Bahnemann Ch. 5 – Excess Claims

Chapter 5 focuses on claims excess of some fixed positive amount. We will be introducing a number of formulas and notation in this chapter. Fortunately, much of this has already been covered in some degree on earlier exams.

I. Excess Claim Size

Let X and a be an unlimited claim-size random variable and non-negative constant, respectively. Then, the random variable Y defined by

$$Y = \begin{cases} 0, & 0 \leq X \leq a \\ X - a, & X > a \end{cases}$$

represents size of claims modified by a limit amount a.

The distribution function of Y is obtained from the distribution function of X as follows:

$$F_Y(y) = Pr(Y \leq y) = \begin{cases} 0, & y < 0 \\ F_X(y + a), & y \geq 0 \end{cases}$$

Assuming $E[X]$ exists, it can be shown that $E[Y]$ equals the following:

$$E[Y] = E[X] - E[X; a]$$

where $E[X; a] = E[min(X, a)]$ is the limited expected value of X at a limit amount a.

Now, based on the definition of Y, the insurer pays nothing if $X \leq a$. Thus, insurers tend to focus on X_a, defined only for $X > a$:

$$X_a = X - a$$

X_a represents the **excess of X over the limit a.** We can think of X_a as the variable X **truncated from below and shifted by a.**

The distribution function of X_a is obtained from the distribution function of X as follows:

$$F_{X_a}(x) = Pr(X - a \leq x | X > a) = \begin{cases} 0, & x < 0 \\ \frac{F_X(x+a) - F_X(a)}{1 - F_X(a)}, & x \geq 0 \end{cases}$$

Assuming $E[X]$ exists, it can be shown that $E[X_a]$ equals the following:

$$E[X_a] = \frac{E[X] - E[X; a]}{1 - F_X(a)}$$

We will revisit $E[X_a]$ in the next section.

We can calculate the **limited expected value of X_a at a limit amount l** as follows:

$$E[X_a; l] = \frac{E[X; a+l] - E[X; a]}{1 - F_X(a)}$$

Next, Bahnemann introduces two familiar claim-size distributions for X:

1) Exponential distribution
2) Pareto distribution

Exponential Distribution

We can summarize the key formulas for the exponential distribution with parameter β as follows (assume $x \geq 0$):

- $F_X(x) = 1 - e^{\frac{-x}{\beta}}$
- $E[X] = \beta$
- $Var(X) = \beta^2$

- $E[X; x] = \beta \left(1 - e^{\frac{-x}{\beta}}\right)$

- $E[X_d] = E[X] = \beta$

Pareto Distribution

We can summarize the key formulas for the pareto distribution with parameters α and β as follows (assume $x > 0$):

- $F_X(x) = 1 - \left(\frac{\beta}{x+\beta}\right)^{\alpha}$

- $E[X] = \frac{\beta}{\alpha-1}$

- $Var(X) = \frac{\alpha\beta^2}{(\alpha-1)^2(\alpha-2)}$

- $E[X; x] = \frac{\beta}{\alpha-1}\left[1 - \left(\frac{\beta}{x+\beta}\right)^{\alpha-1}\right]$

- $E[X_d] = \frac{d+\beta}{\alpha-1}$

The last thing Bahnemann covers in this section is **how to use the chi-squared statistic to fit a distribution**.

Example: Fitting a Lognormal Distribution

Given the following grouped claim-size data from a sample of 300 claims from an unlimited population with an unknown distribution:

Size Group	# Claims
$0 - 5,000$	139
$5,001 - 10,000$	68
$10,001 - 15,000$	32
$15,001 - 20,000$	15
$20,001 - 25,000$	11
$25,001 - 30,000$	8
$30,001 - 35,000$	5

35,001 – 40,000	4
40,001 – 45,000	3
45,001 – 48,500	15
Total	**300**

Before the data above was tabulated, the claims were **censored by a \$50,000 policy limit** and then **subjected to a \$1,500 straight deductible**. Notice that the policy limit was applied first! Then, the deductible was applied to the capped claims.

Our goal is to use the minimum chi-square approach to fit a lognormal distribution for the population of the **unlimited** (i.e., non-truncated and non-censored) claims.

Let's define a few things:

- $F_{(\mu,\sigma)}(x)$ is the $Pr(X \leq x)$ for a lognormal distribution with parameters μ and σ

- $c_k = 5000k$ represents our size group boundaries, where $k = 0,1,\ldots,9$ and $c_{10} = \infty$

- n_k is the number of claims in each size group (ex. $n_{10} = 15$)

The probability $P_k(\mu, \sigma)$ of a claim being less than or equal to c_k is as follows:

$$P_k(\mu,\sigma) = \begin{cases} \dfrac{F_{\mu,\sigma}(c_k + 1,500) - F_{\{\mu,\sigma\}}(1,500)}{1 - F_{\mu,\sigma}(1,500)}, & k = 0,1,\ldots,9 \\ 1, & k = 10 \end{cases}$$

Notice that we must divide by the probability of being greater than or equal to 1,500 since the data is truncated from below by the \$1,500 deductible.

Once we define $P_k(\mu, \sigma)$, we can define the expected number of claims for each size group as follows:

$$\phi_k(\mu,\sigma) = (300)\big(P_k(\mu,\sigma) - P_{k-1}(\mu,\sigma)\big)$$

In the formula above, we multiply the total number of claims by the probability of being in a size group. This produces the expected number of claims in each size group.

The chi-square statistic is defined as follows:

$$\chi^2(\mu, \sigma) = \sum_{k=1}^{10} \frac{(actual - expected)^2}{expected}$$

$$= \sum_{k=1}^{10} \frac{\left(n_k - \phi_k(\mu, \sigma)\right)^2}{\phi_k(\mu, \sigma)}$$

The parameters (μ, σ) that minimize the chi-square statistic above are $(\hat{\mu}, \hat{\sigma}) = (8.67593, 1.18109)$.

Here is the calculation of the final chi-square statistic using the estimated parameters:

Size Group	# Claims (i.e., n_k)	$P_k(\mu, \sigma)$	$\phi_k(\mu, \sigma)$	$\dfrac{\left(n_k - \phi_k(\mu, \sigma)\right)^2}{\phi_k(\mu, \sigma)}$
$0 - 5{,}000$	139	0.4690	140.69	0.0202
$5{,}001 - 10{,}000$	68	0.6756	62.00	0.5808
$10{,}001 - 15{,}000$	32	0.7826	32.09	0.0003
$15{,}001 - 20{,}000$	15	0.8452	18.79	0.7644
$20{,}001 - 25{,}000$	11	0.8850	11.94	0.07377
$25{,}001 - 30{,}000$	8	0.9118	8.04	0.0002
$30{,}001 - 35{,}000$	5	0.9307	5.65	0.0755
$35{,}001 - 40{,}000$	4	0.9444	4.11	0.0031
$40{,}001 - 45{,}000$	3	0.9546	3.08	0.0019
$45{,}001 - 48{,}500$	15	1.0000	13.61	0.1410
Total	**300**		**300**	**1.6610**

In the table above, the $P_k(\mu, \sigma)$ terms are found using the LOGNORM.DIST() function in Excel. For example:

$$P_1(\mu, \sigma) = \frac{LOGNORM.DIST(5{,}000+15{,}000,8.67593,1.18109) - LOGNORM.DIST(1500,8.67593,1.18109)}{1 - LOGNORM.DIST(1500,8.67593,1.18109)} = 0.4690$$

The $\phi_k(\mu, \sigma)$ terms are found using the formula above. For example, $\phi_2(\mu, \sigma) = (300)(0.6756 - 0.4690) = 62.00$.

To actually solve for the estimated parameters $(\hat{\mu}, \hat{\sigma})$ in Excel, we start with an initial parameter set and use a Solver Add-In to find the parameters that minimize the sum of the $\frac{(n_k - \phi_k(\mu, \sigma))^2}{\phi_k(\mu, \sigma)}$ column. Now, **on the exam,** it's unlikely that you would be asked to fit a two-parameter distribution given that a Pearson VUE does not have a solver routine. But it's still important to understand the example above conceptually.

Before we move on, suppose we wanted to **estimate the number of claims eliminated by the deductible.** Since we estimated the parameters for the unlimited claim size distribution, we can estimate the number of claims eliminated by the deductible as follows:

- \# Population Claims $\leq 1{,}500 \approx \frac{300}{1 - Pr(X \leq 1{,}500)} Pr(X \leq 1{,}500) =$

  $\frac{300}{1 - LOGNORM.DIST(1500, 8.67593, 1.18109)} \left(LOGNORM.DIST(1500, 8.67593, 1.18109) \right) = 43$ claims

The $\frac{300}{1 - Pr(X \leq 1{,}500)}$ portion of the calculation is estimating the size of the sample before claims were eliminated by the deductible. Then, we multiply that number by $Pr(X \leq 1{,}500)$ to determine the number of claims eliminated by the deductible.

II. Excess Severity

We defined $E[X_a]$ mathematically in the previous section. $E[X_a]$ is known as the **mean excess claim size at a** or **excess severity at a**. If we replace the limit "a" with the limit "x," we end up with $E[X_x]$. This is often denoted as $e_X(x)$:

$$e_X(x) = \frac{E[X] - E[X; x]}{1 - F_X(x)}$$

Then, we can graph $e_X(x)$ for various parametric families and study their characteristics. Here is a plot of $e_X(x)$ for a number of common distributions:

Excess Severity Function Graphs
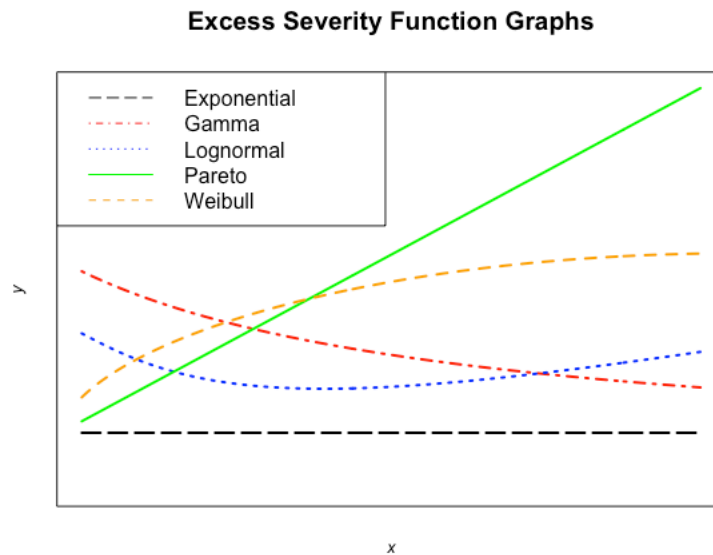
We can summarize the excess severity function graphs as follows:

- Exponential – $e_X(x)$ is a constant function of x
- Gamma – $e_X(x)$ decreases toward a horizontal asymptote as $x \to \infty$
- Lognormal – $e_X(x)$ increases without bound as $x \to \infty$
- Pareto – $e_X(x)$ is an increasing linear function of x (for $\alpha > 1$)
- Weibull – $e_X(x)$ is increases with x at a decreasing rate

The plots above are **useful when fitting distributions to empirical data**:

- If the empirical excess severity function is approximately linear with a positive slope, then this suggests a Pareto distribution should be used for the data
- If the empirical excess severity function is nearly constant for **large x**, then this suggests either a gamma or exponential distribution should be used for the data
- If the empirical excess severity lies in between the two extremes above, then this suggests either a lognormal or Weibull distribution should be used for the data

One key thing to keep in mind is that using empirical excess severity plots to inform the distribution choice only works **asymptotically** (i.e., large values of x). Thus, it is imperative that the claim data contain enough large values of x so that the empirical excess severity function can be reliably calculated.

Example: Fitting a Distribution Using Empirical Excess Severity Plots

Given the following grouped sample claim data for $n = 1,000$ claims:

Size Group	# Claims	Total Loss	Severity
0 – 100	100	6,000	60
101 – 500	300	95,000	317
501 – 1,000	240	145,000	604
1,001 – 2,000	185	260,000	1,405
2,001 – 3,000	140	450,000	3,214
4,001 – 5,000	15	66,000	4,400
5,001 – 10,000	20	150,000	7,500
Total	1,000	1,172,000	1,172

Using the data above, let's construct the empirical excess severity function and use it to select a distribution.

First, let's compute the empirical excess severity for each size group:

Size x	$F_n(x)$	$E_n[\hat{X}; x]$	$e_n(x)$
0	0	0	1,172
100	0.100	96	1,196
500	0.400	401	1,285
1,000	0.640	606	1,572
2,000	0.825	856	1,806
4,000	0.965	1,096	2,171
5,000	0.980	1,122	2,500
10,000	1.000	1,172	-

Let's discuss the table above:

- The first column represents the endpoint of the size group. So, the "Size 500" row represents the size group $500 - 1,001$

- The $F_n(x)$ column represents the empirical CDF. For example, $F_{1,000}(500) = \frac{100+300}{1,000} = 0.400$

- The $E_{1,000}[\hat{X}; x]$ represents the empirical limited expected value. For example, $E_{1,000}[\hat{X}; 500] = \frac{6,000+95,000+500(1,000-100-300)}{1,000} = 401$. Notice that we cap all claims greater than 500 at 500

- The $e_n(x)$ represents the empirical excess severity function. For example, $e_{1,000}(500) = \frac{E_{1,000}[\hat{X}]-E_{1,000}[\hat{X};500]}{1-F_{1,000}(500)} = \frac{1,172-401}{1-0.400} = 1,285$

Second, let's plot the empirical excess severity function and fit a trendline:
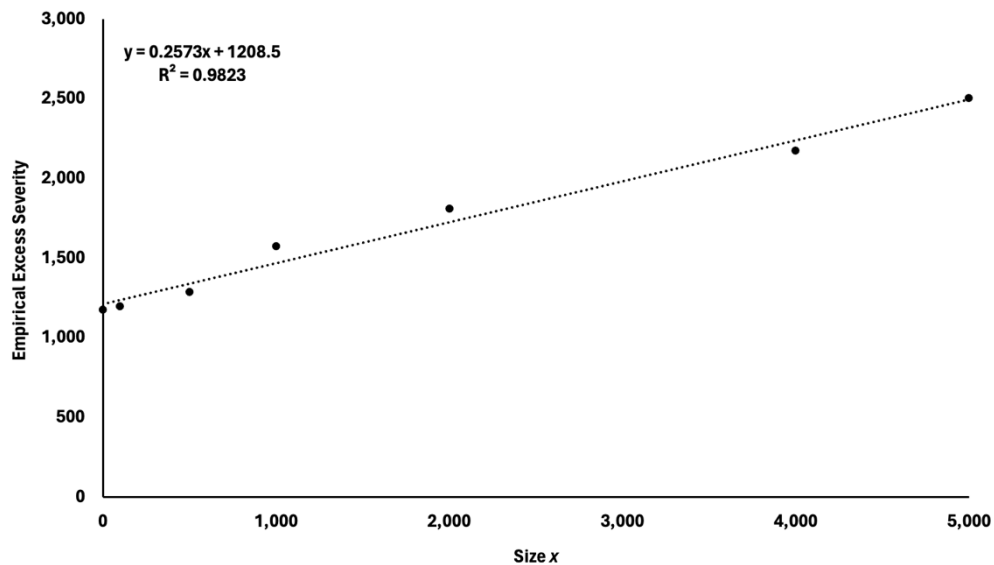
As we can see from the R^2 value, the trendline fits the data well. Since the trendline is increasing as x increases, it suggests that a Pareto distribution would be a good distribution to model the loss data.

Third, let's fit a Pareto distribution to the data using the trendline:

- The excess severity function of a Pareto distribution is $\frac{x+\beta}{\alpha-1} = \frac{1}{\alpha-1}x + \frac{\beta}{\alpha-1}$

- The expanded equation of the trendline above is $0.257335x + 1{,}208.50$. Thus, $\frac{1}{\alpha-1} = 0.257335$, which means $\alpha = 4.88599$. We also know that $\frac{\beta}{\alpha-1} = \frac{\beta}{4.88599-1} = 1{,}208.50$. Thus, $\beta = 4{,}696.22$

- The parameters of the fitted Pareto distribution are $(\alpha, \beta) = (4.88599, 4{,}696.22)$

Fourth, let's compute the excess severity values using the Pareto distribution and compare them to the empirical excess severity values:

Size x	$e_n(x)$	$e_X(x)$
0	1,172	1,208
100	1,196	1,234
500	1,285	1,337
1,000	1,572	1,466
2,000	1,806	1,723
4,000	2,171	2,238
5,000	2,500	2,495
10,000	-	3,782

In the table above, the $e_X(x)$ is calculated using the Pareto excess severity formula of $\frac{x+\beta}{\alpha-1} = \frac{x+4{,}696.22}{4.88599-1}$. As we can see, the Pareto values are very similar to the empirical values.

Caution! Be careful when using a regression line to estimate the parameters of Pareto distribution. The **slope of the regression line** is sensitive to the size of the largest claims, so the calculated distribution parameters could be significantly impacted by changes in just a few of the larger claim sizes.

III. Layers of Coverage

Consider an excess policy written over underlying coverage (ex. umbrella, excess liability). Then, the policy is subject to a lower limit a and an upper limit l. The ground-up loss x is **first decreased by a** and then **limited by l**. These claims belong the **layer of coverage** defined by a and l, where a is the lower limit or **attachment point** and l is the upper limit or **width** of the layer. The layer defined by a and l is sometimes **denoted by the "interval" notation $(a, a + l]$.**

Things change slightly when considering a policy with a **deductible a** and a **policy limit l.** Since the deductible is **typically applied AFTER the policy limit**, the insured layer is simply $(a, l]$. The deductible effectively reduces the policy limit from l to $l - a$.

Example: Assigning Losses to Layers

Consider two insurance policies:

1) Personal auto policy a $3,000 policy limit and a $100 straight deductible
2) Excess policy with a $100 attachment point and $3,000 limit

Suppose each policy incurs the following four claims during the policy period: $50, $600, $1,800, and $4,000

First, let's assign losses to layers based on the personal auto policy:

Layer	Claim 1	Claim 2	Claim 3	Claim 4	Total
(0, 100]	50	100	100	100	350
(100, 3000]	**0**	**500**	**1,700**	**2,900**	**5,100**
(3000, ∞)	0	0	0	1,000	1,000
Total	50	600	1,800	4,000	6,450

Since this policy has a limit and a deductible, the insured layer is (100, 3000], which is the bolded row above. As shown in Claim 4, we cap the loss at the policy limit first (i.e., $4,000 loss capped at $3,000) and then we apply the deductible (i.e., $3,000 - $100 = $2,900). In summary,

of the $6,450 in ground-up losses, $350 is in the deductible layer, $5,100 is in the insured layer, and $1,000 is in the uninsured layer. This means the insured is responsible for $350 + #1,000 = $1,350 of the $6,450 in ground-up losses.

Second, let's assign losses to layers based on the excess policy:

Layer	Claim 1	Claim 2	Claim 3	Claim 4	Total
(0, 100]	50	100	100	100	350
(100, 3100]	**0**	**500**	**1,700**	**3,000**	**5,200**
(3100, ∞)	0	0	0	900	900
Total	50	600	1,800	4,000	6,450

Since this is an excess policy, the insured layer (100, 3100]. In this case, the insured loss is $5,200 instead of $5,100.

Now, let's look at some theoretical results. Let $X_{a,l}$ be the claim size random variable in the layer $(a, a + l]$. Then, $X_{a,l}$ is as defined as follows:

$$X_{a,l} = \begin{cases} X - a, & a < X \leq a + l \\ l, & X > a + l \end{cases}$$

The CDF of $X_{a,l}$ is defined as follows:

$$F_{X_{a,l}}(x) = \begin{cases} 0, & x < 0 \\ \dfrac{F_X(x + a) - F_X(a)}{1 - F_X(a)}, & 0 \leq x < l \\ 1, & x \geq l \end{cases}$$

The expected value and second moment of $X_{a,l}$ are defined as follows:

$$E[X_{a,l}] = E[X_a; l] = \frac{E[X; a + l] - E[X; a]}{1 - F_X(a)}$$

$$E[X_{a,l}^2] = E[X_a^2; l] = \frac{E[X_2; a + l] - E[X^2; a] - 2a(E[X; a + l] - E[X; a])}{1 - F_X(a)}$$

The expected value of $X_{a,l}$ is known as the **layer mean** for the layer that runs from a to $a + l$.

Recall that the **coefficient of variation** of a random variable X is defined as $CV(X) = \frac{SD(X)}{E[X]}$.

Imposing a lower or upper limit (or both) to claims results in a **reduced coefficient of variation** of the claim-size variable.

___

Example: Layer Means and CVs

A claim-size variable X is lognormally distributed with parameters $(\mu, \sigma) = (5.9809, 1.800)$. Given the following information at limits \$3,000 and \$8,000:

Limit l	$F_X(l)$	$E[X; l]$	$E[X^2; l]$
3,000	0.869761	891	1,853,050
8,000	0.952557	1,276	5,774,970
∞	1.000000	2,000	102,134,385

Consider three layers:

- Unlimited layer that runs from $(0, \infty)$
- Layer that runs from $(3000, \infty)$
- Layer that runs from $(3000, 8000)$

First, let's compute the layer mean and CV for the unlimited layer:

- Layer mean = $E[X] = \frac{E[X; \infty] - E[X; 0]}{1 - F_X(0)} = \frac{2,000 - 0}{1 - 0} = 2,000$

- $E[X^2] = E[X^2; \infty] = \frac{E[X^2; \infty] - E[X^2; 0] - 2(0)(E[X; \infty] - E[X; 0])}{1 - F_X(0)} = \frac{102,134,385 - 0}{1 - 0} = 102,134,385$

- $CV(X) = \frac{SD(X)}{E[X]} = \frac{\sqrt{E[X^2] - E[X]^2}}{E[X]} = \frac{\sqrt{102,134,385 - 2,000^2}}{2,000} = 4.9531$

Second, let's compute the layer mean and CV for the layer that runs from $(3000, \infty)$:

- Layer mean = $E[X_{3000}] = E[X; 3000] = \frac{E[X;\infty] - E[X;3000]}{1 - F_X(3000)} = \frac{2{,}000 - 891}{1 - 0.869761} = 8{,}515$

- $E[X^2_{3000}] = E[X^2; 3000] = \frac{E[X^2;\infty] - E[X^2;3000] - 2(3000)(E[X;\infty] - E[X;3000])}{1 - F_X(3000)} =$

  $\frac{102{,}134{,}385 - 1{,}853{,}050 - 2(3000)(2{,}000 - 891)}{1 - 0.869761} = 718{,}885{,}926$

- $CV(X_{3000}) = \frac{SD(X_{3000})}{E[X_{3000}]} = \frac{\sqrt{E[X_{3000}{}^2] - E[X_{3000}]^2}}{E[X_{3000}]} = \frac{\sqrt{718{,}885{,}926 - 8{,}515^2}}{8{,}515} = 2.9858$

Third, let's compute the layer mean and CV for the layer that runs from $(3000, 8000)$:

- Layer mean = $E[X_{3K,8K}] = E[X_{3K}; 8K] = \frac{E[X;8K] - E[X;3K]}{1 - F_X(3K)} = \frac{1{,}276 - 891}{1 - 0.869761} = 2{,}956$

- $E[X^2_{3K,8K}] = E[X^2_{3K}; 8K] = \frac{E[X^2;8K] - E[X^2;3K] - 2(3000)(E[X;8K] - E[X;3K])}{1 - F_X(3K)} =$

  $\frac{5{,}774{,}970 - 1{,}853{,}050 - 2(3000)(1{,}276 - 891)}{1 - 0.869761} = 12{,}376{,}584$

- $CV(X_{3K,8K}) = \frac{SD(X_{3K,8K})}{E[X_{3K,8K}]} = \frac{\sqrt{E[X_{3K,8K}{}^2] - E[X_{3K}]^2}}{E[X_{3K,8K}]} = \frac{\sqrt{12{,}376{,}584 - 2{,}956^2}}{2{,}956} = 0.6452$

As we see from the numbers above, the **CV decreases as we impose additional limits** on the claim size variable.

Another important thing noted in this section is the **difference between interval means and layer means.** Interval means simply divide the total losses for all claims in a certain size group by the number of total claims in a certain size group. Layer means are computed after truncating and censoring losses by the corresponding lower and upper limits. See the Cookbook Recipe "Claim Intervals vs. Layers" for an example of the difference.

IV. Excess Claim Counts

Given the following:

- N_a is the number of claims excess an underlying limit a
- N is the number of ground-up claims

Then, we define the probability mass function for N_a as follows:

$$Pr(N_a = n) = f_{N_a}(n)$$

$$= \sum_{k=n}^{\infty} Pr(n \text{ excess claims}|N = k\} \cdot Pr(N = k)$$

$$= \sum_{k=n}^{\infty} \binom{k}{n} p^n (1-p)^{k-n} f_N(k)$$

where $p = 1 - F_X(a)$ and $f_N(k)$ is the probability mass function for the number of ground-up claims N. The mass function for N_a above is derived from the fact that the number n of excess claims, given the occurrence of k ground-up claims, has a binomial distribution with parameters (k, p).

Bahnemann proves the following results for the expected value and variance of N_a:

$$\boxed{\begin{array}{l} E[N_a] = pE[N] \\ Var(N_a) = p^2 Var(N) + p(1-p)E[N] \end{array}}$$

The probability mass function of N_a can be used to prove out a couple of **special cases**:

- If $N \sim Pois(\lambda)$, then $N_a \sim Pois(p\lambda)$
- If $N \sim NegBin(\alpha, \nu)$, then $N_a \sim NegBin(\alpha, p\nu)$

Example: Excess Claim Counts

Suppose that the number of ground-up claims is Poisson distributed with $\lambda = 15$ and that the unlimited claim-size variable X is lognormally distributed with parameters $(\mu, \sigma) = (5.9809, 1.800)$. This is the same distribution from the prior example. Recall that $F_X(3000) = 0.869761$.

An insured has an excess policy with an attachment point of $3,000 and a limit of $8,000.

First, let's calculate the expected number of claims excess of $a = 3,000$:

- $F_X(3000) = 0.869761$
- $p = 1 - F_X(3000) = 1 - 0.869761 = 0.130239$
- $E[N] = 15$
- $E[N_{3000}] = pE[N] = (0.130239)(15) = 1.9536$
- Notice that the upper limit is not needed here. The number of claims in the layer $(3000, 8000)$ is equal to the number of claims in the layer $(3000, \infty)$. The only difference is that any claim greater than 8,000 is capped at 8,000

Second, let's calculate the coefficient of variation of the number of claims excess of $a = 3,000$:

- $Var(N_{3000}) = p^2 Var(N) + p(1-p)E(N) = 0.130239^2(15) + 0.130239(1 - 0.130239)(15) = 1.9536$. We could also use the fact that N_a is Poisson distributed to jump straight to the answer. The variance of a Poisson distribution equals the mean of a Poisson distribution. Hence, $Var(N_{3000}) = E[N_{3000}] = 1.9536$
- $CV(N_{3000}) = \frac{SD(N_{3000})}{E[N_{3000}]} = \frac{\sqrt{1.9536}}{1.9536} = 0.715$

The CV for N is $\frac{\sqrt{15}}{15} = 0.258$. Thus, the CV for the excess claims is larger than the CV for the ground-up claims.

V. Inflation Effects

Let r_X be a positive rate of inflation for ground-up claim sizes. Then, we define the severity inflation factor $\tau_X = 1 + r_X$.

The impact of inflation on **excess claim size** depends on whether there is a lower limit only OR a both a lower limit and an upper limit.

Trended Excess Claim Size – Lower Limit Only

The average excess claim size with a lower limit a **after trending** is as follows:

$$E[\tau_X X_a] = \frac{\tau_X E[X] - \tau_X E[X; a/\tau]}{1 - F_X(a/\tau_X)}$$

The **effective trend factor** $\tilde{\tau}_X$ is defined as follows:

$$\tilde{\tau}_X = \frac{E[\tau_X X_a]}{E[X_a]}$$
$$= \tau_X \cdot \frac{E[X] - E[X; a/\tau_X]}{E[X] - E[X; a]} \cdot \frac{1 - F_X(a)}{1 - F_X(a/\tau_X)}$$

Trended Excess Claim Size – Lower & Upper Limits

The average excess claim size with a lower limit a and upper limit l **after trending** is as follows:

$$E[\tau_X X_{a,l}] = \frac{\tau_X E[X; (a+l)/\tau_X] - \tau_X E[X; a/\tau_X]}{1 - F_X(a/\tau_X)}$$

The **effective trend factor** $\tilde{\tau}_X$ is defined as follows:

$$\tilde{\tau}_X = \frac{E[\tau_X X_{a,l}]}{E[X_{a,l}]}$$
$$= \tau_X \cdot \frac{E[X; (a+l)/\tau_X] - E[X; a/\tau_X]}{E[X; a+l] - E[X; a]} \cdot \frac{1 - F_X(a)}{1 - F_X(a/\tau_X)}$$

We can also analyze the **impact of inflation** on the **number of excess claims**.

Trended Excess Claim Counts

The expected number of claims after trending are equal to $1 - F_X(a/\tau_X)E[N]$. Then, the **effective trend factor** $\tilde{\tau}_N$ for the excess claim count **due solely to claim size inflation** is defined as follows:

$$\tilde{\tau}_N = \frac{1 - F_X(a/\tau_X)}{1 - F_X(a)}$$

We might expect, a positive severity inflation rate **generally increases the excess claim count**.

If there is an underlying frequency trend as well (denoted as τ_{Freq}), then the effective trend factor is $\tilde{\tau}_{N,Total} = \tau_{Freq} \cdot \tilde{\tau}_N$.

We can also study the aggregate excess losses S. Again, this depends on whether there is a lower limit only OR a both a lower limit and an upper limit.

Trended Aggregate Excess Losses – Lower Limit Only

The effective trend factor for aggregate excess losses $S = N_a X_a$ **due solely to claim size inflation** is as follows:

$$\tilde{\tau}_S = \tau_X \frac{E[X] - E[X; a/\tau_X]}{E[X] - E[X; a]}$$

If there is an underlying frequency trend as well, then the effective trend factor $\tilde{\tau}_{S,Total} = \tau_{Freq} \cdot \tilde{\tau}_S$.

Trended Aggregate Excess Losses – Lower & Upper Limits

The effective trend factor for aggregate excess losses $S = N_a X_{a,l}$ **due solely to claim size inflation** is as follows:

$$\tilde{\tau}_S = \tau_X \frac{E[X; (a + l)/\tau_X] - E[X; a/\tau_X]}{E[X; a + l] - E[X; a]}$$

If there is an underlying frequency trend as well, then the effective trend factor $\tilde{\tau}_{S,Total} = \tau_{Freq} \cdot \tilde{\tau}_S$.

There are a couple of key relationships between $\tilde{\tau}_S$ and τ_X:

- If $\tau_X > 1$, then $\tilde{\tau}_S \geq \tau_X$
- If $\tau_X < 1$, then $\tilde{\tau}_S \leq \tau_X$

Example: Impact of Inflation

Let X be a claim-size random variable that is Pareto distributed with parameters $(\alpha, \beta) = (2,3000)$. Suppose that the severity inflation rate for ground-up claims is $r_X = 10\%$ and that the frequency trend for ground-up claim counts is 5%.

An actuary wants to analyze the impact of inflation claims with an attachment point of $5,000 and an upper limit of $4,000.

First, let's calculate a number of important input values:

- $\tau_X = 1.10$

- $F_X(5000) = 1 - \left(\frac{\beta}{5000+\beta}\right)^\alpha = 1 - \left(\frac{3000}{5000+3000}\right)^2 = 0.8594$

- $F_X(5000/\tau_X) = 1 - \left(\frac{\beta}{5000/\tau_X+\beta}\right)^\alpha = 1 - \left(\frac{3000}{5000/1.10+3000}\right)^2 = 0.8419$

- $E[X; 5000] = \frac{\beta}{\alpha-1}\left[1 - \left(\frac{\beta}{5000+\beta}\right)^{\alpha-1}\right] = \frac{3000}{2-1}\left[1 - \left(\frac{3000}{5000+3000}\right)^{2-1}\right] = 1,875$

- $E[X; 5000/\tau_X] = \frac{\beta}{\alpha-1}\left[1 - \left(\frac{\beta}{5000/\tau_X+\beta}\right)^{\alpha-1}\right] = \frac{3000}{2-1}\left[1 - \left(\frac{3000}{5000/1.10+3000}\right)^{2-1}\right] = 1,807$

- $E[X; 9000] = \frac{\beta}{\alpha-1}\left[1 - \left(\frac{\beta}{9000+\beta}\right)^{\alpha-1}\right] = \frac{3000}{2-1}\left[1 - \left(\frac{3000}{9000+3000}\right)^{2-1}\right] = 2,250$

- $E[X; 9000/\tau_X] = \frac{\beta}{\alpha-1}\left[1 - \left(\frac{\beta}{9000/\tau_X+\beta}\right)^{\alpha-1}\right] = \frac{3000}{2-1}\left[1 - \left(\frac{3000}{9000/1.10+3000}\right)^{2-1}\right] = 2,195$

Second, let's study the impact of inflation on claim size:

- $E\left[X_{5K,9K}\right] = \frac{E[X;9K]-E[X;5K]}{1-F_X(5K/\tau_X)} = \frac{1.10(2,250)-1.10(1,875)}{1-0.8594} = 2,667$. This is the average untrended claim size in the layer that runs from (5000, 9000]

- $E\left[\tau_X X_{5K,9K}\right] = \frac{\tau_X E[X;9K/\tau_X]-\tau_X E[X;5K/\tau_X]}{1-F_X(5K/\tau_X)} = \frac{1.10(2,195)-1.10(1,807)}{1-0.8419} = 2,699$. This is the average trended claim size in the layer that runs from (5000, 9000]

- The effective severity trend factor $\tilde{\tau}_X = \frac{E[\tau_X X_{5K,9K}]}{E[X_{5K,9K}]} = \frac{2,699}{2,667} = 1.0122$. Hence, the effective severity trend for this layer is 1.22%

Third, let' study the impact of inflation on excess claim counts:

- $\tilde{\tau}_N = \frac{1-F_X(5000/\tau_X)}{1-F_X(5000)} = \frac{1-F_X(5000/1.10)}{1-F_X(5000)} = \frac{1-0.8419}{0.8594} = 1.1241$. This is the effective trend factor for claim counts in this layer solely due to claim size inflation

- $\tilde{\tau}_{N,Total} = \tau_{Freq} \cdot \tilde{\tau}_N = (1.05)(1.1241) = 1.1803$. Hence, the effective trend for claim counts in this layer due to claim size inflation and frequency trend is 18.03%

Fourth, let's study the impact the impact of inflation on aggregate excess losses:

- $\tilde{\tau}_S = \tau_X \frac{E[X;(9K)/\tau_X]-E[X;5K/\tau_X]}{E[X;9K]-E[X;5K]} = (1.10)\frac{2,195-1,807}{2,250-1,875} = 1.1378$. This is the effective trend factor for aggregate excess losses in this layer solely due to claim size inflation. Alternatively, we could have calculated this as $\tilde{\tau}_S = \tilde{\tau}_X \cdot \tilde{\tau}_N = (1.0122)(1.1241) = 1.1378$

- $\tilde{\tau}_{S,Total} = \tau_{Freq} \cdot \tilde{\tau}_S = (1.05)(1.1378) = 1.1947$. This is the effective trend factor for aggregate excess losses in this layer due to claim size inflation and frequency trend. Alternatively, we could have calculated this as $\tilde{\tau}_{S,Total} = \tilde{\tau}_X \cdot \tilde{\tau}_{N,Total} = (1.0122)(1.1803) = 1.1947$. Hence, the effective trend for aggregate excess losses in this layer due to claim size inflation and frequency trend is 19.47%

VI. Aggregate Layer Claims

Suppose that the ground-up claim counts N has mean $E[N] = \lambda$ and **contagion parameter γ** such that $Var(N) = \lambda + \gamma\lambda^2$ (the contagion parameter reflects the fact the occurrence of a claim often leads to additional claims). Then, the layer mean and variance for the aggregate excess claims $S = N_a X_{a,l}$ are as follows:

$$E[S] = \lambda(E[X; a + l] - E[X; a])$$
$$Var(s) = \lambda(E[X^2; a + l] - E[X^2; a]) - 2aE[S] + \gamma(E[S])^2$$

Example: Aggregate Excess Claims

Suppose that N is Poisson distributed with mean $\lambda = 15$ and contagion parameter $\gamma = 0.05$. Further suppose that the claim-size variable X is lognormally distributed with parameters $(\mu, \sigma) = (5.9809, 1.800)$.

Given the following information at limits $3,000 and $8,000:

Limit l	$F_X(l)$	$E[X; l]$	$E[X^2; l]$
3,000	0.869761	891	1,853,050
8,000	0.952557	1,276	5,774,970
∞	1.000000	2,000	102,134,385

First, let's calculate the layer mean for S for the layer than runs from (3000,8000):

- $E[S] = \lambda(E[X; a + l] - E[X; a]) = 15(E[X; 8,000] - E[X; 3,000]) = 15(1,276 - 891) = 5,775$
- The expected value of the aggregate excess loss in the layer $5,000 excess of $3,000 is $5,775

Second, let's calculate the variance for S for the layer than runs from $(3000,8000)$:

- $Var(s) = \lambda(E[X^2; a+l] - E[X^2; a]) - 2aE[S] + \gamma(E[S])^2 = 15(E[X^2; 8,000] - E[X^2; 3,000]) - 2(3,000)E[S] + (0.05)(E[S])^2 = 15(5,774,970 - 1,853,050) - 2(3,000)(5,775) + 0.05(5,775)^2 = 25,846,331$

The last thing that Bahnemann mentions in this section is that the cumulative distribution function for S may have large jump discontinuities at the lower end of its distribution due to the fact that the number of excess claim counts is often small (this is especially true for small portfolios). The large jumps make it more challenging to fit continuous distributions to the data. Fortunately, the tail of the CDF (i.e., the region of the CDF we care the most about) is typically more reasonable.