Introduction to Aggregate Loss Models

Chapter 9

Stat 346 - Short-term Actuarial Math

Introduction to Aggregate Loss Random Variable S

- Definition: Aggregate Loss Random Variable $S = X_1 + X_2 + \cdots + X_n$
- Represents the total loss over n policies or events. For now we assume that n is fixed.
- Finding the distribution of S is key in actuarial science.

Known Distributions of Sums of Random Variables

Common cases where the sum's distribution is known:

- Normal distribution: If $X_i \sim N(\mu, \sigma)$, then $\sum X_i \sim N(n\mu, \sqrt{n}\sigma)$.
- Gamma distribution: If $X_i \sim \text{Gamma}(\alpha, \beta)$, then $\sum X_i \sim \text{Gamma}(n\alpha, \beta)$.
- Poisson distribution: If $X_i \sim \text{Poisson}(\lambda)$, then $\sum X_i \sim \text{Poisson}(n\lambda)$.
- Geometric distribution: If $X_i \sim \text{Geometric}(\beta)$, then $\sum X_i \sim \text{Negative Binomial}(r, \beta)$.

Review of Moment Generating Functions (MGFs)

- MGF of a random variable X: $M_X(t) = E[e^{tX}]$.
- **Finding Moments**: The k-th moment is given by $E[X^k] = M_X^{(k)}(0)$, where $M_X^{(k)}(t)$ is the k-th derivative of $M_X(t)$.
- **Central Moments**: For central moments, use $E[(X - E[X])^k]$. For example, the variance (second central moment) is $M_X^{(2)}(0) - [M_X^{(1)}(0)]^2$.
- More commonly you. can find the k-th central moooment as the k-th derivative of the log of the moment. generating function at 0, $E[(X-\mu)^k] = \frac{d^k}{dt^k} \log\left(M_X(t)\right)|_{t=0}$
- Useful as a review of these properties for understanding how MGFs are applied in practice.

Using MGFs to Find Sums of Distributions

- MGFs greatly simplify finding the distribution of sums.
- The MGF of a sum of independent random variables is the product of their individual MGFs.
- If $M_{X_i}(t)$ is the moment generating function for X_i then the. moment generating function of $S = X_1 + X_2 + ... + X_n$ is

$$M_S(t) = \prod_{i=1}^{n} M_{X_i}(t) = M_X(t)^n$$

- This property enables us to determine the distribution type of the sum.

Proof: Sum of Exponentials is a Gamma Distribution

- Let X_i be Exponential with scale parameter θ (or rate $\lambda = 1/\theta$).
- MGF of X_i: $M_{X_i}(t) = \frac{1}{1-\theta t}$, for $t < 1/\theta$.
- For the sum $S = X_1 + \cdots + X_n$:

$$M_S(t) = \left(\frac{1}{1-\theta t}\right)^n$$

- This MGF corresponds to a Gamma distribution with shape parameter $\alpha = n$ and scale parameter θ.
- Hence, the sum of n independent Exponential variables each with scale θ is a Gamma(n, θ) distribution.

Convolution Method for $n = 2$

- When $n = 2$, the distribution of $S = X_1 + X_2$ can be found by convolution.
- For continuous variables: $f_S(s) = \int f_{X_1}(s - x) f_{X_2}(x) dx$
- For discrete variables: $p_S(s) = \sum p_{X_1}(s - x) p_{X_2}(x)$

Creating Probability Distribution for $S = X_1 + X_2$

Given X_i can be 100, 200, or 300 with probabilities 0.25, 0.5, and 0.25:

- Possible sums of X_1 and X_2 are 100, 200, and 300.
- Probability distribution of S:
  - $\Pr(S = 100) = 0.25$
  - $\Pr(S = 200) = 0.5$
  - $\Pr(S = 300) = 0.25$
- Specifically, $\Pr(S = 400) = 0.375$.
- This demonstrates the method to calculate and plot the probability distribution of S.

Full Probability Distribution for $S = X_1 + X_2$

Given X_i can be 100, 200, or 300 with probabilities 0.25, 0.5, and 0.25, the full probability distribution of S is:

Sum (S)	Probability
200	0.0625
300	0.25
400	0.375
500	0.25
600	0.0625

Using Software for Higher n

- For larger n, manual calculation becomes impractical.
- Statistical software can automate the process.

Individual risk model

- Consider a portfolio of n insurance policies.
- Denote the loss, for a fixed period, for each policy i by X_i, for $i = 1, \ldots, n$.
- Assume these lossses are independent and (possibly) identically distributed.
- The aggregate loss, S, is defined by the sum of these losses:

$$S = X_1 + X_2 + \cdots + X_n.$$

- It is possible that a policy does not incur a loss so that each X_i has a mixed distribution with a probability mass at zero.

Alternative representation

- Assume that the losses are also identically distributed say as X. Then we can write X as the product of a Bernoulli I and a positive (continuous) random variable Y:

$$X = IY$$

- $I = 1$ indicates there is a claim, otherwise $I = 0$ means no claim. Let $\Pr(I = 1) = q$ and hence $\Pr(X = 0) = 1 - q$.
- In addition, assume $\mathsf{E}(Y) = \mu_Y$ and $\mathsf{Var}(Y) = \sigma_Y^2$.
- Typically it is assumed that I and Y are independent so that

$$\mathsf{E}(X) = \mathsf{E}(I)\mathsf{E}(Y) = q\mu_Y$$

and

$$\mathsf{Var}(X) = q(1-q)\mu_Y^2 + q\sigma_Y^2.$$

Mean and variance in the individual risk model

- The mean of the aggregate loss can thus be written as

$$\mathsf{E}(S) = nq\mu_Y$$

and its variance is

$$\mathsf{Var}(S) = nq(1-q)\mu_Y^2 + nq\sigma_Y^2.$$

- **Example:** Consider a portfolio on 1,000 insurance policies where each policy has a probability of a claim of 0.15. When a claim occurs, the amount of claim has a Pareto distribution with parameters $\alpha = 3$ and $\theta = 100$. Calculate the mean and variance of the aggregate loss.

Approximating the individual risk model

- For large n, according to the Central Limit Theorem, S can be approximated with a Normal distribution.
- Then, probabilities can be computed using Normal as follows:

$$\begin{aligned} \Pr(S \leq s) &= \Pr\left[\frac{S - \mathsf{E}(S)}{\sqrt{\mathsf{Var}(S)}} \leq \frac{S - \mathsf{E}(S)}{\sqrt{\mathsf{Var}(S)}}\right] \\ &\approx \Pr\left[Z \leq \frac{S - \mathsf{E}(S)}{\sqrt{\mathsf{Var}(S)}}\right] \\ &= \Phi\left(\frac{S - \mathsf{E}(S)}{\sqrt{\mathsf{Var}(S)}}\right) \end{aligned}$$

Illustrative examples

- **Example 1:** An insurable event has a 10% probability of occurring and when it occurs, the amount of the loss is exactly 1,000. Market research has indicated that consumers will pay at most 115 for insuring this event. How many policies must a company sell in order to have a 95% chance of making money (ignoring expenses)? Assume Normal approximation.

Example 2

An insurer has a portfolio consisting of 25 one-year life insurance policies grouped as follows:

Insured amount b_k	Number of policies n_k
1	10
2	5
3	10

The probability of dying within one year is $q_k = 0.01$ for each insured, and the policies are independent.

The insurer sets up an initial capital of \$1 to cover its future obligations.

Using Normal approximation, calculate the probability that the insurer will be able to meet its financial obligation.

Collective risk model

- Let X_i be the claim payment made for the ith policyholder and let N be the random number of claims. The insurer's aggregate loss is

$$S = X_1 + \cdots + X_N = \sum_{i=1}^{N} X_i.$$

- This is called the Collective Risk Model.
- If N, X_1, X_2, ... are independent and the individual claims X_i are i.i.d., then S has a *compound distribution*.
- N: frequency of claims; X: the severity of claims.
- Central question is finding the probability distribution of S.

Properties of the collective risk model

- X is called the individual claim and assume has moments denoted by $\mu_k = \mathsf{E}(X^k)$.
- Mean of S: $\mathsf{E}(S) = \mathsf{E}(X)\mathsf{E}(N) = \mu_1 \mathsf{E}(N)$
- Variance of S: $\mathsf{Var}(S) = \mathsf{E}(N)\mathsf{Var}(X) + \mathsf{Var}(N)\mu_1^2$
- CDF of S: $\mathsf{Pr}(S \le s) = \sum_{n=0}^{\infty} \mathsf{Pr}(S \le s | N = n)\mathsf{Pr}(N = n)$

Illustrative example 1

Suppose that the number of claims follows the distribution in the following table:

n	Pr$(N = n)$
1	0.35
2	0.45
3	0.20

When there is a claim, the amount lost has an equal chance of being 100 or 300. Let S be the aggregate loss random variable for total losses. Find the probability that the total claims is less than 450.

Find the probability mass function of S, i.e. $Pr(S = s)$ for all possible values of S.

Illustrative example 2

Let $S = \sum_{i=1}^{N} X_i$ be a aggregate loss random variable where the number of claims, N, is a negative binomial random variable with $\beta = 4$ and $r = 2$ and the severity of each claim is a Lognormal random variable with $\mu = 4$ and $\sigma = 2$. What is the mean and variance of S?

Poisson number of claims

- If $N \sim \text{Poisson}(\lambda)$, so that λ is the average number of claims, then the resulting distribution of S is called a Compound Poisson.
- It can then be shown that:
  - Mean of S: $\mathsf{E}(S) = \lambda \mathsf{E}(X) = \lambda\mu_1$
  - Variance of S: $\mathsf{Var}(S) = \lambda \mathsf{E}(X^2) = \lambda\mu_2$

Illustrative example 3

Suppose S has a compound Poisson distribution with $\lambda = 0.8$ and individual claim amount distribution

x	$\Pr(X = x)$
1	0.250
2	0.375
3	0.375

What is $E(S)$ and $Var(S)$?

Use CLT to approximate the compound distribution

- If the average number of claims is large enough, we may use the Normal approximation to estimate the distribution of S.
- All you need are the mean and variance of the aggregate loss