac{\lambda}{c} \int_0^t \phi(u) \, \mathrm{d}u \\
- \frac{\lambda}{c} \left[\int_{u=0}^{t} \left\{ \int_{x=0}^{u} \phi(u - x) f_X(x) \, \mathrm{d}x \right\} \mathrm{d}u \right]
\end{aligned} \qquad (*)$$

Consider the integral in curly braces separately and integrate it by parts using $f_X(x) = \frac{\mathrm{d}}{\mathrm{d}x}\left(-(1 - F_X(x)\right)$

$$\begin{aligned}
\{ \quad \} &= \left[-\phi(u - x)(1 - F_X(x)\right]_0^u \\
&\quad - \int_0^u \phi'(u - x)\left(1 - F_X(x)\right) \, \mathrm{d}x \\
&= -\phi(0)\left(1 - F_X(u)\right) + \phi(u)\left(1 - F_X(0)\right) \\
&\quad - \int_0^u \phi'(u - x)\left(1 - F_X(x)\right) \, \mathrm{d}x
\end{aligned}$$

Since X is strictly positive, $F_X(0) = 0$ and so

$$\begin{aligned}
&= -\phi(0)\left(1 - F_X(u)\right) + \phi(u) \\
&\quad - \int_0^u \phi'(u - x)\left(1 - F_X(x)\right) \, \mathrm{d}x
\end{aligned}$$

And so the integral which appears in square brackets in (*) becomes

$$\left[\quad\right] = -\phi(0)\int_{u=0}^{t}\Big(1-F_X(u)\Big)\ \mathrm{d}u + \int_{u=0}^{t}\phi(u)\ \mathrm{d}u - \int_{u=0}^{t}\int_{x=0}^{u}\phi'(u-x)\Big(1-F_X(x)\Big)\ \mathrm{d}x\ \mathrm{d}u$$

Interchanging the order of integration in the last term, we get

$$= -\phi(0)\int_{u=0}^{t}\Big(1-F_X(u)\Big)\ \mathrm{d}u + \int_{u=0}^{t}\phi(u)\ \mathrm{d}u - \int_{x=0}^{t}\Big(1-F_X(x)\Big)\int_{u=x}^{t}\phi'(u-x)\ \mathrm{d}u\ \mathrm{d}x$$

$$= \cancel{-\phi(0)\int_{u=0}^{t}\Big(1-F_X(u)\Big)\ \mathrm{d}u} + \int_{u=0}^{t}\phi(u)\ \mathrm{d}u - \int_{x=0}^{t}\Big(1-F_X(x)\Big)\Big\{\phi(t-x) - \cancel{\phi(0)}\Big\}\ \mathrm{d}x$$

Clearly, the indicated terms cancel

$$= -\phi(0)\int_{u=0}^{t}\Big(1-F_X(u)\Big)\ \mathrm{d}u + \int_{u=0}^{t}\phi(u)\ \mathrm{d}u - \int_{x=0}^{t}\Big(1-F_X(x)\Big)\int_{u=x}^{t}\phi'(u-x)\ \mathrm{d}u\ \mathrm{d}x$$

$$= \int_{u=0}^{t}\phi(u)\ \mathrm{d}u - \int_{x=0}^{t}\Big(1-F_X(x)\Big)\phi(t-x)\ \mathrm{d}x$$

Substituting this back into (*), we get

$$\phi(t) = \phi(0) + \cancel{\frac{\lambda}{c}\int_{0}^{t}\phi(u)}\ \mathrm{d}u - \frac{\lambda}{c}\left[\cancel{\int_{u=0}^{t}\phi(u)}\ \mathrm{d}u - \int_{x=0}^{t}\Big(1-F_X(x)\Big)\phi(t-x)\ \mathrm{d}x\right]$$

Once again, the indicated terms cancel

$$\phi(t) = \phi(0) + \frac{\lambda}{c}\int_{x=0}^{t}\phi(t-x)\Big(1-F_X(x)\Big)\ \mathrm{d}x$$

which is precisely statement (4.4), which we wanted to prove.

We can find $\phi(0)$ by a slightly informal argument. Let $t\to\infty$ in (4.4). We get

$$\begin{aligned}
\phi(\infty) &= \phi(0) + \frac{\lambda}{c}\int_{x=0}^{\infty}\phi(\infty - x)\big(1 - F_X(x)\big)\ \mathrm{d}x \\
&= \phi(0) + \frac{\lambda}{c}\phi(\infty)\int_{x=0}^{\infty}\big(1 - F_X(x)\big)\ \mathrm{d}x \\
&= \phi(0) + \frac{\lambda}{c}\phi(\infty)\int_{x=0}^{\infty}\int_{y=x}^{\infty} f_X(y)\ \mathrm{d}y\ \mathrm{d}x \\
&= \phi(0) + \frac{\lambda}{c}\phi(\infty)\int_{y=0}^{\infty} f_X(y)\int_{x=0}^{y} 1\ \mathrm{d}x\ \mathrm{d}y \\
&= \phi(0) + \frac{\lambda}{c}\phi(\infty)\int_{y=0}^{\infty} y f_X(y)\ \mathrm{d}y \\
&= \phi(0) + \frac{\lambda}{c}\phi(\infty)\mathbb{E}(X_1) \\
&= \phi(0) + \frac{\lambda\mu}{c}\phi(\infty)
\end{aligned}$$

However,

$$\phi(\infty) = \mathbb{P}(\text{No ruin} \mid \text{start with } \infty \text{ capital}) = 1$$

And so

$$\boxed{\phi(0) = 1 - \frac{\lambda\mu}{c}}$$

as required. ■

Note that $\psi(0) = 1 - \phi(0) = \lambda\mu / c$

We are now ready to derive the main theorem of this section:

Theorem 4.4 (Cramér-Lundberg approximation): Assume positive safety loading and condition C in the classical risk model. Then

$$\lim_{u\to\infty} e^{Ru}\psi(u) = A$$

where

$$A = \left\{\frac{R}{p}\int_0^{\infty} x e^{Rx}\,\frac{1 - F_X(x)}{\mu}\ \mathrm{d}x\right\}^{-1}$$

And R is the adjustment coefficient.

Proof: Let

$$f_I(x) = \frac{1 - F_X(x)}{\mu}$$

Then $f_I(x) \geq 0$ is a probability density on $(0,\infty)$ because $\int_0^{\infty} f_I(x)\ \mathrm{d}x = 1$.

Now, recall that $\frac{\lambda\mu}{c} < 1$ by positive safety loading. Let's now play around with equation (4.4)

$$\begin{aligned}\phi(u) &= \phi(0) + \frac{\lambda}{c}\int_0^u \phi(t-x)\big(1-F_X(x)\big)\ \mathrm{d}x \\ &= 1 - \frac{\lambda\mu}{c} + \frac{\lambda\mu}{c}\int_0^u \phi(u-x) f_I(x)\ \mathrm{d}x\end{aligned}$$

We can now replace our survival probabilities ϕ with ruin probabilities ψ:

$$\begin{aligned}\psi(u) &= \frac{\lambda\mu}{c}\left(1 - \int_0^u \big(1-\psi(u-x)\big) f_I(x)\ \mathrm{d}x\right) \\ &= \frac{\lambda\mu}{c} - \frac{\lambda\mu}{c}\int_0^u f_I(x)\ \mathrm{d}x \\ &\qquad + \frac{\lambda\mu}{c}\int_0^u \psi(u-x) f_I(x)\ \mathrm{d}x \\ &= \frac{\lambda\mu}{c} - \frac{\lambda\mu}{c}\left(1 - \int_0^\infty f_I(x)\ \mathrm{d}x\right) \\ &\qquad + \frac{\lambda\mu}{c}\int_0^u \psi(u-x) f_I(x)\ \mathrm{d}x \\ &= \frac{\lambda\mu}{c}\int_u^\infty f_I(x)\ \mathrm{d}x + \frac{\lambda\mu}{c}\int_0^u \psi(u-x) f_I(x)\ \mathrm{d}x \qquad (4.6)\end{aligned}$$

Therefore

$$\begin{aligned}\psi(u)e^{Ru} &= \frac{\lambda\mu}{c} e^{Ru}\int_u^\infty f_I(x)\ \mathrm{d}x \\ &\qquad + \frac{\lambda\mu}{c}\int_0^u e^{R(u-x)}\psi(u-x) e^{Rx} f_I(x)\ \mathrm{d}x\end{aligned}$$

This is of the form

$$Z(u) = z(u) + \int_0^u z(u-x) g(x)\ \mathrm{d}x \qquad (*)$$

where

- $Z(u) = e^{Ru}\psi(u)$
- $z(u) = \frac{\lambda\mu}{c} e^{Ru}\int_u^\infty f_I(x)\ \mathrm{d}x$
- $g(x) = \frac{\lambda\mu}{c} e^{Rx} f_I(x)$. Note that this is a density because $g(x) \geq 0$ and $\frac{\lambda\mu}{c}\int_0^\infty e^{Rx} f_I(x)\ \mathrm{d}x = \frac{\lambda}{c}\int_0^\infty e^{Rx}\big(1-F_X(x)\big)\ \mathrm{d}x = 1$, by the definition of the Lundberg exponent, which states that

$$\begin{aligned}
M_X(R) &= \frac{cR}{\lambda} + 1 \\
\int_0^\infty e^{Rx} f_X(x)\ \mathrm{d}x &= \frac{cR}{\lambda} + 1 \\
\int_0^\infty e^{Rx} \tfrac{\mathrm{d}}{\mathrm{d}x}\Big(-(1 - F_X(x)\Big)\ \mathrm{d}x &= \frac{cR}{\lambda} + 1 \\
1 + R\int_0^\infty e^{Rx} \Big(1 - F_X(x)\Big)\ \mathrm{d}x &= \frac{cR}{\lambda} + 1 \\
\frac{\lambda}{c}\int_0^\infty e^{Rx} \Big(1 - F_X(x)\Big)\ \mathrm{d}x &= 1
\end{aligned}$$

Now, (*) is a "renewal type equation". We use a small result from renewal theory (see Feller, Vol 2 Chap 11) which states that if z is integrable and equals the difference of two non-decreasing functions, then

$$Z(u) \to A = \frac{\int_0^\infty z(x)\ \mathrm{d}x}{\int_0^\infty xg(x)\ \mathrm{d}x} \qquad \text{as } u \to \infty \qquad (\#)$$

In our case

$$\begin{aligned}
z(u) &= \frac{\lambda\mu}{c} e^{Ru} \int_u^\infty f_I(x)\ \mathrm{d}x \\
&= \frac{\lambda\mu}{c} e^{Ru} \left(\int_0^\infty f_I(x)\ \mathrm{d}x - \int_0^u f_I(x)\ \mathrm{d}x \right) \\
&= \frac{\lambda\mu}{c} e^{Ru} - \frac{\lambda\mu}{c} e^{Ru} \int_0^u f_I(x)\ \mathrm{d}x
\end{aligned}$$

Both functions are non-decreasing, so the result above applies.

Now

$$\begin{aligned}
\int_0^\infty z(x)\ \mathrm{d}x &= \frac{\lambda\mu}{c} \int_{x=0}^\infty e^{Rx} \int_{t=x}^\infty f_I(t)\ \mathrm{d}t\ \mathrm{d}x \\
&= \frac{\lambda\mu}{c} \int_{t=0}^\infty f_I(t) \int_{x=0}^t e^{Rx}\ \mathrm{d}x\ \mathrm{d}t \\
&= \frac{\lambda\mu}{c} \int_{t=0}^\infty f_I(t) \frac{1}{R}\left\{ e^{Rt} - 1 \right\}\ \mathrm{d}t \\
&= \frac{\lambda\mu}{cR} \left\{ \int_0^\infty e^{Rt} f_I(t)\ \mathrm{d}t - 1 \right\} \\
&= \frac{1}{R} \int_0^\infty g(t)\ \mathrm{d}t - \frac{\lambda\mu}{cR} \\
&= \frac{1}{R}\left(1 - \frac{\lambda\mu}{c} \right)
\end{aligned}$$

Therefore, using result (#), we have

$$\begin{aligned} A^{-1} &= \frac{\int_0^\infty xg(x)\ \mathrm{d}x}{\int_0^\infty z(x)\ \mathrm{d}x} \\ &= \frac{\frac{\lambda\mu}{c}\int_0^\infty xe^{Rx} f_X(x)\ \mathrm{d}x}{\frac{1}{R}\left(1-\frac{\lambda\mu}{c}\right)} \\ &= \frac{R}{\rho}\int_0^\infty xe^{Rx} f_X(x)\ \mathrm{d}x \end{aligned}$$

As required. ■

Note that this implies $\psi(u) \to Ae^{-Ru}$ as $u \to \infty$.

Note also that A can be written

$$A = \left(\frac{M_X'(R) - \frac{c}{\lambda}}{\mu\rho}\right)^{-1}$$

<u>Example</u>: If $X_1 \sim \exp(1/\mu)$, we find $\psi(u) = \frac{1}{1+\rho}\exp\left(-\frac{\rho}{(1+\rho)\mu}u\right)$, and

$$R = \frac{\rho}{(1+\rho)\mu}.$$ □

Credibility Theory

Credibility theory is used when we wish to estimate the expected aggregate claim (or the expected number of claims) in the coming time period for a **single risk** (ie: a single policy or group of policies) based on

- An estimate A based on data from the risk itself
- An estimate B based on "collateral information" from somewhere else (for example, information from similar but not identical risks).

The credibility approach is to use the **credibility formula**

$$C = zA + (1-z)B \qquad z \in (0,1]$$

Where C is the expected aggregate claim, and z is known as the **credibility factor**. We expect z to

- Increase as the number of data points for the risk itself increases.
- Decrease for "more relevant" collateral information

Our work in this area will require some Bayesian concepts. We briefly review those here:

- If X has density $f(x;\theta)$ and θ is a random variable with **prior distribution** $\pi(\theta)$, the **posterior density** for θ given $X = x$ is given by

$$\pi(\theta \mid x) = \frac{\pi(\theta) f(x;\theta)}{\int \pi(\theta) f(x;\theta)\ \mathrm{d}\theta}$$

- To estimate θ on the basis of data $\boldsymbol{x} = \left(x_1, \cdots, x_n\right)$, we define $L\left(\theta, g(\boldsymbol{x})\right)$ to be the loss incurred when $g(\boldsymbol{x})$ is used as an estimator for θ. The **Bayes' estimator** minimizes the **expected posterior loss**

$$\int L\left(\theta, g(\boldsymbol{x})\right) \pi(\theta \mid \boldsymbol{x})\ \mathrm{d}\theta$$

if we use a **quadratic loss function** $L\left(\theta, g(\boldsymbol{x})\right) = \left(g(\boldsymbol{x}) - \theta\right)^2$, the Bayes' estimator is the **posterior mean of** θ:

$$g(\boldsymbol{x}) = \mathbb{E}\left(\theta \mid \boldsymbol{x}\right)$$

- We will also need the **conditional covariance formula**

$$\begin{aligned}
\text{cov}(X,Y) &= \mathbb{E}\Big(\big(X-\mu_X\big)\big(Y-\mu_S\big)\Big) \\
&= \mathbb{E}\big(XY - X\mu_Y - Y\mu_X + \mu_X\mu_Y\big) \\
&= \mathbb{E}\big(XY\big) - \mathbb{E}\big(X\big)\mathbb{E}\big(Y\big) \\
&= \mathbb{E}\big\{\mathbb{E}(XY \mid Z)\big\} - \mathbb{E}\big\{\mathbb{E}(X \mid Z)\big\}\mathbb{E}\big\{\mathbb{E}(Y \mid Z)\big\} \\
&= \mathbb{E}\big\{\text{cov}\big(X,Y \mid Z\big) + \mathbb{E}(X \mid Z)\mathbb{E}(Y \mid Z)\big\} \\
&\qquad - \mathbb{E}\big\{\mathbb{E}(X \mid Z)\big\}\mathbb{E}\big\{\mathbb{E}(Y \mid Z)\big\} \\
&= \mathbb{E}\big\{\text{cov}(X,Y \mid Z\big\} + \mathbb{E}\big\{\mathbb{E}(X \mid Z)\mathbb{E}(Y \mid Z)\big\} \\
&\qquad - \mathbb{E}\big\{\mathbb{E}(X \mid Z)\big\}\mathbb{E}\big\{\mathbb{E}(Y \mid Z)\big\} \\
&= \mathbb{E}\big\{\text{cov}(X,Y \mid Z\big\} + \text{cov}\big\{\mathbb{E}(X \mid Z), \mathbb{E}(Y \mid Z)\big\}
\end{aligned}$$

(This reduces to the conditional variance formula when $X = Y$)

Bayesian Credibility Theory (*Exact Credibility*)

In Bayesian credibility, the concept of "collateral information" is formalised in terms of a **period density** of θ, which is chosen to reflect subjective degrees of belief about the value of θ. We set up our model as follows:

- Let X be yearly aggregate claims with density $f(x;\theta)$
- Let $\pi(\theta)$ be the prior density of θ.
- Suppose that we have n observations of X, $\boldsymbol{x} = \big(x_1, \cdots, x_n\big)$, and that $X_1 \mid \theta, \cdots, X_n \mid \theta$ are independent – these are data from the **particular risk itself**.

We are interested in the aggregate claims for the coming year. If θ was known, then our answer would be $\mu(\theta) = \mathbb{E}(X \mid \theta)$.

We do not, however, know θ. This means that we can do two things:

- ***Estimate X based only on the prior $\boldsymbol{\pi(\theta)}$***. In this case, we're basing our estimate of X on the **collateral data** only, encapsulated in $\pi(\theta)$:

$$C = \mathbb{E}_{\pi(\theta)}\big[\mu(\theta)\big]$$

- ***Estimate X based on the prior $\boldsymbol{\pi(\theta)}$ as well as on $\boldsymbol{x}$***. In this case, we're basing our estimate of X on **collateral data** encapsulated in $\pi(\theta)$ *as well as* on **specific data** from the risk itself, encapsulated in $\boldsymbol{x}$:

  To do this, we use the **posterior mean**, which is the optimal Bayesian estimator under quadratic loss

$$C = \mathbb{E}_{\pi(\theta)}\big[\mu(\theta) \mid \boldsymbol{x}\big] = \mathbb{E}_{\pi(\theta \mid \boldsymbol{x})}\big[\mu(\theta)\big] = \int \mu(\theta)\pi(\theta \mid \boldsymbol{x})\ \mathrm{d}\theta$$

For certain special choices of f and π, we find that this estimate takes the form of a credibility estimate:

$$C = z\begin{pmatrix}\text{Something based on}\\ \text{the data only}\end{pmatrix} + (1-z)\begin{pmatrix}\text{Something based on}\\ \text{the prior/collateral only}\end{pmatrix}$$

with a specific formula for z.

<u>EXAMPLE</u>: Consider a situation in which

- $X \mid \theta \sim N(\theta, \sigma_1^2)$, σ_1 known
- $\theta \sim N(\mu, \sigma_2^2)$, μ, σ_2 known
- This means that

$$\begin{aligned}\pi(\theta \mid \boldsymbol{x}) &\propto \pi(\theta) f(\boldsymbol{x} \mid \theta)\\ &\propto \exp\left\{-\frac{1}{2\sigma_1^2}\sum_{i=1}^{n}(x_i - \theta)^2 - \frac{(\theta-\mu)^2}{2\sigma_2^2}\right\}\\ &\propto \exp\left\{-\frac{1}{2}\left[\left(\frac{n}{\sigma_1^2} + \frac{1}{\sigma_2^2}\right)\theta^2 - 2\left(\frac{x_+}{\sigma_1^2} + \frac{\mu}{\sigma_2^2}\right)\theta\right]\right\}\end{aligned}$$

By completing the square on the denominator, we find that

$$\pi(\theta \mid \boldsymbol{x}) \sim N\left(\mathrm{M}, \Sigma^2\right)$$

$$M = \frac{n\overline{x}\sigma_2^2 + \mu\sigma_1^2}{n\sigma_2^2 + \sigma_1^2} \qquad \Sigma^2 = \frac{\sigma_1^2\sigma_2^2}{n\sigma_2^2 + \sigma_1^2}$$

Now, back to the example. It's clear that $\mu(\theta) = \mathbb{E}(X \mid \theta) = \theta$. In other words, if we know θ, our best estimate for X is θ. But we don't know θ, so let's see what we can do:

- Based only on collateral information,

$$C = \mathbb{E}_\pi\left[\mu(\theta)\right] = \mathbb{E}_\pi\left(\theta\right) = \mu$$

- Based on the data as well as the collateral

$$C = \mathbb{E}_\pi\left[\mu(\theta) \mid \boldsymbol{x}\right] = \mathbb{E}_{\pi(\theta|\boldsymbol{x})}(\theta) = \frac{n\overline{x}\sigma_2^2 + \mu\sigma_1^2}{n\sigma_2^2 + \sigma_1^2}$$

Which can be written as

$$\boxed{C = z\overline{x} + (1-z)\mu}$$

Where

$$z = \frac{n}{n + \frac{\sigma_1^2}{\sigma_2^2}} = \frac{n}{n + \frac{\mathrm{var}(X|\theta)}{\mathrm{var}(\mu(\theta))}}$$

This puts C precisely in the form we were interested in, since $\overline{x}$ depends only on the data, whereas μ depends only on the collateral (since μ was our estimate of X using only collateral information, in the previous bullet point).

Note that our expression for n meets our intuitive expectations:

- As n increases (more actual data) z increases (more weight on the actual data)
- As σ_2 increases (collateral data less precise) z increases (more weight on the actual data).

Note also that although $X_1 \mid \theta, \cdots, X_n \mid \theta$ are independent, the X themselves are not necessarily independent, since

$$\begin{aligned}
\mathbb{E}\left(X_1 X_2\right) &= \mathbb{E}\left[\mathbb{E}\left(X_1 X_2 \mid \theta\right)\right] \\
&= \mathbb{E}\left[\mathbb{E}(X_1 \mid \theta)\mathbb{E}(X_2 \mid \theta)\right] \\
&= \mathbb{E}\left(\theta^2\right) \\
&= \operatorname{var}(\theta) + \mathbb{E}(\theta)^2 \\
&= \sigma_2^2 + \mu^2 \\
\mathbb{E}(X_1)\mathbb{E}(X_2) &= \mathbb{E}\left[\mathbb{E}(X_1 \mid \theta)\right]\mathbb{E}\left[\mathbb{E}(X_2 \mid \theta)\right] \\
&= \mathbb{E}(\theta)^2 \\
&= \mu^2
\end{aligned}$$

These are not generally equal, unless $\sigma_2 = 0$. □

We can also get exact credibility if X is the **number of claims** in a given time period, $X \mid \theta \sim \text{Po}(\theta)$ and $\theta \sim \Gamma(\alpha, \lambda)$. In general, we do not get exact credibility.

Empirical Bayesian Credibility – The Buhlman Model

Usually, we know neither f nor π. All we have is

- n observations X pertaining to the risk in question.
- Some observations pertaining to other, similar policies.

We would like to get a credibility estimate out of these data. We define the following quantities:

- $\mu(\theta) = \mathbb{E}\left(X_1 \mid \theta\right)$, the expected claim amount *assuming* θ is known. (This is a random variable, since θ is a random variable).
    - Note that $\text{var}\left(\mu(\theta)\right) = \text{var}\left(\mathbb{E}(X_1 \mid \theta)\right)$ is a measure of how different the various models are – in other words, it's a measure of how reliable the *collateral* data is.
- $\sigma^2\left(\theta\right) = \text{var}\left(X_1 \mid \theta\right)$, the expected claim amount variance *assuming* θ is known (this is also a random variable).
    - Note that $\sigma^2(\theta)$ is effectively the variance for our data on a *given* risk. Thus, $\mathbb{E}\left(\sigma^2(\theta)\right) = \mathbb{E}\left(\text{var}\left(X_1 \mid \theta\right)\right)$ is a measure of how reliable the *specific* data for a risk is.

And:

- $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$ – this is the mean of all observations available for our specific risk. It only contains information about the risk itself.
- $m = \mathbb{E}\left(X_1\right) = \mathbb{E}\left(\mu(\theta)\right)$, the expected average claim amount – this is the premium we would charge to a new claim with no history. It is based entirely on **collateral information**.

(Once again, we assume all moments are finite as needed. Note also that this analysis is valid if the X are claim frequencies).

Our method is as follows:

- We'll first derive a credibility estimate of the form

$$C = z\begin{pmatrix}\text{Something based on} \\ \text{the specific data } X\end{pmatrix} + \left(1-z\right)\begin{pmatrix}\text{Something based on} \\ \text{collateral data } \left\{X_{js}\right\}\end{pmatrix}$$

$$C = z\bar{X} + (1-z)m$$

- We'll then use the data available to estimate the following quantities

$$\mathbb{E}\left(\sigma^2(\theta)\right) \qquad \text{var}\left(\mu(\theta)\right) \qquad m$$

  m is directly needed in C, and the other quantities are needed to work out z.

Step 1

As we saw above, the Bayesian estimator minimising squared error loss is the posterior mean

$$C = \mathbb{E}\left(\mu(\theta) \mid \text{all data}\right)$$

However, this sometimes does not take the form of a credibility estimate. To ensure we obtain something of the form $C = z\overline{X} + (1-z)m$, we restrict our attention to

$$C = C_0 + \sum_{j=1}^{n} C_j X_j$$

chosen so as to minimize the squared error loss

$$L = \mathbb{E}\left\{\left(\mu(\theta) - C_0 - \sum_{j=1}^{n} C_j X_j\right)^2\right\}$$

Taking derivatives, we obtain

$$\frac{\partial L}{\partial C_0} = \mathbb{E}\left\{\mu(\theta) - C_0 - \sum_{j=1}^{n} C_j X_j\right\} = 0 \qquad (1)$$

$$\frac{\partial L}{\partial C_r} = \mathbb{E}\left\{X_r\left(\mu(\theta) - C_0 - \sum_{j=1}^{n} C_j X_j\right)\right\} = 0 \qquad \forall r \qquad (2)$$

Time for some acrobatics to find C_r and C_0

- (*Finding* C_r) Taking $(2) - \mathbb{E}(X_r)(1)$ gives

$$\mathbb{E}\left\{X_r\left(\mu(\theta) - C_0 - \sum_{j=1}^{n} C_j X_j\right)\right\} - \mathbb{E}(X_r)\mathbb{E}\left\{\mu(\theta) - C_0 - \sum_{j=1}^{n} C_j X_j\right\} = 0$$

$$\mathbb{E}\left(X_r\mu(\theta)\right) - \mathbb{E}(X_r)\mathbb{E}(\mu(\theta)) = \mathbb{E}\left(X_r \sum_{j=1}^{n} C_j X_j\right) - \mathbb{E}(X_r)\mathbb{E}\left(\sum_{j=1}^{n} C_j X_j\right)$$

$$\operatorname{cov}\left(\mu(\theta), X_r\right) = \sum_{j=1}^{n} C_j \operatorname{cov}\left(X_r, X_j\right) \qquad \forall r \qquad (3)$$

We can use the conditional covariance formula on both the LHS and the RHS of (3):

$$\begin{aligned}
\operatorname{cov}(X_r, Y_j) &= \mathbb{E}\left\{\operatorname{cov}(X_r, X_j \mid \theta)\right\} + \operatorname{cov}\left\{\mathbb{E}(X_r \mid \theta), \mathbb{E}(X_j \mid \theta)\right\} \\
&= \mathbb{E}\left\{\delta_{rj}\sigma^2(\theta)\right\} + \operatorname{cov}\left\{\mu(\theta), \mu(\theta)\right\} \\
&= \delta_{rj}\mathbb{E}\left(\sigma^2(\theta)\right) + \operatorname{var}\left(\mu(\theta)\right) \\
\operatorname{cov}\left(\mu(\theta), X_r\right) &= \mathbb{E}\left\{\operatorname{cov}\left(\mu(\theta), X_r \mid \theta\right)\right\} + \operatorname{cov}\left\{\mathbb{E}(\mu(\theta) \mid \theta), \mathbb{E}(X_r \mid \theta)\right\} \\
&= \mathbb{E}\left\{\mu(\theta)\operatorname{cov}\left(1, X_r \mid \theta\right)\right\} + \operatorname{var}\left(\mu(\theta)\right) \\
&= \operatorname{var}\left(\mu(\theta)\right)
\end{aligned}$$

Where δ_{rj} is the Kronecker delta, equal to 1 if $r = j$ and 0 otherwise.

So (3) becomes

$$\operatorname{var}\Big(\mu(\theta)\Big) = C_r \mathbb{E}\Big(\sigma^2(\theta)\Big) + \operatorname{var}\Big(\mu(\theta)\Big)\sum_{j=1}^{n} C_j \qquad \forall r \tag{4}$$

Directly from (4), we get

$$C_r = \frac{\operatorname{var}\Big(\mu(\theta)\Big)}{\mathbb{E}\Big(\sigma^2(\theta)\Big)}\left(1 - \sum_{j=1}^{n} C_j\right) \tag{4b}$$

Add up (4) from 1 to n to get

$$n \operatorname{var}\Big(\mu(\theta)\Big) = \Big\{\mathbb{E}\Big(\sigma^2(\theta)\Big) + n \operatorname{var}\Big(\mu(\theta)\Big)\Big\}\sum_{j=1}^{n} C_j$$

$$\sum_{j=1}^{n} C_j = \frac{1}{1 + \frac{\mathbb{E}\left(\sigma^2(\theta)\right)}{n \operatorname{var}\left(\mu(\theta)\right)}} \tag{5}$$

Feeding (5) into (4b), we get

$$\boxed{C_r = \frac{1}{n}\left[1 + \frac{\mathbb{E}\left(\sigma^2(\theta)\right)}{n \operatorname{var}\left(\mu(\theta)\right)}\right]^{-1}}$$

- (*Finding* C_0) From (1), we get

$$\mathbb{E}\Big(\mu(\theta)\Big) - C_0 - \sum\nolimits_{j=1}^{n} C_j \mathbb{E}\Big(X_j\Big) = 0$$
$$C_0 = m\Big(1 - \sum\nolimits_{j=1}^{n} C_j\Big)$$

once again, feeding (5) into this gives

$$\boxed{C_0 = m\left(1 - \left[1 + \frac{\mathbb{E}\left(\sigma^2(\theta)\right)}{n \operatorname{var}\left(\mu(\theta)\right)}\right]^{-1}\right)}$$

Feeding C_0 and C_r into $C_0 + \sum_{j=1}^{n} C_j X_j$, we obtain

$$z\overline{X} + (1 - z)m \tag{5.4}$$

Where

$$z = \frac{1}{1 + \frac{\mathbb{E}\left(\sigma^2(\theta)\right)}{n \operatorname{var}\left(\mu(\theta)\right)}} = \frac{n}{n + \frac{\mathbb{E}\left(\sigma^2(\theta)\right)}{\operatorname{var}\left(\mu(\theta)\right)}} = \frac{n}{n + \frac{\text{Var of individual risk}}{\text{Var of collateral info}}}$$

Which is indeed in the form of a credibility estimate.

Note that:

- In this case, C_r does not depend on r, because every $\boldsymbol{X}$ is weighed identically. We could have used this to simplify the equations above.
- As n increases, z increases to 1.

- As $\text{var}\big(\mu(\theta)\big)$ increases, z increases (ie: as the collateral information becomes less relevant, we put more weight on the data pertaining to the risk itself).

Step 2

We now estimate the quantities needed for the credibility estimate. We will assume that the data we have is in the form $\left\{X_{js}\right\}$, pertaining to k other policies over n time periods, where

- X_{js} is the claim amount (or number of claims) in time period s for policy j. We also write $\boldsymbol{X}_j = \left(X_{j1}, \cdots, X_{jn}\right)$ for the data about a single policy.
- Each of the k policies has its own structure variable $\theta_1, \cdots, \theta_k$, which are IID with (unknown) distribution $\pi(\theta)$.

We assume the following dependence structure:

- Within any policy, $X_{j1} \mid \theta_j, \cdots, X_{jn} \mid \theta_j$ are IID (but the X_{ji} may not be independent themselves.)
- All the $\left(\theta_j, \boldsymbol{X}_j\right)$ are IID

Now, let $\mu(\theta_j) = \mathbb{E}\left(X_{js} \mid \theta_j\right)$, and note that from our assumptions, we have $\text{cov}\left(\boldsymbol{X}_j \mid \theta_j\right) = \sigma^2(\theta_j)\mathrm{I}_n$. We then use the following estimators for our quantities of interest:

$$\mu(\theta_j) = \mathbb{E}(X_{j1} \mid \theta_j) = M_j = \frac{1}{n}\sum_{s=1}^{n} X_{js}$$

$$m = \mathbb{E}\left[\mu(\theta_j)\right] = M_0 = \frac{1}{k}\sum_{j=1}^{k} M_j = \frac{1}{kn}\sum_{j=1}^{k}\sum_{s=1}^{n} X_{js}$$

$$\mathbb{E}\left[\sigma^2(\theta)\right] = \mathbb{E}\left(\text{var}(X_{j1} \mid \theta_j)\right) = \frac{1}{k}\sum_{j=1}^{k}\frac{1}{n-1}\sum_{s=1}^{n}\left(X_{js} - M_j\right)^2 = s^2$$

$$\text{var}\left(\mu(\theta)\right) = \frac{1}{k-1}\sum_{j=1}^{k}\left(M_j - M_0\right)^2 - \frac{1}{n}s^2$$

Empirical Bayesian Credibility – The Buhlman-Straub Model

We now consider a more complicated model which takes into account the volume of business for the different risks and time periods. We set up our model as follows:

- Let $Y_1, \cdots, Y_n$ be the claim amounts for a particular risk
- Let $p_1, \cdots, p_n$ be the (known) volumes of business for each of these claim amounts (for example, number of policies or premium income)

We then define $X_j = Y_j / p_j$, the income per unit volume for each period, and we assume that

- $X_1 \mid \theta, \cdots, X_n \mid \theta$ are independent
- $\mathbb{E}\left(X_j \mid \theta\right)$ and $p_j \operatorname{var}\left(X_j \mid \theta\right)$ do not depend on j

We define

- $\mu(\theta) = \mathbb{E}(X_j \mid \theta)$
- $\sigma^2(\theta) = p_j \operatorname{var}\left(X_j \mid \theta\right)$

<u>Example</u>: Suppose a particular risk is made up for a number of independent policies. In time period j, there are p_j policies. If claims from a single policy have mean $\mu(\theta)$ and variance $\sigma^2(\theta)$, then the total claim amount for that period has mean $p_j \mu(\theta)$ and variance $p_j \sigma^2(\theta)$.

Thus, $\mathbb{E}(X_j \mid \theta) = \mu(\theta)$ and $\operatorname{var}\left(X_j \mid \theta\right) = \sigma^2(\theta) / p_j$. □

We once again assume the credibility premium takes the form

$$C = a_0 + \sum_{j=1}^{n} a_j X_j$$

chosen so as to minimize the squared error loss

$$L = \mathbb{E}\left\{\left(\mu(\theta) - a_0 - \sum\nolimits_{j=1}^{n} a_j X_j\right)^2\right\}$$

Once again, we take derivatives and obtain equations (1) and (2) above.

We then engage in similar acrobatics to find a_r and a_0

- (*Finding C_r*) Taking $(2) - \mathbb{E}(X_r)(1)$ gives

$$\operatorname{cov}\left(\mu(\theta), X_r\right) = \sum_{j=1}^{n} C_j \operatorname{cov}\left(X_r, X_j\right) \tag{3}$$

Using the conditional variance formula on the LHS gives the same results as above. On the RHS, however,

$$\begin{aligned}\operatorname{cov}(X_r, Y_j) &= \mathbb{E}\left\{\operatorname{cov}(X_r X_j \mid \theta\right\} + \operatorname{cov}\left\{\mathbb{E}(X_r \mid \theta), \mathbb{E}(X_j \mid \theta)\right\} \\ &= \mathbb{E}\left\{\delta_{rj} \operatorname{var}(X_j \mid \theta)\right\} + \operatorname{cov}\left\{\mu(\theta), \mu(\theta)\right\} \\ &= \delta_{rj} \frac{1}{p_j} \mathbb{E}\left(\sigma^2(\theta)\right) + \operatorname{var}\left(\mu(\theta)\right)\end{aligned}$$

Feeding this back into (3), we get

$$\operatorname{var}\left(\mu(\theta)\right) = \frac{a_r}{p_r} \mathbb{E}\left(\sigma^2(\theta)\right) + \operatorname{var}\left(\mu(\theta)\right) \sum_{j=1}^{n} a_j$$

$$p_r \operatorname{var}\left(\mu(\theta)\right) = a_r \mathbb{E}\left(\sigma^2(\theta)\right) + p_r \operatorname{var}\left(\mu(\theta)\right) \sum_{j=1}^{n} a_j \tag{4}$$

Re-arranging (4)

$$a_r = p_r \frac{\operatorname{var}\left(\mu(\theta)\right)}{\mathbb{E}\left(\sigma^2(\theta)\right)} \left\{1 - \sum_{j=1}^{n} a_j\right\} \tag{4b}$$

Adding (4) up from 1 to n, and letting $p_+ = \sum_{i=1}^{n} p_j$, we get

$$p_+ \operatorname{var}\left(\mu(\theta)\right) = \left\{\mathbb{E}\left(\sigma^2(\theta)\right) + p_+ \operatorname{var}\left(\mu(\theta)\right)\right\} \sum_{j=1}^{n} a_j$$

$$\sum_{j=1}^{n} a_j = \frac{p_+}{p_+ + \frac{\mathbb{E}\left(\sigma^2(\theta)\right)}{\operatorname{var}\left(\mu(\theta)\right)}} \tag{5}$$

Feeding (5) into (4b), we get

$$\boxed{a_r = \frac{p_r}{p_+}\left[1 + \frac{\mathbb{E}\left(\sigma^2(\theta)\right)}{p_+ \operatorname{var}\left(\mu(\theta)\right)}\right]^{-1}}$$

- (*Finding* C_0) From equation (1), we get

$$\mathbb{E}\left(\mu(\theta)\right) - a_0 - \sum_{j=1}^{n} a_j \mathbb{E}\left(X_j\right) = 0$$
$$a_0 = m\left(1 - \sum_{j=1}^{n} a_j\right)$$

Feeding (5) into this gives

$$\boxed{a_0 = m\left(1 - \left[1 + \frac{\mathbb{E}\left(\sigma^2(\theta)\right)}{p_+ \operatorname{var}\left(\mu(\theta)\right)}\right]^{-1}\right)}$$

Feeding back into $C = a_0 + \sum_{j=1}^{n} a_j X_j$, our credibility estimate *per unit volume* is

$$C = z\tilde{X} + \left(1 - z\right)\mathbb{E}\left(\mu(\theta)\right)$$

Where

$$z = \left[1 + \frac{\mathbb{E}\left(\sigma^2(\theta)\right)}{p_+ \operatorname{var}\left(\mu(\theta)\right)}\right]^{-1} \qquad \tilde{X} = \frac{\sum_{j=1}^{n} p_j X_j}{p_+}$$

Note that

- If every $p_j = 1$, then $p_j = n$, and we recover the Buhlman credibility factor.
- The quantities $\mathbb{E}\left(\sigma^2(\theta)\right), \operatorname{var}\left(\mu(\theta)\right), \mathbb{E}\left(\mu(\theta)\right)$ must be estimated from data.

No Claims Discount (NCD) Systems

No claims discount systems give the policyholder a discount on the usual premium, the size of the discount being based on the number of claim-free years for the policy holder. For example, a motor insurance scheme may have 3 discount categories. Policyholders in category 0 pay the full previous c, those in category 1 pay $0.7c$ and those in category 2 pay $0.6c$. If a policyholder makes no claims in a particular year, they move up to the next category (or stays in category 2). If they make ≥ 1 claim, they move down a category (or stay in category 0).

Suppose the categories are $0,1,\cdots,d$, and consider a policyholder who takes out a policy at year 0 and enters in category 0. Let X_n be their discount category in year n. Finally, suppose the distribution of the number of claims per year is the same each year. Then $\{X_n\}$ is a discrete-time time-homogeneous Markov chain with finite statespace $0,1,\cdots,d$. Its transition matrix is $P=(p_{ij})$, where $p_{ij}=\mathbb{P}\left(X_{n+1}=j \mid X_n=i\right)$.

<u>EXAMPLE</u>: In our motor insurance example, in which we had three categories

$$P = \text{From}\begin{cases}(0)\\(1)\\(2)\end{cases}\overbrace{\begin{pmatrix} 1-p & p & 0 \\ 1-p & 0 & p \\ 0 & 1-p & p \end{pmatrix}}^{\text{To: }(0)\quad(1)\quad(2)}$$

Where $p=\mathbb{P}\left(\text{No claims in a given year}\right)$. □

Now, let $\pi_i^{(n)}=\mathbb{P}\left(X_n=i\right)$ and $\boldsymbol{\pi}^{(n)}=\left(\pi_0^{(n)},\cdots,\pi_d^{(n)}\right)$. At any given time, this vector contains the probability of being in each state. For example, since the policyholder enters in category 0, we have $\boldsymbol{\pi}^{(0)}=\left(1,0,\cdots,0\right)$.

Now

$$\begin{aligned}\pi_j^{(n+1)} &= \mathbb{P}\left(X_{n+1} = j\right) \\ &= \sum_{i=0}^{d} \mathbb{P}\left(X_{n+1} = j \mid X_n = i\right)\mathbb{P}\left(X_n = i\right). \\ &= \sum_{i=0}^{d} p_{ij}\pi_i^{(n)}\end{aligned}$$

in other words

$$\boldsymbol{\pi}^{(n+1)} = \boldsymbol{\pi}^{(n)}P \tag{6.1}$$

The stochastic evolution of $\{X_n\}$ therefore only depends on P and $\boldsymbol{\pi}^{(0)}$.

Under certain conditions (always satisfied in our examples), $\boldsymbol{\pi}^{(n)} \to \boldsymbol{\pi}$ as $n \to \infty$. To find this *equilibrium distribution*, let $n \to \infty$ in (6.1). This gives $\boldsymbol{\pi} = \boldsymbol{\pi}P$. Solving this (redundant) system of linear equations together with $\sum_i \pi_i = 1$ allows us to find $\boldsymbol{\pi}$.

EXAMPLE: In our motor insurance example, the system of equations $\boldsymbol{\pi} = \boldsymbol{\pi}P$ is

$$\begin{pmatrix}\pi_0 & \pi_1 & \pi_2\end{pmatrix}\begin{pmatrix}1-p & p & 0 \\ 1-p & 0 & p \\ 0 & 1-p & p\end{pmatrix}$$

$$\Downarrow$$

$$\begin{aligned}\pi_0(1-p) + \pi_1(1-p) &= \pi_0 \\ \pi_0 p + \pi_2(1-p) &= \pi_1 \\ \pi_1 p + \pi_2 p &= \pi_2\end{aligned}$$

This gives

$$\boldsymbol{\pi} = \pi_0\left(1, \frac{p}{1-p}, \frac{p^2}{(1-p)^2}\right)$$

Using $\sum_i \pi_i = 1$, we find that

$$\pi_0 = \frac{\frac{p}{1-p} - 1}{\left(\frac{p}{1-p}\right)^3 - 1}$$

For example, if $p = 0.9$, then

$$\boldsymbol{\pi} = \tfrac{1}{91}\left(1, 9, 81\right)$$

This could then be used to find the expected premium payable. □

Consider a final application of our example:

<u>EXAMPLE</u>: Suppose the driver pays the full premium in year 0 (ie: they are in category 0) and then has an accident in that year with repair cost ℓ.

Assuming no further accidents, compare the cost to the driver if they *do* claim and if they do *not* claim:

Year	0	1	2	3	4
If claim made	0	c	$0.7c$	$0.6c$	$0.6c$
If no claim made	ℓ.	$0.7c$	$0.6c$	$0.6c$	$0.6c$

The driver's decision to pay or not depends on their time horizon:

- **Two year time horizon (0 and 1)** – the driver will claim if
$$0 + c < 0.7c + \ell$$
$$\boxed{\ell > 0.3c}$$
- **Infinite time horizon** – the driver will claim if
$$0 + c + 0.7c > \ell + 0.7c + 0.6c$$
$$\boxed{\ell > 0.4c}$$

In general, the loss ℓ is a random variable (say lognormal), then we could find the probability p of claiming, and use that in the transition matrix.

Of course, it is also true that drivers currently in different categories will have different "thresholds" for claiming. □

Run-off triangles

Delays may occur at various stages in settling claims – for example, incurred but not reported claims, or outstanding reported claims.

EXAMPLE: Imagine the last year for which we have complete data is 2009. Then a *run-off* (or *delay*) triangle might look like this

Development year

Claim payments (£000)	0	1	2	3
2006	300 (*2006*)	500 (*2007*)	200 (*2008*)	100 (*2009*)
2007	500 (*2007*)	700 (*2008*)	300 (*2009*)	*(No data available for 2010)*
2008	400 (*2008*)	600 (*2009*)	*(No data available for 2010)*	
2009	500 (*2009*)	*(No data available for 2010)*		

Accident Year

The diagonals corresponds to payments in a given calendar year.

□

We begin by developing some notation:

- Let Y_{ij} be the amount paid for accident year i in development (not calendar) year j.
- Let $C_{ij} = \sum_{k=0}^{j} Y_{ik}$ be the total amount paid for accident year i up to j development years after i.

We observe Y_{ij} and C_{ij} for $i = 0, \cdots, d$ and $j = 0, \cdots, d - i$, where d is the last full year for which complete information is available. Note also that $i + j$ is the calendar year of a given payment.

Our aim is to obtain projections for the amounts yet to be paid.

The Chain-Ladder Technique

We assume that the expected value in cell (i, j) is

$$\mathbb{E}(Y_{ij}) = n_i r_j \tag{7.1}$$

where

- n_i reflects the volume of claims relating to accident year i.
- r_j is a factor related to the development year j.

We assume that the r_j do not vary over accident years, and we further assume that the claims for year 0 are "fully run off" – ie: they are finally settled by development year d, so that $r_0 + \cdots + r_d = 1$.

Under equation (7.1), we have

$$\begin{aligned}\mathbb{E}\left(C_{ij}\right) &= n_i\left(r_0 + \cdots + r_j\right) \\ &= \frac{r_0 + \cdots + r_j}{r_0 + \cdots + r_{j-1}}\mathbb{E}\left(C_{i,(j-1)}\right) \\ &= \left(1 + \frac{r_j}{r_0 + \cdots + r_{j-1}}\right)\mathbb{E}\left(C_{i,(j-1)}\right)\end{aligned}$$

and we write

$$\mathbb{E}(C_{ij}) = \lambda_j \mathbb{E}\left(C_{i,(j-1)}\right) \tag{7.2}$$

We can use 7.2 to estimate the λ_j by

$$\lambda_j = \frac{\mathbb{E}(C_{ij})}{\mathbb{E}\left(C_{i,(j-1)}\right)}$$

(and equating expected values to observed values).

However, for any given development year j there might be a number of accident years i available to estimate λ_j. The *chain-ladder* technique takes the following *weighed average* of these values

$$\begin{aligned}\hat{\lambda}_j &= \frac{\frac{C_{0j}}{C_{0,(j-1)}}C_{0,(j-1)} + \cdots + \frac{C_{(d-j),j}}{C_{(d-j),(j-1)}}C_{(d-j),(j-1)}}{C_{0,(j-1)} + \cdots + C_{(d-j),(j-1)}} \\ &= \frac{C_{0,j} + \cdots + C_{(d-j),j}}{C_{0,(j-1)} + \cdots + C_{(d-j),(j-1)}}\end{aligned}$$

EXAMPLE: In the example above, we begin by drawing up a table of the cumulative amounts C_{ij}

Development year

Accident Year	Claim payments (£000)	0	1	2	3
	2006	300	800	1000	1100
	2007	500	1200	1500	
	2008	400	1000		
	2009	500			

We then calculate

$$\hat{\lambda}_1 = \frac{800+1200+1000}{300+500+400} = 2.5$$
$$\hat{\lambda}_2 = \frac{1000+1500}{800+1200} = 1.25$$
$$\hat{\lambda}_3 = \frac{1100}{1000} = 1.1$$

We can then calculate the projected C_{ij} for the future. For $j > d - i$:

$$C_{ij} = C_{i,(d-i)}\lambda_{d-1+1}\cdots\lambda_j$$

We can then find Y_{ij} by subtraction. □

This method projects forward using an implicit inflation rate embodied in the λ_j.

The Inflation Adjustment Chain-Ladder Technique

We now assume that the expected value in cell (i, j) is

$$\mathbb{E}(Y_{ij}) = n_i r_j t_{i+j}$$

Where t_{i+j} is the assumed index of claims inflation from year to year. In other words, the inflation from calendar year s to $s + 1$ is $v_s = t_{s+1}/t_s$.

Values from calendar year $d - k$ can be "converted" to calendar year d money, by multiplying by $v_{d-k}\cdots v_{d-1} = t_d / t_{d-k}$ – let the resulting values be $\tilde{Y}_{ij}$. We can then apply the chain-ladder technique to these to obtain projections, in calendar year d money. We can then project even further using future inflation values.