ACTSC 431 — Loss Model I

Classnotes for Fall 2018

by

Johnson Ng

BMath (Hons), Pure Mathematics major, Actuarial Science Minor

University of Waterloo

Table of Contents

List of Definitions

List of Theorems

1 Lecture 1 Sep 06

1.1 Introduction and Overview

Course Objective In Loss Model I, the focus of our study is to learn the basic methods which are used by insurers to quantify risk from mathematical/statistical models, in order for insurers to make various decisions[1]. By quantifying risk, it helps us monitor underlying risks so that not only are we aware of them, but also so that we can take actions or preventive measures against them.

[1] e.g. setting premiums, control expenses, deciding for reinsurance, etc.

Our main interest of this course is:

- to quantify and seek protection against the loss of funds due either to **too many claims** or **a few large claims**;
- to reduce adverse financial impact of random events that prevent the realization of reasonable expectations.

The main model that shall be the focus of this course is **models for liability risk**.

Definition 1 (Liability Risk)

*A **liability risk** is a risk that insurance companies assume by selling insurance contracts.*

In particular, the liability that we shall focus on is **insurance claims**.

Many of the models that we shall see later in the course are also applied for other types of risks, e.g. investment risk, credit risk, liquidity risk, and operational risk.

We are interested in modelling the total amount of claims, i.e. the **aggregate claim amount**, of a group fo insurance policies over a given period of time. In the actuarial literature, there are two main approaches that have been proposed to model the aggrement claim amount of an insurance portfolio, namely:

- individual risk model;
- collective risk model.

1.1.1 Individual Risk Model

Definition 2 (Individual Risk Model)

*In an **individual risk model**, the aggregate claim is modeled by*

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

*where n is a **deterministic**[2] integer that represents the **total number of insurance policies**, and Z_i is a random variable for the **potential loss of the i^{th} insurance policy.***

[2] i.e. fixed

Note

Since a policy may or may not incur a loss[3], we have that

$$P(Z_i = 0) > 0.$$

Thus, in an individual risk model, we may also express the aggregate claim amount as

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

where I_i is the indicator function about the claimant of policy i, while X_i represents the size of the claim(s) for the i^{th} policy provided that there is a claim.[4]

[3] Since a claim may or may not be made!

[4] **This is actually incorrect, despite being in the recommended textbook. See Appendix A.1.**

However, in an individual risk model, according to Dhaene and Vyncke (2010)[5],

> *A third type of error that may arise when computing aggregate claims follows from the fact that the assumption of mutual independency of the individual claim amounts may be violated in practice.*

[5] Dhaene, J. and Vyncke, D. (2010). The individual risk model. https://www.researchgate.net/publication/228232062_The_Individual_Risk_Model

Due to complications such as this, the individual risk model will not be the focus of our studies.

1.1.2 Collective Risk Model

Definition 3 (Collective Risk Model)

*In a **collective risk model**, the aggregate claim is modeled by*

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

*where N is a non-negative integer-valued random variable that denotes **the number of claims among a given set of policies**, while X_i denotes the **size of the i^{th} policy.***

Note

In a collective risk model, we need to determine:

- *the distribution of the total number of claims for the entire portfolio, i.e. the distribution of N; and*
- *the distribution of the loss amount per claim, i.e. the distribution of X_i.*

In this course, the primary focus of our studies will be on **collective risk models**.

Terminologies To end today's lecture, the following terminologies are introduced:

Definition 4 (Severity Distribution)

*The **severity distribution** is the distribution of the loss amount of the amount paid by the insurer on a given loss/claim.*

Definition 5 (Frequency Distribution)

*The **frequency distribution** is the distributino fo the number of losses/claims paid by the insurer over a given period of time.*

Note

The frequency distribution is typically a discrete distribution.

Definition 6 (Aggrement Payment / Loss)

*The **aggregate payment (loss)** is the total amout of all claim payments (losses) over a given period of time.*

Note

There is a distinction between an aggregate payment and an aggregate loss, since an aggregate payment is "essentially" an aggregate loss after certain claim adjustments, such as deductibles, limits, and coinsurance.

2 Lecture 2 Sep 11th

2.1 Review of Probability Theory

Firstly, we shall review the definition of a random variable.

Definition 7 (Random Variable)

*Let Ω be a sample space and $\mathcal{F}$ its σ-algebra[1]. A **random variable** (rv) $X : \Omega \to (\Omega, \mathcal{F})$ is a function from a possible set of outcomes to a measurable space $(\Omega, \mathcal{F})$. Within the context of our interest, X is real-valued, i.e. $(\Omega, \mathcal{F}) = \mathbb{R}$.*

[1] For definitions of Ω and $\mathcal{F}$, see notes on STAT330.

2.1.1 Discrete Random Variables

Definition 8 (Discrete Random Variable)

*A **discrete random variable** (drv) is an rv X that takes only countable (finite) real values.*

Note

Let X be a drv.

- *The **probability mass function** (pmf) of X is: for $i \in \mathbb{N}$,*

$$p(x_i) = P(X = x_i)$$

- *The* ***cumulative distribution function*** *(cdf) of X is*

$$F(x) = P(X \leq x) = \sum_{x_i \leq x} p(x_i).$$

- *The kth* ***moment*** *of X is*[2]

$$E[X^k] = \sum_{i \in \mathbb{N}} x_i^k p(x_i)$$

if $E[X^k]$ is finite.

- *Some commonly seen/introduced discrete distributions are: Poisson, Binomial, Negative Binomial*

[2] This implicitly uses the Law of the Unconcious Statistician.

Example 2.1.1

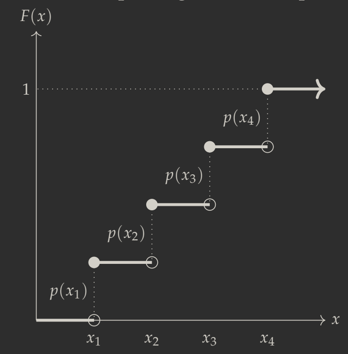
Let X take values from $\{x_1, x_2, x_3, x_4\}$, and

$$p(x_i) = P(X = x_i) \text{ for } i = 1, 2, 3, 4.$$

The cdf of X is

$$F(x) = \begin{cases} 0 & x < x_1 \\ p(x_1) & x_1 \leq x < x_2 \\ p(x_1) + p(x_2) & x_2 \leq x < x_3 \\ 1 - p(x_4) & x_3 \leq x < x_4 \\ 1 & x \geq x_4 \end{cases}$$

It is recommended to visualize the cdf first before putting it down in pencil.



Note

- *It is important that we stress the need for showing* ***right continuity*** *in the graph.*
- *Note that the cdf always sums to 1.*
- *The "**jumps**" at x_i correspond to $p(x_i)$, for $i = 1, 2, 3, 4$.*

Definition 9 (Probability Generating Function)

Suppose a drv X only takes ***non-negative integer values****. The* ***proba-***

bility generating function (pgf) of X is defined as

$$G(z) = E\left[z^X\right] = \sum_{k=1}^{\infty} z^k p(k)$$

where we note that if $\max X = n$, then $p(m) = 0$ for all $m > n$.

Note

- *The pgf uniquely identifies the distribution of the drv*[3].
- *To get the probability for $k \in \{0, 1, 2, \ldots\}$, we simply need to do*

$$p(k) = \frac{1}{k!} G^{(k)}(x)\Big|_{x=0}.$$

[3] This was given as is without proof, and I cannot find any resources that proves this.

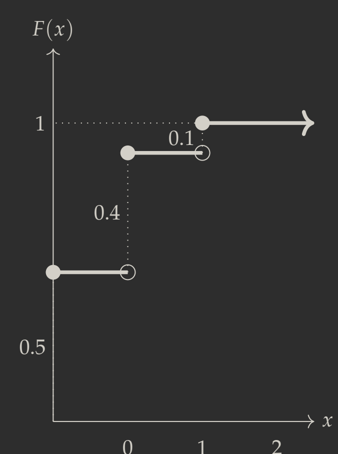
Example 2.1.2 (Lecture Slides: Example 1)

Consider a drv X with pmf

$$p(x) = P(X = x) = \begin{cases} 0.5 & x = 0 \\ 0.4 & x = 1 \\ 0.1 & x = 2 \end{cases}$$

Its cdf is

$$F(x) = P(X \leq x) \begin{cases} 0 & x < 0 \\ 0.5 & 0 \leq x < 1 \\ 0.9 & 1 \leq x < 2 \\ 1 & x \geq 2 \end{cases}$$



and its pgf is

$$G(z) = E\left[z^X\right] = 0.5 + 0.4z + 0.1z^2.$$

2.1.2 Continuous Random Variables

Definition 10 (Continuous Random Variable)

*A **continuous random variable** (crv) takes on a continuum of values.*

Note

Let X be a crv.

- *$\exists f : X \to \mathbb{R}$ called a* ***probability density function*** *(pdf) such that its cdf is*

$$F(x) = \int_{-\infty}^{x} f(y)\, dy,$$

and consequently by the ***Fundamental Theorem of Calculus****, we have*

$$f(x) = F'(x).$$

- *The kth moment of X is*

$$E[X^k] = \int_x x^k f(x)\, dx$$

so long that $E[X^k]$ is defined.

- *Some commonly introduced distributions are: Uniform, Exponential, Gamma, Weibull, and Normal.*

Definition 11 (Moment Generating Function)

Let X be an rv. The ***moment generating function*** *(mgf) of X is, for $t \in \mathbb{R}$ (appropriately so),*

$$M_X(t) = E\left[e^{tX}\right] = \int_x e^{tx} f(x)\, dx$$

provided that the integral is well-defined.

The mgf is also defined for drvs.

Note

- *The mgf uniquely determines the distribution of its rv*[4]
- *With the mgf, we can obtain the kth moment of an rv X by*

$$E\left[X^k\right] = \frac{d^k}{dt^k} M_X(t)\Big|_{t=0}$$

[4] This shall, also, not be proven in this course.

Example 2.1.3 (Lecture Notes: Example 2)

Consider an exponential rv X with pdf[5]

[5] When not explicitly stated, it shall be assumed that domains at which we did not specify x shall have probability 0.

$$f(x) = 0.1e^{-0.1x}, \; x > 0.$$

Its cdf is

$$F(x) = \int_{-\infty}^{x} f(y)\,dy = \begin{cases} 1 - e^{-0.1x} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

and its mgf is

$$\begin{aligned} M_X(t) = E\left[e^{tX}\right] &= \int_0^\infty e^{tx} 0.1 e^{-0.1x}\,dx \\ &= 0.1 \int_0^\infty e^{(t-0.1)x}\,dx \\ &= \frac{0.1}{0.1 - t}, \quad t < 0.1, \end{aligned}$$

where we note that we must have $t < 0.1$, for otherwise the value of the exponent would render the integral undefined.

Definition 12 (Hazard Rate Function)

For a crv X, the ***hazard rate function*** *(aka* ***failure rate****) of X is defined as*

$$h(x) = \frac{f(x)}{\overline{F}(x)} = -\frac{d}{dx} \ln \overline{F}(x),$$

where $\overline{F}(x) = 1 - F(x)$ is the ***survival function***[6]

[6] You should be familiar with this if you have studied for Exam P.

Note

- *We may also express the survival function in terms of the hazard rate by*

$$\overline{F}(x) = e^{-\int_{-\infty}^{x} h(y)\,dy}.$$

- *In terms of limits, we can express the hazard rate function, for small* *enough $\delta > 0$, as*

$$\begin{aligned}
h(x) = \frac{f(x)}{\bar{F}(x)} = \frac{F'(x)}{\bar{F}(x)} \\
&\approx \frac{F(x+\delta) - F(x)}{\delta \bar{F}(x)} \\
&= \frac{P(x < X \leq x + \delta)}{\delta F(X > x)} \\
&= \frac{1}{\delta} P(x < X \leq x + \delta \mid X > x).
\end{aligned}$$

We can make sense of this expression by recalling the notion of the probability of survival from Exam MLC[7]*, where if a life has survived over x, the hazard rate is the probability that the life does not survive beyond another δ*[8].

[7] This also tells us that the hazard rate gets its name from life insurance.

[8] From the perspective of life insurance, the greater the probability, the more likely the claim is going to happen.

3 Lecture 3 Sep 13th

3.1 Review of Probability Theory (Continued)

3.1.1 Continuous Random Variables (Continued)

Example 3.1.1 (Lecture Notes: Example 3 — Hazard Rate of Weibull Distribution)

Suppose $X \sim \text{Wei}(\theta, \tau)$ with pdf

$$f(x) = \frac{\tau \left(\frac{x}{\theta}\right)^{\tau} e^{-\left(\frac{x}{\theta}\right)^{\tau}}}{x}, \quad x > 0,$$

where $\theta, \tau > 0$. Find its hazard rate function.

Solution

We first require the survival function[1]:

[1] Weibull Survival Function

$$\begin{aligned} \bar{F}(x) &= \int_x^\infty \frac{1}{y} \tau \left(\frac{y}{\theta}\right)^{\tau} e^{-\left(\frac{y}{\theta}\right)^{\tau}} \, dy \\ &= \int_{\frac{x}{\theta}}^\infty \frac{1}{u} \tau u^{\tau} e^{-u^{\tau}} \, du \qquad \text{where } u = \frac{y}{\theta} \\ &= \int_{\frac{x}{\theta}}^\infty \tau u^{\tau-1} e^{-u^{\tau}} \, du \\ &= -e^{-u^{\tau}} \Big|_{\frac{x}{\theta}}^{\infty} = e^{-\left(\frac{x}{\theta}\right)^{\tau}} \end{aligned}$$

The hazard rate is therefore

$$h(x) = \frac{f(x)}{\bar{F}(x)} = \frac{\tau}{x} \left(\frac{x}{\theta}\right)^{\tau}$$

3.1.2 Mixed Random Variable

Definition 13 (Mixed Random Variable)

*We call X a **mixed random variable** (mixed rv) if it has both discrete and continuous components.*

Note

- *Mixed rvs are important in modeling insurance claims, e.g., the loss amount is usually a continuous random variable with a probability mass at 0.*

The following is a type of mixed random variable:

Definition 14 (Deductibles)

Let X be an rv and d be a fixed value.

$$[X-d]_+ = \begin{cases} X-d & x \geq d \\ 0 & otherwise \end{cases}$$

Note

If X be an rv and d a fixed value, the deductible $[X-d]_+$ has a mass point at 0 since

$$P([X-d]_+ = 0) = P(X < d) > 0$$

Note

*Let $\{x_1, x_2, \ldots\}$ be a sequence of real numbers in an increasing order. Suppose X is a rv that takes on values on the real, and has a **density function** f on each interval (x_i, x_{i+1}), and has **discrete mass points** at the boundaries of these intervals, i.e.*

$$P(X = x_i) = p(x_i) > 0 \quad i \in \mathbb{N}.$$

Since X is an rv, it must be the case that

$$\sum_{i\in\mathbb{N}} p(x_i) + \sum_{i\in\mathbb{N}} \int_{x_i}^{x_{i+1}} f(x)\, dx = 1.$$

In other words, we treat the discrete and continuous part of a mixed rv separately.

The cdf of a mixed rv X is

$$F(x) = P(X \leq x) = \sum_{i \in \mathbb{N}} p(x_i) \mathbb{1}_{\{x_i \leq x\}} + \sum_{i \in \mathbb{N}} \int_{x_i}^{x_{i+1}} f(y) \mathbb{1}_{\{y \leq x\}} \, dy.$$

The kth moment of X is

$$E\left[X^k\right] = \sum_{i \in \mathbb{N}} (x_i)^k p(x_i) + \sum_{i \in \mathbb{N}} \int_{x_i}^{x_{i+1}} x^k f(x) \, dx.$$

The mgf of X is

$$M_X(t) = E\left[e^{tX}\right] = \sum_{i \in \mathbb{N}} e^{tx_i} p(x_i) + \sum_{i \in \mathbb{N}} \int_{x_i}^{x_{i+1}} e^{tx} f(x) \, dx.$$

Example 3.1.2 (Lecture Notes: Example 4)

Assume a claim amount of an insurance policy is modeled by a non-negative rv X which has probability mass of p and 0, and otherwise continuous with a pdf f over $(0, \infty)$. Find its cdf, kth moment, and mgf.

Solution

The cdf of X is

$$F(x) = \begin{cases} p + \int_0^x f(y) \, dy & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

The kth moment of X is

$$E\left[X^k\right] = \int_0^\infty x^k f(x) \, dx.$$

The mgf of X is

$$M_X(t) = p + \int_0^\infty e^{tx} f(x) \, dx.$$

3.2 Distributional Quantities and Risk Measures

This chapter introduces us to some **distributional quantities** for a given rv X. These distributional quantities are informative values to describe the characteristics of a risk.

3.2.1 Distributional Quantities

Definition 15 (Central Moment)

*The **kth central moment** of an rv X is defined as*

$$E\left[(X - E(X))^k\right].$$

Note

*The second central moment is the **variance**. The square root of the variance is the **standard deviation**.*

Example 3.2.1 (Lecture Notes: Example 5)

Consider an rv $Y = \begin{cases} Y_1 & U = 1 \\ Y_2 & U = 2 \end{cases}$ [2], where $Y_1 = 0$, $Y_2 \sim \text{Exp}(10)$, and $P(U = 1) = P(U = 2) = 0.5$.

[2] This notation is just syntatic sugar for saying $Y_1 = Y \mid (U = 1)$ and $Y_2 = Y \mid (U = 2)$.

1. Find the cdf of Y.
2. Find the mean and variance of Y.
3. Let $Z = \frac{1}{2}Y_1 + \frac{1}{2}Y_2$. Does Z have the same distribution as Y? Answer this by solving the mean and variance of Z.

Solution

1. Note that

$$F(y) = P(Y_1 \leq y \mid U = 1)P(U = 1) + P(Y_2 \leq y \mid U = 2)P(U = 2).$$

Observe that

$$P(Y_1 \leq y \mid U = 1) = \begin{cases} 1 & y \geq 0 \\ 0 & y < 0 \end{cases}$$

and

$$P(Y_2 \leq y \mid U = 2) = \begin{cases} 1 - e^{-10y} & y \geq 0 \\ 0 & y < 0 \end{cases}$$

Therefore

$$F(y) = \begin{cases} 1 - \frac{1}{2}e^{-10y} & y \geq 0 \\ 0 & y < 0 \end{cases}$$

2. The mean of Y is

$$E(Y) = E(Y \mid U = 1)P(U = 1) + E(Y \mid U = 2)P(U = 2) = 10 \cdot \frac{1}{2} = 5.$$

To calculate the variance of Y, we require

$$\begin{aligned} E\left[Y^2\right] &= E\left[Y^2 \mid U = 1\right] P(U = 1) + E\left[Y^2 \mid U = 2\right] P(U = 2) \\ &= (\mathrm{Var}(Y_2) + E(Y_2)^2) \cdot \frac{1}{2} = 100. \end{aligned}$$

Therefore

$$\mathrm{Var}(Y) = 100 - 5^2 = 75.$$

3. The mean of Z is

$$E[Z] = E[\frac{1}{2}Y_1 + \frac{1}{2}Y_2] = 5.$$

The variance of Z is

$$\mathrm{Var}(Z) = \frac{1}{4}\mathrm{Var}(Y_1) + \frac{1}{4}\mathrm{Var}(Y_2) = 25.$$

Therefore, Z does not have the same distribution as Y.

Definition 16 (Quantiles)

*The $100p\%$ **quantile** (or **percentile**) of an rv X is a set π_p such that*

$$\pi_p = \{x \in X \mid P(X < x) \leq p \leq P(X \leq x)\}.$$

This definition may also be presented as: any number π_p such that

$$P(X < \pi_p) \leq p \leq P(X \leq \pi_p).$$

Note

- *If X is a continuous random variable, we have that $P(X < \pi_p) = P(X \leq \pi_p)$ and so we have to define the quantile as*

$$\pi_p = F^{-1}(p)$$

where F^{-1} is the inverse function of F, the cdf of X.

- *A quantile **can be a set of numbers**.*
- *$\pi_{0.5}$ is called the **median** of X.*

Graphical method to interpret this notion will be included.

Example 3.2.2 (Lecture Notes: Example 1)

Find the $100p\%$ quantile of the loss distribution $F(x) = 1 - e^{-\frac{x}{\theta}}$, $x > 0$.

Solution

Note that F is the cdf of an exponential distribution, which is a continuous distribution. Therefore,

$$F(\pi_p) = 1 - e^{-\frac{\pi_p}{\theta}} = p \implies \pi_p = -\theta \ln(1-p).$$

Example 3.2.3 (Lecture Notes: Example 2)

Find the median $\pi_{0.5}$ for the following cdf

$$F(x) = \begin{cases} 0 & x < 0 \\ 0.6 + 0.4(1 - e^{-\frac{x}{3}}) & x \geq 0 \end{cases}$$

Solution

Since $F(0) = 0.6$ and F is an increasing function, we have that $F(x) = 0$ for all $x < 0$. Therefore

$$\pi_{0.5} = 0.$$

Example 3.2.4 (Lecture Notes: Example 3)

Find the median $\pi_{0.5}$ for a loss X with pmf

$$p(0) = 0.25,\ p(1) = 0.25,\ p(2) = 0.5.$$

Solution

The cdf of X is

$$F(x) = \begin{cases} 0 & x < 0 \\ 0.25 & 0 \leq x < 1 \\ 0.5 & 1 \leq x < 2 \\ 1 & x \geq 2 \end{cases}$$

since $F(x) = 0.5$ when $1 \leq x < 2$, we have that

$$\pi_{0.5} = [1, 2].$$

4 Lecture 4 Sep 18th

4.1 Distributional Quantities and Risk Measures (Continued)

4.1.1 Risk Measures

Definition 17 (Risk Measure)

*A **risk measure** is a mapping from the loss rv to the real line $\mathbb{R}$.*

Klugman, Panjer & Wilmot (2012) [1] on risk measure:

> *The level of exposure to risk is often described by one number, or at least a small set of numbers. These numbers are necessarily functions of the model and are often called 'key risk indicators'. Such key risk indicators indicate to risk managers the degree to which the company is subject to particular aspects of risk.*

[1] Klugman, S. A., Panjer, H. H., and Willmot, G. E. (2012). *Loss Models: From Data to Decisions*. John Wiley & Sons, Inc., 4th edition

To ensure its solvency, insurers will have to charge on these risks, i.e. we have to **price these exposures to risks**.

Definition 18 (Premium Principle)

*A **premium principle** (or **insurance pricing**) is a rule for assigning a premium to an insurance risk.*

Note

The following are some of the common principles used by insurers:

- ***Expectation Principle***

$$\Pi(X) = (1+\theta)E(X), \quad \theta > 0$$

- ***Standard Deviation Principle***

$$\Pi(X) = E(X) + \theta\sqrt{\mathrm{Var}(X)}, \quad \theta > 0$$

- ***Dutch Principle***

$$\Pi(X) = E(X) + \theta E([X - E(X)]_{+}), \quad \theta > 0$$

One particular measure is known as the **Value-at-Risk** (VaR).

4.1.1.1 *Value-At-Risk*

Definition 19 (Value-at-Risk (VaR))

The ***Value-at-Risk (VaR)*** *is a* ***quantile*** *of the distribution of aggregate losses, i.e. the VaR of a risk X at the 100%p level is defined as*[2]

$$\begin{aligned} \pi_p = \mathrm{VaR}_p(X) &= \inf\{x \in \mathbb{R} : P(X > x) \leq 1 - p\} \\ &= \inf\{x \in \mathbb{R} : P(X \leq x) \geq p\}. \end{aligned}$$

[2] I must find out why we define using inf instead of min (see following remark), and I will not take "safe definition" as an answer without full justification.

Note

- VaR *is often called a* ***quantile risk measure****.*
- VaR *is the standard risk measure used to evaluate exposure to risks.*
- VaR *measures the amount of capital required by the insurer to remain solvent, with high certainty, in the face of large claims.*
- *In practice, p is generally high:* 99.95% *or as low as* 95%*.*

Remark
Observe that

$$B = \{x \in \mathbb{R} \mid F_X(x) \geq p\} = (A, \infty) \text{ or } [A, \infty)$$

This remark basically points out that the left endpoint of the interval B is always included, which should be quite clear by right-continuity of F.

for some $A \in \mathbb{R}$, since F is an increasing function. Now let $x_0 \in B$ such that

$$F(x_0) = P(X \leq x_0) \geq p \quad \wedge \quad F(x_0-) = P(X < x_0) \leq p,$$

i.e. it is not necessary that $P(X = x_0) = p$ (see the two example graphs on the margin).

Let $\{x_n\}_{n \in \mathbb{N}}$ be a decreasing sequence of points on $\mathbb{R}$ such that $x_n \to x_0$ as $n \to \infty$. Since F is right-continuous, we have that $F(x_n) \to F(x_0)$ as $n \to \infty$. Therefore,

$$B = [x_0, \infty)$$

This justifies the definition of π_p.

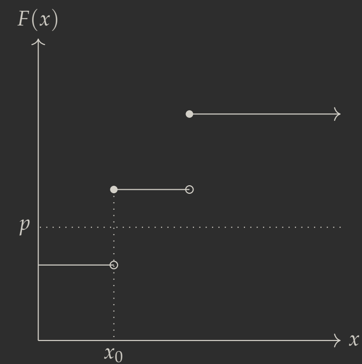


Figure 4.1: Discrete cdf

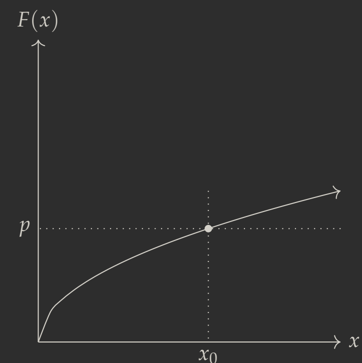


Figure 4.2: Continuous cdf

The lecturer asserts that we can really define VaR using min instead of inf, but even with this, I am not completely satisfied or convinced.

Note

- *Note that by definition, we have*

$$P(X < \pi_p) \leq p \leq P(X \leq \pi_p)$$

- *If X is a crv whose cdf is strictly increasing, i.e. no constant points, then*

$$\pi_p = F^{-1}(p)$$

  since $P(X < \pi_p) = P(X \leq \pi_p)$.

Warning (Shortcomings of VaR)

- VaR *cannot tell us the size of the potential loss in the $100(1-p)\%$ cases, making it difficult for us to prepare the right amount in order to safeguard against insolvency.*
- VaR *actually fails to satisfy properties to be a* ***coherent risk measure***[3]*, for example,* ***subadditivity***.
- VaR *is extensively used in financial risk management of trading risk over a fixed (usually short) time period, which are usually normally distributed, and* VaR *satisfies all coherency requirements.*
- *In insurance losses, instead of normal distributions, in general, skewed distributions are used, and in this cases,* VaR *is flawed as it lacks subadditivity.*

[3] See Appendix A.2.

Example 4.1.1

Suppose that X has a Pareto distribution with cdf

$$F(x) = 1 - \left(\frac{\theta}{x+\theta} \right)^{\alpha}, \quad x > 0$$

where $\alpha, \theta > 0$. Find $\mathrm{VaR}_p(X)$.

Solution

Since F is continuous and strictly increasing, we have that

$$\pi_p = F^{-1}(p) = \theta \left[(1-p)^{-\frac{1}{\alpha}} - 1 \right]$$

Example 4.1.2

Find $\mathrm{VaR}_{0.95}(X)$, $\mathrm{VaR}_{0.5}(X)$, and $\mathrm{VaR}_{0.3}(X)$ for a random loss with pmf

$$p(0) = 0.25,\ p(1) = 0.25, \text{ and } p(2) = 0.5.$$

Solution

Note that the cdf of X is

$$F(x) = \begin{cases} 0 & x < 0 \\ 0.25 & 0 \leq x < 1 \\ 0.5 & 1 \leq x < 2 \\ 1 & x \geq 2 \end{cases}.$$

Therefore,

$$\mathrm{VaR}_{0.95}(X) = 2,\ \mathrm{VaR}_{0.5}(X) = 1, \text{ and } \mathrm{VaR}_{0.3}(X) = 1.$$

4.1.1.2 *Tail-Value-at-Risk*

To compensate for the weakness of VaR at giving us the size of the loss X of which we cannot measure, we use the **Tail-Value-at-Risk**.

Definition 20 (Tail-Value-at-Risk (TVaR))

*Let X be an rv. The **Tail-Value-at-Risk (TVaR)** of X at the $100p\%$ level, denoted as $\mathrm{TVaR}_p(X)$, is defined as the average of all VaR values above the level p, and expressed as*

TVaR also has the following names, used by different regions:

- **Conditional Tail Expectation** (CTE) — NA
- **Tail Conditional Expectation** (TCE)
- **Expected Shortfall** (ES) — EU

$$\mathrm{TVaR}_p(X) = \frac{1}{1-p}\int_p^1 \mathrm{VaR}_\alpha(X)\,d\alpha = \frac{1}{1-p}\int_p^1 \pi_\alpha\,d\alpha$$

Remark

By considering the average of VaR *from* p*'s going up to* 1*, we take into account even the extreme cases of which* VaR *fails to account for.*

Perhaps a clearer definition would be the following, although the expression is only sensible if X is a crv:

Definition 21 (Tail-Value-at-Risk (TVaR))

Let X *be an rv. The* ***Tail-Value-at-Risk (TVAR)*** *of* X *at the* $100p\%$ *level, denoted* $\mathrm{TVaR}_p(X)$*, is the expected loss given that the loss exceeds the* $100p$ *percentile (or quantile) of the distribution of* X*, expressible as*

$$\mathrm{TVaR}_p(X) = E[X \mid X > \pi_p] = \frac{1}{\bar{F}(\pi_p)}\int_{\pi_p}^\infty x f(x)\,dx.$$

Note that the two definitions agree with one another:

$$\begin{aligned}\frac{1}{1-p}\int_p^1 \pi_\alpha\,d\alpha &= \frac{1}{1-F(\pi_p)}\int_p^1 F^{-1}(\alpha)\,d\alpha \\ &= \frac{1}{\bar{F}(\pi_p)}\int_{\pi_p}^1 x f(x)\,dx\end{aligned}$$

where we let $\alpha = F(x)$ as substitution.

Note

While it is not difficult to notice that

$$\mathrm{TVaR}_p(X) \geq \mathrm{VaR}_p(X),$$

the proof is also simple:

$$\begin{aligned}\mathrm{TVaR}_p(X) &= \frac{1}{1-p}\int_p^1 \pi_\alpha\,d\alpha \\ &\geq \frac{1}{1-p}\pi_p\int_p^1 d\alpha = \pi_p = \mathrm{VaR}_p(X).\end{aligned}$$

Example 4.1.3

Find $\mathrm{TVaR}_p(X)$ for $X \sim \mathrm{Exp}(\theta)$.

Solution

Since X is a crv, and $F(x) = 1 - e^{-\frac{x}{\theta}}$, we have that

$$\pi_p = F^{-1}(p) = -\theta \ln(1-p).$$

Therefore,

$$\begin{aligned}
\mathrm{TVaR}_p(X) &= \frac{1}{1-p} \int_p^1 \pi_\alpha \, d\alpha = \frac{-\theta}{1-p} \int_p^1 \ln(1-\alpha) \, d\alpha \\
&= \frac{-\theta}{1-p} \int_{-\infty}^{\ln(1-p)} u e^u \, du \quad \text{let } u = \ln(1-\alpha) \\
&= \frac{-\theta}{1-p} \left[u e^u \Big|_{-\infty}^{\ln(1-p)} - \int_{-\infty}^{\ln(1-p)} e^u \, du \right] \text{ by IBP} \\
&= \frac{-\theta}{1-p} \left[(1-p)\ln(1-p) - (1-p) \right] \\
&= \theta[1 - \ln(1-p)]
\end{aligned}$$

Note

From the last example, by the memoryless property of $\mathrm{Exp}(\theta)$, *notice that we may also do*

$$\begin{aligned}
\mathrm{TVaR}_p(X) &= E[X \mid X > \pi_p] = E[X - \pi_p + \pi_p \mid X > \pi_p] \\
&= E[X - \pi_p \mid X > \pi_p] + E[\pi_p \mid X > \pi_p] \qquad (4.1) \\
&= E[X] + \pi_p
\end{aligned}$$

5 Lecture 5 Sep 20th

5.1 Distrbutional Quantities and Risk Measures (Continued 2)

5.1.1 Risk Measures (Continued)

Before ending this section, we introduce a notion that is related to TVaR.

Definition 22 (Mean Excess Loss)

Let X be an rv, and $d \in \mathbb{R}$. The ***mean excess loss****, denoted $e_X(d)$, is defined as*

$$e_X(d) = E[X - d \mid X > d]$$

and $e_X(d) = 0$ for those d such that $P(X > d) = 0$.

Proposition 1 (Relation of $\mathrm{TVaR}_p(X)$ and $e_X(d)$)

For a crv X, we have

$$\mathrm{TVaR}_p(X) = e_X(\pi_p) + \mathrm{VaR}_p(X)$$

Proof

By Equation (4.1), we have that

$$\mathrm{TVaR}_p(X) = E[X - \pi_p \mid X > \pi_p] + \pi_p = e_X(\pi_p) + \pi_p.$$

$\square$

Proposition 2 (Expection from Survival Function)

Let X be a non-negative rv such that $E[X^k] < \infty$, for any $k \in \mathbb{N} \setminus \{0\}$. Then[1]

$$E\left[X^k\right] = k \int_0^\infty x^{k-1} \bar{F}(x)\, dx$$

[1] Note that this works for the discrete case as well, by replacing $\int$ with $\sum$.

Proof

Firstly, note that since $E[X^k] < \infty$ for all $k \in \mathbb{N} \setminus \{0\}$, we have that $\bar{F}(x)$ decays faster than x^k as $x \to \infty$. Now

$$\begin{aligned}
E\left[X^k\right] &= \int_0^\infty x^k f(x)\, dx \quad \because \text{ Law of the Unconscious Statistician} \\
&= \int_0^\infty x^k\, dF(x) \quad \because dF(x) = f(x)\, dx \\
&= -\int_0^\infty x^k\, d\bar{F}(x) \\
&= -\left[x^k \bar{F}(x) \Big|_0^\infty - \int_0^\infty k x^{k-1} \bar{F}(x)\, dx \right] \quad \because \text{ IBP} \\
&= k \int_0^\infty x^{k-1} \bar{F}(x)\, dx
\end{aligned}$$

□

Example 5.1.1

Calculate $e_X(d)$ and $\mathrm{TVaR}_p(X)$ for a Pareto distribution X with cdf

$$F(x) = 1 - \left(\frac{\theta}{x + \theta} \right)^\alpha, \quad x > 0,$$

where $\alpha > 1$ and $\theta > 0$.

Solution

Using Proposition 2,

$$\begin{aligned}
e_X(d) &= \int_0^\infty P(X-d>x \mid X>d)\,dx = \int_0^\infty \frac{P(X-d>x, X>d)}{P(X>d)}\,dx \\
&= \int_0^\infty \frac{P(X>x+d)}{P(X>d)}\,dx = \int_0^\infty \frac{\bar{F}(x+d)}{\bar{F}(d)}\,dx \\
&= \int_0^\infty \left(\frac{d+\theta}{x+d+\theta}\right)^\alpha dx = \frac{(d+\theta)^\alpha}{1-\alpha}\left(\frac{1}{x+d+\theta}\right)^{\alpha-1}\Big|_0^\infty \\
&= \frac{d+\theta}{\alpha-1}
\end{aligned}$$

By Example 4.1.1, we have

$$\pi_p = \theta\left[(1-p)^{-\frac{1}{\alpha}} - 1\right]$$

and so

$$\begin{aligned}
\mathrm{TVaR}_p(X) &= e_X(\pi_p) + \pi_p \\
&= \frac{\theta\left[(1-p)^{-\frac{1}{\alpha}}-1\right]+\theta}{\alpha-1} + \theta\left[(1-p)^{-\frac{1}{\alpha}}-1\right] \\
&= \frac{\theta(1-p)^{-\frac{1}{\alpha}}}{\alpha-1} + \frac{\theta(\alpha-1)(1-p)^{-\frac{1}{\alpha}}}{\alpha-1} - \theta \\
&= \frac{\theta\alpha(1-p)^{-\frac{1}{\alpha}}}{\alpha-1} - \theta
\end{aligned}$$

Proposition 3 (Expected Deductible)

We have

$$E([X-d]_+) = \int_d^\infty \bar{F}(x)\,dx$$

Proof

By the Law of the Unconscious Statistician and IBP on the last step,

$$E([X-d]_+) = \int_d^\infty (x-d)\,dF(x) = -\int_d^\infty (x-d)\,d\bar{F}(x) = \int_d^\infty \bar{F}(x)\,dx$$

$\square$

Proposition 4 (An Expression for Mean Excess Value)

If $\bar{F}(d) > 0$, we have

$$e_X(d) = \frac{\int_d^\infty \bar{F}(x)\,dx}{\bar{F}(d)}$$

Proof

Observe that by Proposition 3, we have

$$\begin{aligned} e_X(d) &= E[X - d \mid X > d] = \frac{E[(X-d)\mathbb{1}_{X>d}]}{P(X > d)} \\ &= \frac{E([X-d]_+)}{\bar{F}(d)} = \frac{\int_d^\infty \bar{F}(x)\,dx}{\bar{F}(d)} \end{aligned}$$

□

5.2 *Severity Distributions — Creating Severity Distributions*

Recall the definition of a severity distribution.

Definition (Severity Distribution)

A ***severity distribution*** *is a distribution used to describe single random losses in an insurance portfolio.*

When a loss occurs, the full amount of the loss is not necessarily the amount paid by the insurer, since an insurance policy typically involves some form of adjustment (e.g. **deductible, limit, coinsurance**). A distinction needs to be made between the actual loss prior to any of the adjustments (aka **ground-up loss**) and the amount ultimately paid by the insurer.

Our goal is to find a reasonable model for the **ground-up loss** rv X. The following are two desirable properties for X:

- $\mathrm{Im}(X) = \mathbb{R}_{>0}$, since losses are positive;
- pf of X is right-skewed, since we want the “tail” of the distribution to be not heavy.

- The motivation for this property is due to the **20-80 rule**: 20% of the largest claims accountn for 80% of the total claim amount.

THERE ARE two approaches to constructing a severity distribution:

- **Parametric approach**[2]: specify a "form" for the distribution with a finite number of parameters.
- Nonparametric approach: no form is specified; the distribution is constructed directly from the empirical data.

[2] This approach shall be the focus of this course.

A weakness of the **Nonparametric approach** is, if there is not enough data, such as in catasthropic risks, is becomes difficult to obtain reliable information. We shall look at one such example in this approach.

Definition 23 (Empirical Distribution Function)

Let $\{X_1, \ldots, X_n\}$ be an iid sample of a risk X. Then its ***empricial distribution function (edf)*** *is defined as*

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{X_i \leq x\}}, \quad x \in \mathbb{R}.$$

Remark

Simply put, the edf assigns a probability of $\frac{1}{n}$ to each sample point X_i.

Example 5.2.1

Consider a random sample of a risk with size 5: $\{30, 80, 150, 150, 200\}$. Find the edf of the risk.

Solution

The edf is given by

$$\hat{F}_n(x) = \frac{1}{5} \sum_{i=1}^{5} \mathbb{1}_{\{X_i \leq x_i\}} = \begin{cases} 0 & x < 30 \\ \frac{1}{5} & 30 \leq x < 80 \\ \frac{2}{5} & 80 \leq x < 150 \\ \frac{4}{5} & 150 \leq x < 200 \\ 1 & x \geq 200 \end{cases}$$

6 Lecture 6 Sep 25th

6.1 Severity Distributions — Creating Severity Distributions (Continued)

The Parametric Approach The following is a graph showing the process of a parametric approach:

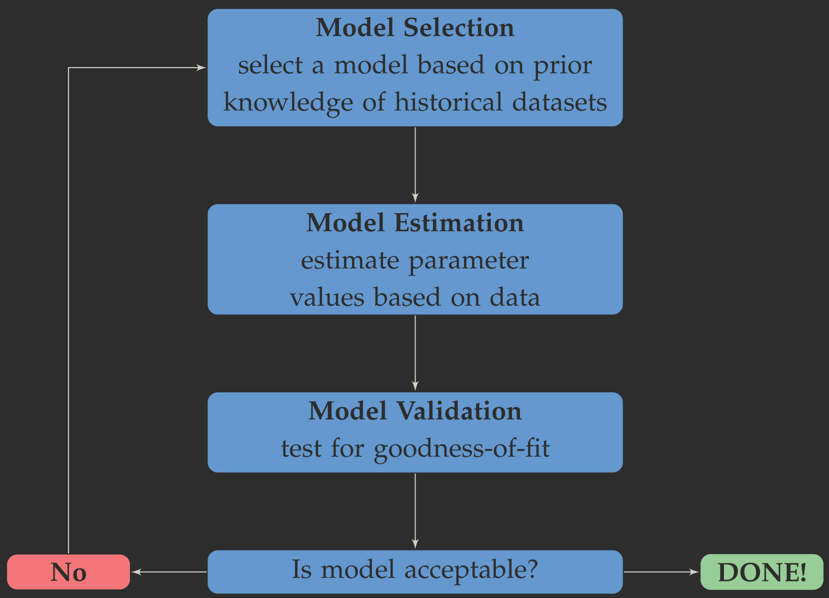


Figure 6.1: Process of a Parametric Approach

Common Techniques in Creating New Parametric Distributions Before diving into the topic, first, a definition:

Definition 24 (Parametric Distribution)

*A **parametric distribution** is a set of distribution functions, of which each member is determined by specifying one or more parameters.*

Some common techniques are the following:

- Multiplication by a constant
- Raising to a power
- Exponentiation
- Mixture of distributions

6.1.1 Multiplication By A Constant

This transformation is equivalent to applying inflation uniformly across all loss levels, and is known as a change of scale.

Proposition 5 (Multiplication by a Constant)

Let X be a crv with cdf F_X and pdf f_X. Let $Y = cX$ for some $c > 0$. Then

$$F_Y(y) = F_X\left(\frac{y}{c}\right), \quad f_Y(y) = \frac{1}{c} f_X\left(\frac{y}{c}\right).$$

Proof

$$F_Y(y) = P(Y \leq y) = P(cX \leq y) = P\left(X \leq \frac{y}{c}\right) = F_X\left(\frac{y}{c}\right)$$
$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy} F_X\left(\frac{y}{c}\right) = \frac{1}{c} f_X\left(\frac{y}{c}\right)$$

□

Definition 25 (Scale Distribution)

*We say that a parametric distribution is a **scale distribution** if $Y = cY$ for any positive constant c is from the same set of distributions as X.*

It is clear that we have the following result:

Corollary 6

The parameter c in Proposition 5 is a scale parameter, and Y is a scale distribution.

Example 6.1.1

Let $X \sim \mathrm{Exp}(\theta)$ with pdf

$$f_X(x) = \frac{1}{\theta} e^{-\frac{x}{\theta}}, \quad x > 0.$$

Let $y = cX$ with $c > 0$, it follows that

$$f_Y(y) = \frac{1}{c} f_X\left(\frac{y}{c}\right) = \frac{1}{c\theta} e^{-\frac{y}{c\theta}}, \quad y > 0.$$

Thus $Y \sim \mathrm{Exp}(c\theta)$ and so Y is a scale distribution. In particular, the exponential distribution belongs to a family of scale distributions.

Definition 26 (Scale Parameter)

A parameter θ is called a ***scale paramter*** *of a parametric distribution X if it satisfies the following condition: the parametric value of cX is $c\theta$ for any positive constant c, and other parameters (if any) remain unchanged.*

Example 6.1.2

From Example 6.1.1, we had that

$$f_X(x) = \frac{1}{\theta} e^{-\frac{x}{\theta}}, \quad x > 0.$$

We showed that $Y = cX \sim \mathrm{Exp}(c\theta)$. Therefore, the parameter θ is a scale parameter.

Example 6.1.3

Determine whether the lognormal distribution $X \sim \mathrm{LogN}(\mu, \sigma^2)$, i.e. $\ln(X) \sim \mathrm{N}(\mu, \sigma^2)$, is a scale distribution or not. If yes, determine whether it has any scale parameter.

Solution

Let $Y = cX$ for some $c > 0$. Observe that

$$\ln Y = \ln cX = \ln c + \ln X \sim \mathrm{N}(\mu + \ln c, \sigma^2).$$

For the last equation, note that if we let $Z = \ln X \sim \mathrm{N}(\mu, \sigma^2)$

$$E\left[e^{t(Z+\ln c)}\right] = e^{t\ln c}e^{\mu t + \frac{\sigma^2 t^2}{2}} = e^{t(\mu+\ln c)+\frac{\sigma^2 t^2}{2}}$$

we see that the above is the mgf of $\mathrm{N}(\mu + \ln c, \sigma^2)$. Thus we have that Y has the same distribution as X and so it is a scale distribution. However, we also see that it has no scale parameters.

6.1.2 *Raising to a Power*

Proposition 7 (Raising to a Power)

Let X be a crv with pdf f_X and cdf F_X with $F_X(0) = 0$. Let $Y = X^{\frac{1}{\tau}}$. If $\tau > 0$, then

$$F_Y(y) = F_X(y^\tau), \quad f_Y(y) = \tau y^{\tau-1} f_X(y^\tau), \quad y > 0,$$

while if $\tau < 0$, then

$$F_Y(y) = 1 - F_X(y^\tau), \quad f_Y(y) = -\tau y^{\tau-1} f_X(y^\tau), \quad y > 0.$$

Proof

When $\tau > 0$,

$$F_Y(y) = P(Y \leq y) = P\left(X^{\frac{1}{\tau}}\right) = P\left(X \leq y^\tau\right) = F_X\left(y^\tau\right)$$

and

$$f_Y(y) = \frac{d}{dy}F_Y(y) = \frac{d}{dy}f_X\left(y^\tau\right) = \tau y^{\tau-1} f_X(y^\tau).$$

When $\tau < 0$,

$$F_Y(y) = P(Y \leq y) = P\left(X^{\frac{1}{\tau}} \leq y\right) = P\left(X \geq y^\tau\right) = \overline{F}_X(y^\tau)$$

and

$$f_Y(y) = \frac{d}{dy}F_Y(y) = \frac{d}{dy}(1 - F_X\left(y^\tau\right)) = -\tau y^{\tau-1} f_X\left(y^\tau\right).$$

□

Example 6.1.4

Let $X \sim \text{Exp}(\theta)$ and $Y = X^{\frac{1}{\tau}}$ for $\tau > 0$, we have

$$F_Y(y) = F_X\left(t^\tau\right) = 1 - e^{\frac{-y^\tau}{\theta}} = 1 - e^{-\left(\frac{y}{\alpha}\right)^\tau},$$

where $\alpha = \theta^{\frac{1}{\tau}}$. In particular, we have that $Y \sim \text{Wei}(\alpha, \tau)$.

6.1.3 *Exponentiation*

Proposition 8 (Exponentiation Method)

Let X be a crv with pdf f_X and cdf F_X. Let $Y = e^X$. Then

$$F_Y(y) = F_X(\ln y), \quad f_Y(y) = \frac{1}{y} f_X(\ln y).$$

Proof

We have

$$F_Y(y) = P\left(e^X \leq y\right) = P(X \leq \ln y) = F_X(\ln y)$$

and

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy} F_X(\ln y) = \frac{1}{y} f_X(\ln y).$$

□

Exercise 6.1.1 (Lognormal Distribution)

Let $X \sim \mathrm{N}(\mu, \sigma^2)$. The cdf and pdf of $Y = e^X$ is

$$F_Y(y) = F_X(\ln y) = \Phi\left(\frac{\ln y - \mu}{\sigma}\right)$$

$$f_Y(y) = \frac{1}{y} f_X(\ln y) = \frac{1}{y} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2} \cdot \left(\frac{\ln y - \mu}{\sigma}\right)^2}$$

6.1.4 Mixing Distributions

The rationale behind mixing distributions is to define an rv X conditional on a second rv, say Θ (aka **mixing rv**). The mixing rv Θ can either be discrete or be continuous, which leads to two types of mixtures:

- **discrete mixture**: when Θ is discrete; and
- **continuous mixture**: when Θ is continuous.

Definition 27 (Discrete Mixed Distribution)

Let Θ be a drv taking values on $\{\theta_1, \theta_2, \dots, \theta_n\}$ with

$$P(\Theta = \theta_i) = p_i > 0, \quad i = 1, \dots, n,$$

and the rv $Y_i := X \mid \Theta = \theta_i$ has cdf

$$F_{Y_i}(x) = P(X \leq x \mid \Theta = \theta_i), x \in \mathbb{R}.$$

*Then X is called a **discrete mixed distribution** with cdf*

$$F_X(x) = \sum_{i=1}^{n} P(X \leq x \mid \Theta = \theta_i) P(\Theta = \theta_i) = \sum_{i=1}^{n} p_i F_{Y_i}(x).$$

Following the above definition, by the Law of the Unconscious Statistician, we have

$$E[g(X)] = \sum_{i=1}^{n} E[g(X) \mid \Theta = \theta_i] P(\Theta = \theta_i) = \sum_{i=1}^{n} p_i E[g(Y_i)],$$

for any function g such that the expectation exists. In particular, we have

$$E[X] = \sum_{i=1}^{n} p_i E[Y_i] \text{ and } E\left[X^2\right] = \sum_{i=1}^{n} p_i E\left[Y_i^2\right].$$

Example 6.1.5

Let $Y_i \sim \mathrm{Exp}(i)$ for $i = 1, 2, 3$. Define X to be an equal mixture of these three exponential rvs. Fidn the cdf, pdf, and mean of X.

Solution

The cdf of X is

$$\begin{aligned} F_X(x) &= \sum_{i=1}^{3} \frac{1}{3} F_{Y_i}(x) = \frac{(1-e^{-x}) + (1-e^{-x/2}) + (1-e^{-x/3})}{3} \\ &= 1 - \frac{1}{3}\left(e^{-x} + e^{-\frac{x}{2}} + e^{-\frac{x}{3}}\right), x > 0. \end{aligned}$$

The pdf of X is

$$f_X(x) = \frac{1}{3}\left(e^{-x} + \frac{1}{2}e^{-\frac{x}{2}} + \frac{1}{3}e^{-\frac{x}{3}}\right), x > 0.$$

The mean of X is therefore

$$E[X] = \sum_{i=1}^{3} E[Y_i] = \frac{1}{3}(1+2+3) = 2.$$

7 Lecture 7 Sep 27th

7.1 Severity Distributions — Creating Severity Distributions (Continued 2)

7.1.1 Mixing Distributions (Continued)

Definition 28 (Continuous Mixture)

Let Θ be a crv with density f_Θ, and the cdf and pdf of $X \mid \Theta = \theta$ are given by

$F_{X|\Theta}(x \mid \theta) = P(X \leq x \mid \Theta = \theta)$ *and* $f_{X|\Theta}(x \mid \theta) = P(X = x \mid \Theta = \theta)$.

The unconditional distribution of X is said to be a ***continuous mixed distribution*** *with cdf and pdf*

$$F_X(x) = \int_{-\infty}^{\infty} F_{X|\Theta}(x \mid \theta) f_\Theta(\theta)\, d\theta$$
$$f_X(x) = \int_{-\infty}^{\infty} f_{X|\Theta}(x \mid \theta) f_\Theta(\theta)\, d\theta.$$

Furthermore, for any function H,

$$E[H(X)] = \int_{-\infty}^{\infty} E[H(X) \mid \Theta = \theta] f_\Theta(\theta)\, d\theta.$$

Example 7.1.1

Suppose that $X \mid \Lambda = \lambda$ is exponentially distributed with mean $\frac{1}{\lambda}$, and let Λ be a gamma distributed rv with mean α/θ and variance α/θ^2, i.e.

$$f_\Lambda(\lambda) = \frac{\theta^\alpha \lambda^{\alpha-1} e^{-\theta\lambda}}{\Gamma(\alpha)}, \lambda > 0,$$

where $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1}e^{-t}\,dt$ is the gamma function. Determine the conditional pdf of X.

Solution

We have

$$\begin{aligned}
f_X(x) &= \int_0^\infty f_{X|\Lambda}(x \mid \lambda) f_\Lambda(\lambda)\,d\lambda \\
&= \int_0^\infty \lambda e^{-x\lambda} \frac{\theta^\alpha \lambda^{\alpha-1}e^{-\theta\lambda}}{\Gamma(\alpha)}\,d\lambda \\
&= \frac{\theta^\alpha}{\Gamma(\alpha)} \int_0^\infty \lambda^\alpha e^{-\lambda(x+\theta)}\,d\lambda \\
&= \frac{\theta^\alpha}{\Gamma(\alpha)(x+\theta)} \int_0^\infty \left(\frac{y}{x+\theta}\right)^\alpha e^{-y}\,dy \quad \text{where } y = \lambda(x+\theta) \\
&= \frac{\theta^\alpha}{\Gamma(\alpha)(x+\theta)^{\alpha+1}} \int_0^\infty y^\alpha e^{-y}\,dy \\
&= \frac{\theta^\alpha \Gamma(\alpha+1)}{\Gamma(\alpha)(x+\theta)^{\alpha+1}} = \frac{\alpha\theta^\alpha}{(x+\theta)^{\alpha+1}}.
\end{aligned}$$

Proposition 9 (Total Expectation and Total Variance)

For any rvs X and Θ, provided that the repsective expectation and variance exist, we have

$$\begin{aligned}
E[X] &= E[E[X \mid \Theta]] \\
\mathrm{Var}(X) &= E[\mathrm{Var}(X \mid \Theta)] + \mathrm{Var}(E[X \mid \Theta])
\end{aligned}$$

Proof

$$\begin{aligned}
E[X] &= E\left(\int_X x f_{X|\Theta}(x \mid \Theta)\,dx\right) \\
&= \int_\Theta \int_X x f_{X|\Theta}(x \mid \theta) f_\Theta(\theta)\,dx\,d\theta \\
&= \int_X x \int_\Theta f_{X,\Theta}(x,\theta)\,d\theta\,dx \quad \because \text{ Fubini's Theorem} \\
&= \int_X x f_X(x)\,dx = E[X].
\end{aligned}$$

Note that

$$\mathrm{Var}(X \mid \Theta) = E[X^2 \mid \Theta] + E[X \mid \Theta]^2.$$

And so

$$\begin{aligned}
&E[\operatorname{Var}(X \mid \Theta)] + \operatorname{Var}(E[X \mid \Theta]) \\
&= E[E[X^2 \mid \Theta]] - E\left[E[X \mid \Theta]^2\right] + E\left[E[X \mid \Theta]^2\right] - E[E[X \mid \Theta]]^2 \\
&= E\left[X^2\right] - E[X]^2 = \operatorname{Var}(X)
\end{aligned}$$

□

Example 7.1.2

Suppose that $X \mid \Theta = \theta \sim \operatorname{Exp}(\theta)$ and $p_\Theta(\theta) = \frac{1}{3}$ for $\theta = 1, 2, 3$. Find the mean and variance of X.

Solution

The mean of X is

$$E[X] = EE[X \mid \Theta] = E[\Theta] = \frac{1}{3}(1 + 2 + 3) = 2.$$

The variance of X is

$$\begin{aligned}
\operatorname{Var}(X) &= E[\operatorname{Var}(X \mid \Theta)] + \operatorname{Var}(E[X \mid \Theta]) \\
&= E[\Theta^2] + \operatorname{Var}(\Theta) = 2E[\Theta^2] - E[\Theta]^2 \\
&= \frac{2}{3}(1 + 4 + 9) - 4 = \frac{28}{3} - \frac{12}{3} = \frac{16}{3}
\end{aligned}$$

Example 7.1.3

Suppose that $X \mid \Lambda = \lambda \sim \operatorname{Exp}(\lambda)$ and $\Lambda \sim \operatorname{Gam}(\alpha, \theta)$ with mean $\alpha\theta$ and variance $\alpha\theta^2$. Find the mean and variance of X.

Solution

The mean of X is

$$E[X] = EE[X \mid \Lambda] = E[\Lambda] = \alpha\theta.$$

The variance of X is

$$\begin{aligned}
\operatorname{Var}(X) &= E[\operatorname{Var}(X \mid \Lambda)] + \operatorname{Var}(E[X \mid \Lambda]) \\
&= E[\Lambda^2] + \operatorname{Var}(\Lambda) = 2\operatorname{Var}(\Lambda) + E[\Lambda]^2 \\
&= 2\alpha\theta^2 + \alpha^2\theta^2.
\end{aligned}$$

7.2 Severity Distributions — Tail of Distributions

Definition 29 (Tail)

The ***tail*** *of a distribution (usually the right tail) is the portion of the distribution corresponding to large values of the random variable.*

It is important that we understand large possible loss values as they have the greatest impact on the total losses that we may have to endure. In general, a loss rv is said to be **heavy-tailed** if it has a large probability to take large values.

Two measurements of tail weight:

- **relative**: comparing "sizes" of the tails of two distributions;
- **absolute**: classifying distributions as heavy or light-tailed.

The following is a set of criteria to measure or compare the heaviness of the tails of loss distributions:

- Existence of moments
- Limiting ratios
- Hazard rate function
- Mean excess loss function

7.2.1 Existence of Moments

Recall that the kth moment of a loss X is

$$E\left[X^k\right] = \int_0^\infty x^k f_X(x)\,dx.$$

Now if f_X takes on large values for large x, we may have $E\left[X^k\right]$ blow up to infinity, and so it is desirable to find/use some distribution with a **decaying** probability function, one at which its rate of decay is faster than the growth of $x^{-(k+1)}$.

8 Lecture 8 Oct 02nd

8.1 Severity Distributions — Tail of Distributions (Continued)

8.1.1 Existence of Moments (Continued)

Example 8.1.1

For a Pareto distribution, as $x \to \infty$, we have that $f_X(x) \sim x^{-(\alpha+1)}$, so its moments are finite if and only if $k < \alpha$.

We say that the Pareto distribution has a **power tail**.

Example 8.1.2

Given the transformed Gamma distribution, with pdf

$$f_X(x) = \frac{\left(\frac{x}{\theta}\right)^{\alpha} e^{-\frac{x}{\theta}}}{x\Gamma(\alpha)}.$$

Now as $x \to \infty$, we have

$$f_X(x) \sim x^{\alpha-1} e^{-\frac{x}{\theta}}$$

We see that the exponential term decays faster than the rate of growth of $x^{\alpha-1}$ for any $\alpha > 0$. Thus all moments of the Gamma distribution exists.

We say that the Gamma distribution has a **exponential tail**.

Exercise 8.1.1

The Normal distribution has an exponential tail.

Definition 30 (Heavy-Tails Light-Tails)

We say that a distribution is a ***heavy-tailed distribution*** *if* ***its moments only exist up to some*** $k \in \mathbb{N} \setminus \{0\}$.

We say that a distribution is a ***light-tail distribution*** *if* ***its moments exist for all*** $k \in \mathbb{N} \setminus \{0\}$.

The actual definition, or should I say notion, of tail-heaviness comes from talking about the boundedness of the tail of the distribution, with reference to the exponential distribution. If a distribution has a tail that has greater value than the tail of the exponential distribution, then we say that the distribution has a heavy-tail.

Note

We may also use the mgf to determine if a distribution has a heavy or light tail; the inexistence of the kth moment implies the inexistence of the mgf, i.e. if the mgf does not exist, then the moments of the distribution is only finite up to some $k \in \mathbb{N} \setminus \{0\}$.

8.1.1.1 *Limiting Ratio: Survival Functions*

Definition 31 (Limiting Ratio)

The ***limiting ratio*** *of* ***two survival functions*** *is used to compare the heaviness of tails of the two losses. Consider two losses X and Y, and consider the limit of the ratio*

$$\lim_{x \to \infty} \frac{\overline{F}_X(x)}{\overline{F}_Y(x)}.$$

If the limit does not exist, we say that the comparison is inconclusive. Otherwise, we have 3 cases:

- *If $c = 0$, then $\overline{F}_X(x)$ decays faster than $\overline{F}_Y(x)$ as $x \to \infty$, i.e. Y has a heavier tail than X;*
- *If $0 < c < \infty$, then $\overline{F}_X(x)$ and $\overline{F}_Y(x)$ decays at the smae rate, as $x \to \infty$, i.e. X and Y have similar tails;*
- *If $c = \infty$, then $\overline{F}_X(x)$ decays slower than $\overline{F}_Y(x)$ as $x \to \infty$, i.e. X has a heavier tail than Y;*

where we let

$$c := \lim_{x \to \infty} \frac{\overline{F}_X(x)}{\overline{F}_Y(x)}$$

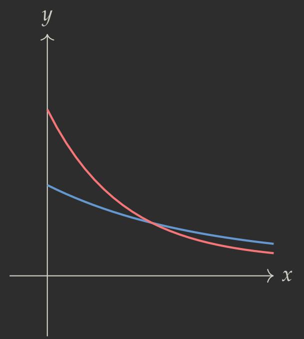


Figure 8.1: Limiting Ratio

Note
Not all distributions have an explicit survival function, but they will always have a pdf/pmf. Fortunately, by ***L'Hôpital's Rule****, the above definition can be applied to the pdfs of X and Y, i.e.*

$$c = \lim_{x\to\infty} \frac{\bar{F}_X(x)}{\bar{F}_Y(x)} = \lim_{x\to\infty} \frac{-f_X(x)}{-f_Y(x)} = \lim_{x\to\infty} \frac{f_X(x)}{f_Y(x)}$$

Example 8.1.3

Show that the Pareto distribution has a heavier tail than the Gamma distribution using limiting ratio.

Solution
Let $X \sim \text{Pareto}(\alpha, \theta)$ and $Y \sim \text{Gam}(\tau, \lambda)$. We have

$$c = \lim_{x\to\infty} \frac{f_X(x)}{f_Y(x)} = \lim_{x\to\infty} \frac{\frac{\alpha\theta^\alpha}{(x+\theta)^{\alpha+1}}}{\frac{x^{\tau-1}e^{-\frac{x}{\lambda}}}{\lambda^\tau\Gamma(\tau)}} = \alpha\theta^\alpha\lambda^\tau\Gamma(\tau) \lim_{x\to\infty} \frac{e^{\frac{x}{\lambda}}}{x^{\tau-1}(x+\theta)^{\alpha+1}}$$

Since the exponential term grows faster than the term in the denominator, we have $c = \infty$, i.e. X has a heavier tail than Y, as required.

Example 8.1.4

For two losses X and Y, suppose that $f_X(x) = \frac{2}{\pi(1+x^2)}$ and $f_Y(x) = \frac{1}{(1+x^2)}$ for $x > 0$. Compare the tail heaviness of the two losses.

Solution
Notice that

$$c = \lim_{x\to\infty} \frac{f_X(x)}{f_Y(y)} = \lim_{x\to\infty} = \frac{2}{\pi} < \infty,$$

i.e. X and Y have similar tails.

8.1.1.2 *Hazard Rate*

Recall Definition 12. We had

$$h(x) = \frac{f(x)}{\bar{F}(x)} = -\frac{d}{dx} \ln \bar{F}(x),$$
$$h_X(x)\Delta x \approx P(X \leq x + \Delta x \mid X > x)$$

and the hazard rate function relates to the survival function as

$$\bar{F}(x) = e^{-\int_{-\infty}^{x} h(y)\,dy}.$$

Notice that

- if the hazard rate function is a **decreasing** function, that implies that the probability of the occurrence of $X \leq x + \Delta x$ decreases given $X > x$, as x increases, i.e. it is more likely that we have $X > x + \Delta x \mid X > x$. So X has a **heavy tail**.
- if the hazard rate function is a **increasing** function, that implies that the probability of the occurrence of $X \leq x + \Delta x$ increases given $X > x$, as x increases, i.e. it is less likely that $X > x + \Delta x \mid X > x$. So X has a **light tail**.

Definition 32 (Decreasing and Increasing Failure Rates)

Let X be a loss with hazard rate function h_X. We say that[1]

- *X or F_X has a* ***decreasing failure rate (DFR)*** *if h_X is decreasing;*
- *X or F_X has a* ***increasing failure rate (IFR)*** *if h_X is increasing.*

[1] The following source claims that the **failure rate** and hazard rate are, in fact, not always interchangable terms: `https://nomtbf.com/2013/11/difference-hazard-failure-rate/`. Perhaps this is worth looking into.

Note

Consequently,

- *Distributions that have a DFR are heavy-tailed;*
- *Distributions that have an IFR are light-tailed.*

Proposition 10 (Exponential has Constant Hazard Rate)

The exponential distribution has a constant hazard rate.

Proof

The pdf and survival function of $X \sim \mathrm{Exp}(\lambda)$ is

$$f_X(x) = \lambda e^{-\lambda x} \text{ and } \bar{F}_X(x) = e^{-\lambda x},$$

respectively. Thus the hazard rate of X is

$$h(x) = \frac{f_X(x)}{\bar{F}_X(x)} = \lambda,$$

which is a fixed value. □

Note

We say that the exponential distribution is the only distribution which is said to have both DFR and IFR.[2]

[2] **Why?**

Example 8.1.5

Let $X \sim \text{Pareto}(\alpha, \theta)$ with $f_X(x) = \frac{\alpha\theta^\alpha}{(x+\theta)^{\alpha+1}}$ and $\bar{F}_X(x) = \frac{\theta^\alpha}{(x+\theta)^\alpha}$. Determine whether X has a DFR or IFR.

Solution

The hazard rate function of X is

$$h_X(x) = \frac{f_X(x)}{\bar{F}_X(x)} = \frac{\frac{\alpha\theta^\alpha}{(x+\theta)^{\alpha+1}}}{\frac{\theta^\alpha}{(x+\theta)^\alpha}} = \frac{\alpha}{x+\theta}.$$

It is clear that h_X is a decreasing function, and so $X \sim \text{Pareto}(\alpha, \theta)$ has a DFR, i.e. it is heavy-tailed.

It is not always easy to get the survival function. The following is an alternative approach to finding out if the hazard rate function is increasing or decreasing.

Proposition 11 (Ratio Comparison for DFR/IFR)

Let X be an rv, and[3]

$$s(x) = \frac{f_X(x+y)}{f_X(x)}.$$

1. *If $s(x)$ is increasing in x for every y, then X has a DFR;*
2. *If $s(x)$ is decreasing in x for every y, then X has an IFR.*

[3] **Any bounds on y?**

Proof

We shall prove for one case as the other will follow analogously. Notice that

$$h_X(x) = \frac{f_X(x)}{\overline{F}_X(x)} = \frac{f_X(x)}{\int_x^\infty f_X(y)\,dy} = \frac{1}{\int_0^\infty \frac{f_X(x+y)}{f_X(x)}\,dy}$$

by a change of variable in the last equality. We notice that if $\frac{f_X(x+y)}{f_X(x)}$ is increasing, then $h_X(x)$ will be decreasing, and so X has a DFR. □

Example 8.1.6

Let $X \sim \mathrm{Gam}(\alpha, \theta)$ with $\alpha > 1$. Determine whether X is a DFR or IFR distribution.

Solution

The cdf of X is

$$f_X(x) = \frac{x^{\alpha-1}e^{-\frac{x}{\theta}}}{\theta^\alpha \Gamma(\alpha)}.$$

The survival function of X is not explicit, and so we should use Proposition 11. We have

$$\frac{f_X(x+y)}{f_X(x)} == \frac{\frac{(x+y)^{\alpha-1}e^{-\frac{x+y}{\theta}}}{\theta^\alpha\Gamma(\alpha)}}{\frac{x^{\alpha-1}e^{-\frac{x}{\theta}}}{\theta^\alpha\Gamma(\alpha)}} = \left(\frac{x+y}{x}\right)^{\alpha-1} e^{-\frac{y}{\theta}} = \left(1+\frac{y}{x}\right)^{\alpha-1} e^{-\frac{y}{\theta}}.$$

To try to determine if it is increasing or decreasing, we calculate the second derivative of the ratio:

$$\frac{d}{dx}\left(1+\frac{y}{x}\right)^{\alpha-1} e^{-\frac{y}{\theta}} = y(\alpha-1)\left(1+\frac{y}{x}\right)^{\alpha-2} e^{-\frac{y}{\theta}}.$$

It is important to note that y is not completely free: it is bounded below by $-x$, as if $y < -x$, then $x + y < 0$, and f is undefined at these values. Also, if $y = -x$, then the ratio is simply a constant, and we cannot use Proposition 11 to reach a conclusion. To be able to use Proposition 11, we must have $y > -x$. In this case, it is clear that the ratio is increasing as x increases. Thus X has an IFR.

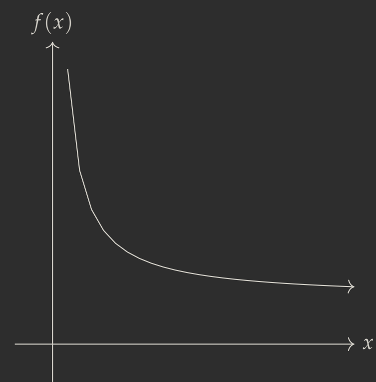


Figure 8.2: Graph of $(1+\frac{y}{x})^{\alpha-1}e^{-\frac{y}{\theta}}$ for $y > -x$ and $x > 0$.

9 Lecture 9 Oct 11th

9.1 Severity Distributions — Tail of Distributions (Continued 2)

9.1.1 Mean Excess Loss

Definition 33 (Excess Loss Random Variable)

For a loss rv X, we define the ***excess loss rv*** *as*

$$T_d = X - d \mid X > d, \quad d > 0.$$

The survival function of T_d is

$$\begin{aligned}\bar{F}_{T_d}(x) &= P(T_d > x) = P(X - d > x \mid X > d) \\ &= \frac{P(X > x + d)}{P(X > d)} = \frac{\bar{F}_X(x+d)}{\bar{F}_X(d)}.\end{aligned}$$

As defined before in Definition 22,

Definition (Mean Excess Loss)

The ***mean excess loss*** *(or* ***mean residual life****) function is defined as*

$$e_X(d) = E[T_d] = \int_0^\infty \bar{F}_{T_d}(x)\,dx = \frac{\int_0^\infty \bar{F}_X(x+d)\,dx}{\bar{F}_X(d)} = \frac{\int_d^\infty \bar{F}_X(y)\,dy}{\bar{F}_X(d)}$$

Essentially, the mean excess loss is the average payment in excess of the threshold d, given that the loss exceeds the threshold.

Definition 34 (Increasing and Decreasing Mean Residual Lifetime)

Given a loss rv X,

1. *we say X or F_X is an* ***increasing mean residual lifetime (IMRL)*** *if $e_X(x)$ is increasing in x;*
2. *we say X or F_X is an* ***decreasing mean residual lifetime (DMRL)*** *if $e_X(x)$ is decreasing in x.*

Note

- *IMRL distributions are* ***heavy-tailed****;*
- *DMRL distributions are* ***light-tailed****.*

The reason of this claim should be rather clear from the context of $e_X(x)$: if $e_X(x)$ is increasing with x, then we expect that the survival probability of T_d to be greater, and so the tail should be a heavy one. The following proposition clarifies this notion.

Proposition 12 (Relation between DFR/IFR and IMRL/DMRL)

A DFR rv is IMRL, and an IFR rv is a DMRL.

Proof

Suppose X has a DFR. The mean excess loss of X is

$$e_X(d) = \frac{\int_0^\infty \bar{F}_X(x+d)\,dx}{\bar{F}_X(d)} = \int_0^\infty \frac{\bar{F}_X(x+d)}{\bar{F}_X(d)}\,dx.$$

Note that by the relationship between the survival function and the hazard rate[1],

$$\frac{\bar{F}_X(x+d)}{\bar{F}_X(d)} = \frac{e^{-\int_0^{x+d} h_X(y)\,dy}}{e^{-\int_0^d h_X(y)\,dy}} = e^{-\int_d^{x+d} h_X(y)\,dy} = e^{-\int_0^x h_X(z+d)\,dz}.$$

Since X has a DFR, h_X is decreasing, and thus $\frac{\bar{F}_X(x+d)}{\bar{F}_X(d)}$ is increasing. Thus $e_X(d)$ is increasing and so X is a IMRL, as required. THe argument is similar for *IFL* being a *DMRL*. □

[1] We use the hazard rate here because it is provided by the assumption.

Example 9.1.1

Let $X \sim \text{Wei}(\theta, \tau)$. Determine whether X is DMRL or IMRL.

Solution

Since

$$f_X(x) = \frac{\tau x^{\tau-1} e^{-\left(\frac{x}{\theta}\right)^\tau}}{\theta^\tau}$$

and from an earlier example, we have

$$\bar{F}_X(x) = e^{-\left(\frac{x}{\theta}\right)^\tau}$$

Then the hazard rate is

$$h_X(x) = \frac{f_X(x)}{\bar{F}_X(x)} = \frac{\tau}{\theta^\tau} x^{\tau-1}.$$

Now if $\tau \geq 1$, then $h_X(x)$ is an increasing function, and so X has an IFR, i.e. X is a DMRL. if $0 < \tau \leq 1$, then $h_X(x)$ is a decreasing function, and so X has a DFR, i.e. X is an IMRL.

Example 9.1.2

Consider a loss X with $f_X(x) = (1 + 2x^2)e^{-2x}$ for $x > 0$.

1. Determine $h_X(x)$.
2. Determine $e_X(x)$.
3. Find $\lim\limits_{x\to\infty} h_X(x)$ and $\lim\limits_{x\to\infty} e_X(x)$.
4. Show that X is DMRL but not IFR.

Solution

Since both $h_X(x)$ and $e_X(x)$ require the survival function, we shall first derive that. Observe that[2]

$$\begin{aligned}
\bar{F}_X(x) &= \int_x^\infty (1+2y^2)e^{-2y}\,dy = \frac{1}{2}e^{-2x} + 2\left[\int_x^\infty y^2 e^{-2y}\,dy\right] \\
&= \frac{1}{2}e^{-2x} + 2\left[-\frac{1}{2}y^2e^{-2y}\Big|_x^\infty + \int_x^\infty ye^{-2y}\,dy\right] \\
&= \frac{1}{2}e^{-2x} + x^2e^{-2x} + 2\left[-\frac{1}{2}ye^{-2y}\Big|_x^\infty + \frac{1}{2}\int_x^\infty e^{-2y}\,dy\right] \\
&= \frac{1}{2}e^{-2x} + x^2e^{-2x} + xe^{-2x} + \frac{1}{2}e^{-2x} \\
&= (x^2+x+1)e^{-2x}.
\end{aligned}$$

[2] It is highly recommended that one gets really used to using integration by parts, to the point that you do not have to repeatedly write down what the u and dv are explicitly every time.

1. It is clear that

$$h_X(x) = \frac{1+2x^2}{1+x+x^2}$$

2. By its definition, we have that

$$e_X(x) = \frac{\int_x^\infty \bar{F}_X(y)\,dy}{\bar{F}_X(x)},$$

and so we need to solve for the integral in the numerator. Using pieces from our derivation of $\bar{F}_X(x)$, we obtain

$$\begin{aligned}
&\int_x^\infty (1+y+y^2)e^{-2y}\,dy \\
&= \frac{1}{2}e^{-2x} + \frac{1}{2}xe^{-2x} + \frac{1}{4}e^{-2x} + \frac{1}{2}x^2e^{-2x} + \frac{1}{2}xe^{-2x} + \frac{1}{4}e^{-2x} \\
&= \left(1 + x + \frac{1}{2}x^2\right)e^{-2x}.
\end{aligned}$$

Thus

$$e_X(x) = \frac{1+x+\frac{1}{2}x^2}{1+x+x^2}.$$

3. The answers are straightforward[3]

$$\begin{aligned}
\lim_{x\to\infty} h_X(x) &= \lim_{x\to\infty} \frac{\frac{1}{x^2}+2}{1+\frac{1}{x}+\frac{1}{x^2}} = 2 \\
\lim_{x\to\infty} e_X(x) &= \lim_{x\to\infty} \frac{\frac{1}{x^2}+\frac{1}{x}+\frac{1}{2}}{\frac{1}{x^2}+\frac{1}{x}+1} = \frac{1}{2}
\end{aligned}$$

[3] **Find out why did we calculate these values.**

4. First, observe that

$$\begin{aligned}
e_X'(x) &= \frac{(1+x)\left(1+x+x^2\right) - (1+2x)\left(1+x+\frac{1}{2}x^2\right)}{(1+x+x^2)} \\
&= -\frac{x+\frac{1}{2}x^2}{(1+x+x^2)^2},
\end{aligned}$$

and we see that $e_X'(x) < 0$ for $x > 0$. Thus X has a DMRL. For $h_X(x)$,

$$\begin{aligned}
h_X'(x) &= \frac{4x\left(1+x+x^2\right) - (1+2x)\left(1+2x^2\right)}{\left(1+x+x^2\right)^2} \\
&= \frac{2x^2+2x-1}{x^4+2x^3+3x^3+2x+1}.
\end{aligned}$$

It may appear as if $h_X'(x)$ is positive, seeing that x^4 should dominate. However, notice that the discriminant is positive:[4]

$$2^2 - 4(2)(-1) = 12 > 0,$$

and so the numerator has a root, i.e. there are critical points on $h_X(x)$. In fact, equating the said numerator to 0, we can obtain that $x = -\frac{1}{2} + \sqrt{\frac{3}{4}}$ (the other case is ruled out as $x > 0$). Since $h'_X(x)$ looks as if it is increasing, let's try out some values of x for $0 < x < \sqrt{\frac{3}{4}} - \frac{1}{2}$. In particular, notice that

$$h_X\left(\frac{1}{10}\right) = \frac{102}{111} \approx 0.9198$$
$$h_X\left(\frac{1}{5}\right) = \frac{27}{31} \approx 0.8710$$

but $\frac{1}{10} < \frac{1}{5}$, and so we notice that X is not IFR.

[4] Lecture notes simply threw the values 1 and $\frac{1}{2}$ for x almost out of nowhere. While the result seems harmless, firstly, $x \neq 0$, since $x > 0$. In fact, since the critical point is $\sqrt{\frac{3}{4}} - \frac{1}{2} \approx 0.366$, $\frac{1}{2}$ is a value that comes after the critical point, so we would not have been able to verify without trying and failing numerous times, especially since the critical point is an irrational value.

Here, we are smart and equipped with the knowledge that by solving the first derivative for x by equating to 0 allows us to find these critical points, which is indicative of a change from positive to negative, or vice versa, slope for $h_X(x)$.

10 Lecture 10 Oct 16th

10.1 Severity Distributions — Policy Adjustments (Continued)

Insurance policies contain various **adjustments** to soften the amount that insurers have to pay, to minimize moral hazards, and for various other reasons. In this section, we shall introduce some common policy adjustments.

In the following definitions, suppose that X is our ground-up loss rv, and H a function incurred by the adjustment.

Definition 35 (Policy Limit)

*A fixed level $u > 0$ is called a **policy limit** if, provided that there are no other adjustments, the insurer shall pay*[1]

$$H(X) = \min\{X, u\} := X \wedge u = \begin{cases} X & X \leq u \\ u & X \geq u \end{cases}.$$

[1] Now that this definition uses the symbol $\wedge$ for denoting a policy limit, I shall refrain from using the same symbol in proofs, unless if the context is clear.

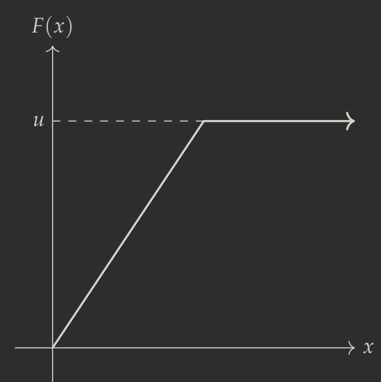


Figure 10.1: Typical graph of a policy limit, without other adjustments.

Note

- *A policy limit protects the insurer from overly large losses.*
- *This is **noteworthy**: in practice, a policy limit may refer to **the maximum amount paid** by the insurer, but in this course, it is the **maximum loss** coverred by the insurer.*

Definition 36 (Ordinary Deductible)

A fixed level $d > 0$ is called an ***ordinary deductible*** *if, given that there are no other adjustments, the insurer pays*

$$Y = H(X) = (X-d)_+ = X \vee d = \begin{cases} 0 & X < d \\ X-d & X \geq d \end{cases}$$

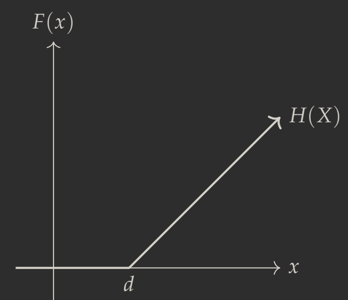


Figure 10.2: Graph of a policy with ordinary deductible without any other adjustments.

Note

- *For any given loss, the first d dollars falls on the insured.*
- *It is a protection against frequent small claims.*

Definition 37 (Franchise Deductible)

A fixed level $d > 0$ is called a ***Franchise Deductible*** *if, given that there are no other adjustments, the insurer pays*

$$\begin{aligned} H(X) = X \cdot \mathbb{1}_{\{X>d\}} &= \begin{cases} 0 & X \leq d \\ X & X > d \end{cases} \\ &= (X-d)\mathbb{1}_{\{X>d\}} + d \cdot \mathbb{1}_{\{X>d\}} \\ &= (X-d)_+ + d \cdot \mathbb{1}_{\{X>d\}} \end{aligned}$$

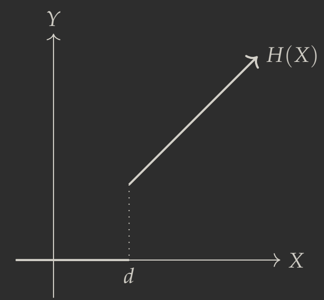


Figure 10.3: Graph of a policy with Franchise deductible without any other adjustments.

Note

- *This differs from the ordinary deductible in that twhen the loss exceeds d, the deductible is waived and the* ***full loss is paid*** *by the insurer.*
- *We are not concerned with whether the payment goes out or not at $X = d$ in this course.*[2]

[2] In the event that a problem of such a nature comes out in either exercises or exams, the point will be explicitly stated.

Remark

This is not a good adjustment as it is prone to ***moral hazard****.*

Definition 38 (Coinsurance)

A fixed rate $\alpha \in [0,1]$ is called a ***coinsurance factor*** *if, given that there are no other adjustments, the insurer pays*

$$H(X) = \alpha X.$$

For any given loss, the ***insurer pays a proportion*** $100\alpha\%$ *of the loss amount the remaining $100(1-\alpha)\%$ falls on the insured.*

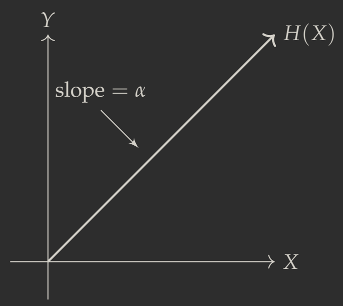


Figure 10.4: Graph of a policy with coinsurance without any other adjustments.

10.1.1 Application Order for Multiple Adjustments

If an insurance policy has more than one adjustment, we assume the adjustments in the following order:

- Policy limit (if any)
- Policy/ordinary deductible (if any)
- Coninsurance (if any)

Note

- *These transformations are not necessarily commutative, so the order must be obeyed.*
- *This ordering is optimal, i.e. it covers for all possible combinations, i.e. any other ways of adjustment can be expressed in this form.*[3]
- *If d is a deductible and u the policy limit, we must have that $d < u$, since if $u < d$, then the insurer will only pay the maximum amount u if the loss exeeds d, which is absurd. Therefore, for all of the cases that we shall consider, we will always assume, and safely so, that $d < u$.*

[3] Claimed by lecturer. Require example.

Applying the ordering, we have

$$X \to X \wedge u \to [(X \wedge u) - d]_+ \to \alpha[(X \wedge u) - d]_+$$

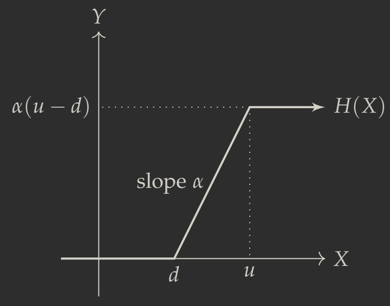


Figure 10.5: Graph of $H(X) = \alpha[(X \wedge u) - d]_+$.

and so

$$H(X) = \alpha[(X \wedge u) - d]_+ = \begin{cases} 0 & X < d \\ \alpha(X - d) & d \leq X < u \\ \alpha(u - d) & X \geq u \end{cases}$$

For the case of applying **Franchise deductible** instead of ordinary deductible, we have

$$X \to X \wedge u \to (X \wedge u)\mathbb{1}_{\{X>d\}} \to \alpha(X \wedge u) \cdot \mathbb{1}_{\{X>d\}}$$

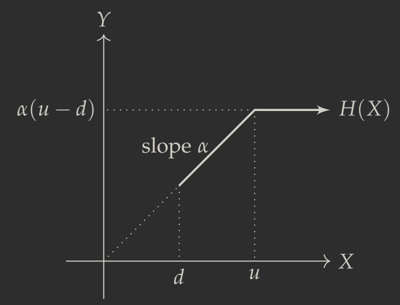


Figure 10.6: Graph of $H(X) = \alpha(X \wedge u) \cdot \mathbb{1}_{\{X>d\}}$.

Notice that $X \wedge u \to (X \wedge u)\mathbb{1}_{\{X>d\}}$, since $X \wedge u > d$ is simply $X > d$ as $u > d$ by assumption. We have that for the case where we consider the Franchise deductible instead of an ordinary deductible,

$$H(X) = \alpha(X \wedge u) \cdot \mathbb{1}_{\{X>d\}} = \begin{cases} 0 & X < d \\ \alpha X & d \leq X < u \\ \alpha u & X \geq u \end{cases}$$

10.1.2 ★ *Reporting Methods*

When we consider the amount paid by the insurer, we typically consider (and distinguish) between two types of reporting methods.

- **Loss basis**: Y_L = amount paid per loss
- **Payment basis**: Y_P = amount paid per payment

It is sensible that

$$Y_P = Y_L \mid Y_L > 0. \tag{10.1}$$

The above relationship shows that we can retrieve Y_P from Y_L, i.e. Y_L holds more information than Y_P. This makes sense; a loss may occur, but the insurer may not have to pay for the loss. For example, if the incurred loss is below a given deductible level.

10.1.2.1 *Loss Basis*

- Each loss is recorded, i.e. each loss has an entry, even if the amount paid is 0.
- For a policy limit u, ordinary deductible d with $d < u$ [4], and

[4] It would be silly if $d > u$.

coinsurance factor α, we have

$$Y_L = \alpha[(X \wedge u) - d]_+.$$

- For a policy with limit u, Franchise deductible d with $d < u$, and coinsurance factor α, we have

$$Y_L = \alpha(X \wedge u)\mathbb{1}_{\{X>d\}}.$$

- Note that in the presence of a deductible d, it is **usually the case** that Y_L has a probability mass at 0, i.e.

$$P(Y_L = 0) = P(X \le d) = F_X(d) > 0.$$

In this case, $Y_P > 0$ almost surely.

10.1.2.2 *Payment Basis*

- Only **non-zero** payments of the insurer are included, and so not every loss will have an entry. Here, we see that Y_P leaves that information behind.[5]

[5] It is still useful to reporting purely on the financial effects of the claims.

- Y_P does not have a probability mass at 0 (**Why**?), i.e.

$$P(Y_P = 0) = 0,$$

or equivalently,

$$Y_P > 0.$$

Example 10.1.1

Let X be the ground-up loss rv and assume that there is an ordinary deductible of 5 applied to the loss. The following table is a typical example illustrating how Y_L and Y_P works.

X	3	2	5	7	9	10
Y_L	0	0	0	2	4	5
Y_P	NA	NA	NA	2	4	5

Table 10.1: Example illustrating the relationship between Y_P and Y_L.

11 Lecture 11 Oct 18th

11.1 Severity Distribution — Policy Adjustments (Continued 2)

11.1.1 Distribution & Moments of Y_P and Y_L

It suffices for us to closely study Y_L due to the following proposition:

Proposition 13 (Y_P is completely determined by Y_L)

The survival function and moments of Y_P are given by

$$\bar{F}_{Y_P}(y) = \begin{cases} 1 & y < 0 \\ \frac{\bar{F}_{Y_L}(y)}{\bar{F}_{Y_L}(0)} & y \geq 0 \end{cases}$$

and

$$E\left[Y_P^k\right] = \frac{E\left[Y_L^k\right]}{\bar{F}_{Y_L}(0)}, \quad k = 1, 2, \ldots.$$

Proof

Using the definition of a survival function, we have

$$\bar{F}_{Y_P}(x) = P(Y_P > x) = P(Y_L > x \mid Y_L > 0) = \begin{cases} 1 & x < 0 \\ \frac{\bar{F}_{Y_L}(x)}{\bar{F}_{Y_L}(0)} & x \geq 0 \end{cases}.$$

Consequently,

$$e\left[Y_P^k\right] = k \int_0^\infty x^{k-1} \bar{F}_{Y_P}(x)\, dx = k \int_0^\infty x^{k-1} \frac{\bar{F}_{Y_L}(x)}{\bar{F}_{Y_L}(0)}\, dx = \frac{E\left[Y_L^k\right]}{\bar{F}_{Y_L}(0)}.$$

□

Note

Proposition 13 tells us that it suffices to discover the distribution of Y_L, since it completely determines Y_P,

Remark

- *If there is a deductible $d > 0$, then the distributions of Y_P and Y_L are **usually**[1] different.*
- *If Y_L has no mass point at 0, i.e. $\bar{F}_{Y_L}(0) = 1$, then Y_P nand Y_L have the same distribution.*

[1] This depends on the distribution of X and Y_L.

11.1.2 Some Important Identities

The following proposition is important for us to venture forward.

Proposition 14 (★★ Expected Value of the Policy Adjustments)

Consider a non-negative rv X and $d > 0$. Then

1. *We have*
$$E[X] = E[[X-d]_+] + E[X \wedge d].$$
2. *For $k = 1, 2, \ldots,$*
$$E\left[X^k\right] = \int_0^\infty x^{k-1}\bar{F}_X(x)\,dx.$$
3. *For $k = 1, 2, \ldots,$*
$$E\left[(X \wedge d)^k\right] = \int_0^d kx^{k-1}\bar{F}_X(x)\,dx.$$
4. *For $k = 1, 2, \ldots,$*
$$E\left[[X-d]_+^k\right] = \int_d^\infty k(x-d)^{k-1}\bar{F}_X(x)\,dx.$$
5. *For $k = 1, 2, \ldots,$ and $\bar{F}_X(d) > 0$,*
$$E\left[(X-d)^k \mid X > d\right] = \frac{E\left[[X-d]_+^k\right]}{\bar{F}_X(d)} = \frac{\int_d^\infty k(x-d)^{k-1}\bar{F}_X(x)\,dx}{\bar{F}_X(d)}.$$

In particular, we have an alternate way to derive the mean excess value

$$e_X(d) = E\left[X - d \mid X > d\right] = \frac{\int_d^\infty \overline{F}_X(x)\,dx}{\overline{F}_X(d)}.$$

Proof

1. For this identity, notice that

$$[X-d]_+ + (X \wedge d) = \begin{cases} 0 + X & X \leq d \\ X - d + d & X > d \end{cases} = X.$$

The result follows from linearity of E.

2. We have proved this earlier on, but it shall be re-proved for exercise, variety, and ease of reference.[2]

$$\begin{aligned} E\left[X^k\right] &= \int_0^\infty x^k f_X(x)\,dx = \int_0^\infty x^k \frac{d}{dx} F_X(x)\,dx \\ &= \int_0^\infty x^k dF_X(x) = \int_0^\infty x^k d(1 - \overline{F}_X(x)) \\ &= -\int_0^\infty x^k d\overline{F}_X(x) \\ &= -x^k \overline{F}_X(x)\Big|_0^\infty + k\int_0^\infty x^{k-1}\overline{F}_X(x)\,dx \quad \because \text{ IBP} \\ &= k\int_0^\infty x^{k-1}\overline{F}_X(x)\,dx, \end{aligned}$$

under the assumption that $\overline{F}_X(x)$ decays faster than x^k [3].

3. Using a similar argument as in the earlier part of the last proof, and by the Law of the Unconscious Statistician, we can arrive at

$$E\left[(X \wedge u)^k\right] = -\int_0^\infty (x \wedge u)^k \, d\overline{F}_X(x).$$

To proceed, use integration by parts as follows[4]:

$$\begin{aligned} u &= (x \wedge u)^k & v &= \overline{F}_X(x) \\ du &= d(x \wedge u)^k & dv &= d\overline{F}_X(x) \end{aligned}$$

We get

$$E\left[(X \wedge u)^k\right] = -(x \wedge u)^k \overline{F}_X(x)\Big|_0^\infty + \int_0^\infty \overline{F}_X(x)\,d(x \wedge u)^k,$$

[2] **Important**: There are two rules that you must use, and it does not depend on any of your existing knowledge as an undergrad whatsoever.
(a) Notice that $\frac{d}{dx}F_X(x) = f_X(x) \implies -d\overline{F}_X(x) = f_X(x)\,dx$;
(b) While using integration by parts, let $dv = d\overline{F}_X(x)$ so that $v = \overline{F}_X(x)$.
Forget any one of these are prepare to be screwed over.

[3] How unlikely is this, I do not know.

[4] This is hopeless. **If you can't remember this**, or somehow make some sense of this monster (without going through a few lectures on Lesbesgue or Riemann-Stieljes integration), **you're screwed.**

and $(x \wedge u)^k \bar{F}_X(x) \Big|_0^\infty = 0$. Next, it is a "fact" that

$$d(x \wedge u)^k = \begin{cases} kx^{k-1}\,dx & x < u \\ 0 & x \geq u \end{cases}$$

[5]Since $x > u$ gives us a 0 term, we are left with

$$E\left[(X \wedge u)^k\right] = \int_0^u kx^{k-1}\bar{F}_X(x)\,dx.$$

4. Using the Law of the Unconscious Statistician and Item 2, we have

$$E\left[[X-d]_+^k\right] = k\int_0^\infty (x-d)_+^k \bar{F}_X(x)\,dx = k\int_d^\infty (x-d)^k \bar{F}_X(x)\,dx.$$

5. Notice once and for all that

$$\begin{aligned} E\left[(X-d)^k \mid X > d\right] &= \frac{E\left[(X-d)^k \mathbb{1}_{\{X>d\}}\right]}{\bar{F}_X(d)} \\ &= \frac{E\left[[X-d]_+^k\right]}{\bar{F}_X(d)} = \frac{\int_d^\infty k(x-d)^k \bar{F}_X(x)\,dx}{\bar{F}_X(d)}. \end{aligned}$$

□

[5] This is yet another monstrosity that makes no sense whatsoever for one that has only taken basic Calculus classes and introductory Analysis. **Remember this "fact"** or **get screwed over**. The only "sense" that I can come up with right now is to consider the cases of when $x < u$ and when $x \geq u$, and determine what $x \wedge u$ should be in these cases, and provide some baseless rationalization.

Example 11.1.1

Consider a ground-up loss X with pdf

$$f_X(x) = 0.0005, \quad 0 \leq x \leq 20.$$

Solve for

1. $\bar{F}_X(x)$;
2. $E[X \wedge 10]$ and $E[X \wedge 25]$;
3. $\mathrm{Var}(X \wedge 10)$;
4. $E\left[[X-10]_+\right]$;
5. $E\left[[X-10]_+^2\right]$; and
6. $e_X(10)$.

Solution

1. We have

$$F_X(x) = \int_0^x 0.005y\,dy = 0.0025x^2$$

and so

$$\bar{F}_X(x) = \begin{cases} 1 & x < 0 \\ 1 - 0.0025x^2 & 0 \leq x \leq 20 \\ 0 & x > 20 \end{cases}.$$

2. Using our identities,

$$E[X \wedge 10] = \int_0^{10} \bar{F}_X(x)\,dx = 10 - \frac{5}{6000}(10)^3 = \frac{55}{6},$$

and

$$E[X \wedge 25] = \int_0^{20} \bar{F}_X(x)\,dx = 25 - \frac{5}{6000}(25)^3 = \frac{40}{3}.$$

3. To get $\mathrm{Var}(X \wedge 10)$, we first need the 2nd moment of $X \wedge 10$:

$$\begin{aligned} E\left[(X \wedge 10)^2\right] &= 2\int_0^{10} x\bar{F}_X(x)\,dx = 2\left[\frac{1}{2}x^2 - \frac{5}{8000}x^4\right]_0^{10} \\ &= 100 - \frac{25}{2} = \frac{175}{2}. \end{aligned}$$

Thus

$$\mathrm{Var}\,(X \wedge 10) = \frac{175}{2} - \left(\frac{55}{6}\right)^2 = \frac{125}{36}.$$

4. We have that

$$\begin{aligned} E\left[[X-10]_+\right] &= \int_{10}^{20} 1 - \frac{1}{400}x^2\,dx \\ &= 10 - \frac{1}{1200}(8000 - 1000) = \frac{25}{6}. \end{aligned}$$

5. We have that

$$\begin{aligned} E\left[[X-10]_+^2\right] &= 2\int_{10}^{20} (x-10)\left(1 - \frac{1}{400}x^2\right)dx \\ &= 2\int_{10}^{20}\left(-10 + x + \frac{1}{40}x^2 - \frac{1}{400}x^3\right)dx \\ &= 2\left[-100 + \frac{300}{2} + \frac{7000}{120} - \frac{150000}{1600}\right] \\ &= \frac{175}{6} \end{aligned}$$

6. We have

$$e_X(10) = \frac{\int_{10}^{20} \left(1 - \frac{1}{400}x^2\right) dx}{1 - \frac{1}{400}(10)^2} = \frac{10 - \frac{7000}{1200}}{\frac{3}{4}} = \frac{50}{9}$$

11.1.2.1 Application of Proposition 14

Policy Limit If a policy limit u is the only adjustment in a contract, then

$$Y_L = X \wedge u \quad \text{and} \quad Y_P = X \wedge u \mid X > 0.$$

Since in most cases $X > 0$, we have that $\bar{F}_X(0) = 1$, and so $Y_L = Y_P = X \wedge u$.

- The survival function is

$$\bar{F}_{Y_P}(y) = \bar{F}_{Y_L}(y) = P(X \wedge u > y) = \begin{cases} 1 & y < 0 \\ \bar{F}_X(y) & 0 \le y < u \\ 0 & y \ge u \end{cases}$$

- The expected value is

$$E[Y_P] = E[Y_L] = E[X \wedge u] = \int_0^u \bar{F}_X(x)\, dx.$$

- The second moment is

$$E\left[Y_P^2\right] = E\left[Y_L^2\right] = E\left[(X \wedge u)^2\right] = 2\int_0^u x\bar{F}_X(x)\, dx$$

Ordinary Deductible If an ordinary deductible d is the only adjustmnet in a contract, then

$$Y_L = [X - d]_+ = (X - d)\mathbb{1}_{\{X > d\}}$$

and

$$Y_P = [X - d]_+ \mid [X - d]_+ > 0 = (X - d)\mathbb{1}_{\{X > d\}} \mid X > d = X - d \mid X > d$$

In most cases, since $\bar{F}_X(d) = P(X > d) < 1$ as it is $P(X < d) \neq 0$, the distribution of Y_L and Y_P differs.

- The survival function is

$$\bar{F}_{Y_L}(y) = P([X-d]_+ > y) = \begin{cases} 1 & y < 0 \\ \bar{F}_X(y+d) & y \geq 0 \end{cases}$$

and

$$\bar{F}_{Y_P}(y) = \frac{\bar{F}_{Y_L}(y)}{\bar{F}_{Y_L}(0)} = \begin{cases} 1 & y < 0 \\ \frac{\bar{F}_X(d+y)}{\bar{F}_X(d)} & y \geq 0 \end{cases}$$

for Y_L and Y_P respectively.

- The mean is

$$E[Y_L] = E[[X-d]_+] = \int_d^\infty \bar{F}_X(x)\,dx$$

and

$$E[Y_P] = E[X-d \mid X > d] = \frac{E[(X-d)\mathbb{1}_{\{X>d\}}]}{\bar{F}_X(d)} = \frac{\int_d^\infty \bar{F}_X(x)\,dx}{\bar{F}_X(d)}$$

for Y_L and Y_P respectively.

- The second moment is

$$E\left[Y_L^2\right] = E\left[[X-d]_+^2\right] = 2\int_d^\infty x\bar{F}_X(x)\,dx$$

and

$$E\left[Y_P^2\right] = \frac{E\left[Y_L^2\right]}{\bar{F}_{Y_L}(0)} = \frac{2\int_d^\infty x\bar{F}_X(x)\,dx}{\bar{F}_X(d)}$$

for Y_L and Y_P respectively.

12 Lecture 12 Oct 23rd

12.1 Severity Distribution — Policy Adjustments (Continued 3)

12.1.1 Some Important Identities (Continued)

12.1.1.1 Application of Proposition 14 (Continued)

Policy Limit + Ordinary Deductible If there is a policy limit u and an ordinary deductible d with $u > d$, then

$$Y_L = [(X \wedge u) - d]_+ = \begin{cases} 0 & X < d \\ X - d & d \leq X < u \\ u - d & X \geq u \end{cases}$$

and so its survival function is[1]

$$\begin{aligned} \bar{F}_{Y_L}(y) &= P(Y_L > y) = P([(X \wedge u) - d]_+ > y) \\ &= P(\max\{0, (X \wedge u) - d\} > y) \\ &= \begin{cases} 1 & y < 0 \\ P((X \wedge u) - d > y) & y \geq 0 \end{cases} \\ &= \begin{cases} 1 & y < 0 \\ P(X > y + d) & 0 \leq y < u - d \\ P(u > y + d) & y \geq u - d \end{cases} \\ &= \begin{cases} 1 & y < 0 \\ \bar{F}_X(y + d) & 0 \leq y < u - d \\ 0 & y \geq u - d \end{cases} \end{aligned}$$

[1] **Important**: Pay attention to the notion here: we are using the actual definition of a deductible to arrive at an explicit solution.

Consequently, its moments are

$$\begin{aligned} E\left[Y_L^k\right] &= \int_0^\infty k y^{k-1} \bar{F}_{Y_L}(y)\,dy \\ &= \int_0^{u-d} k y^{k-1} \bar{F}_X(y+d)\,dy \\ &= \int_d^u k(y-d)^{k-1} \bar{F}_X(y)\,dy \end{aligned}$$

For Y_P, we have

$$\begin{aligned} Y_P = Y_L \mid Y_L > 0 &= [(X \wedge u) - d]_+ \mid [(X \wedge u) - d]_+ > 0 \\ &= (X \wedge u) - d \mid X > d. \end{aligned}$$

Thus by Proposition 13,

$$\bar{F}_{Y_P}(y) = \begin{cases} 1 & y < 0 \\ \frac{\bar{F}_{Y_L}(y)}{\bar{F}_{Y_L}(0)} & 0 \leq y < u-d \\ 0 & y \geq u-d \end{cases}$$

Using the same proposition, the moments of Y_P are

$$E\left[Y_P^k\right] = \frac{E\left[Y_L^k\right]}{\bar{F}_X(d)} = \frac{\int_d^u k(y-d)^{k-1}\bar{F}_X(y)\,dy}{\bar{F}_X(d)}.$$

Franchise Deductible Given a Franchise Deductible $d > 0$, we have

$$Y_L = X \cdot \mathbb{1}_{\{X>d\}}.$$

Its survival function is

$$\begin{aligned} \bar{F}_{Y_L}(y) = P(X \cdot \mathbb{1}_{\{X>d\}} > y) &= \begin{cases} 1 & y < 0 \\ P(X \cdot \mathbb{1}_{\{X>d\}} > y) & y \geq 0 \end{cases} \\ &= \begin{cases} 1 & y < 0 \\ \bar{F}_X(d) & 0 \leq y < d \\ \bar{F}_X(y) & y \geq d \end{cases}. \end{aligned}$$

For the case when $0 \leq y < d$, it is clear that in order for $X \cdot \mathbb{1}_{\{X>d\}} > y$, we first need $X > d$. Now the moments of Y_L are

$$\begin{aligned} E\left[Y_L^k\right] &= \int_0^\infty k y^{k-1} \bar{F}_{Y_L}(y)\, dy \\ &= \bar{F}_X(d) \int_0^d k y^{k-1}\, dy + \int_d^\infty k y^{k-1} \bar{F}_X(y)\, dy \\ &= d^k \bar{F}_X(d) + \int_d^\infty k y^{k-1} \bar{F}_X(y)\, dy. \end{aligned}$$

Now observe that

$$Y_P = Y_L \mid Y_L > 0 = X \cdot \mathbb{1}_{\{X>d\}} \mid X \cdot \mathbb{1}_{\{X>d\}} > 0 = X \mid X > d$$

Again, using Proposition 13,

$$\bar{F}_{Y_P}(y) = \frac{\bar{F}_{Y_L}(y)}{\bar{F}_{Y_L}(0)} = \begin{cases} 1 & y < d \\ \frac{\bar{F}_X(y)}{\bar{F}_X(d)} & y \geq d \end{cases}$$

and its moments are

$$\begin{aligned} E\left[Y_P^k\right] &= \frac{E\left[Y_L^k\right]}{\bar{F}_X(d)} = \frac{d^k \bar{F}_X(d) + \int_d^\infty k y^{k-1} \bar{F}_X(y)\, dy}{\bar{F}_X(d)} \\ &= d^k + \frac{\int_d^\infty k y^{k-1} \bar{F}_X(y)\, dy}{\bar{F}_X(d)}. \end{aligned}$$

Coinsurance Let $\hat{Y}_L$ be the amount to-be-paid per loss without coinsurance. Let $\alpha \in (0,1)$ be a coinsurance factor. Applying coinsurance for adjustment, we have that

$$Y_L = \alpha \hat{Y}_L.$$

The survival function of Y_L is

$$\bar{F}_{Y_L}(y) = P\left(\alpha \hat{Y}_L > y\right) = P\left(\hat{Y}_L > \frac{y}{\alpha}\right) = \begin{cases} 1 & y < 0 \\ \bar{F}_{\hat{Y}_L}\left(\frac{y}{\alpha}\right) & y \geq 0 \end{cases}$$

and its moments are

$$E\left[Y_L^k\right] = E\left[\alpha^k \hat{Y}_L{}^k\right] = \alpha^k E\left[\hat{Y}_L{}^k\right].$$

Then for Y_P, we have that its survival function is

$$\bar{F}_{Y_P}(y) = \frac{\bar{F}_{Y_L}(y)}{\bar{F}_{Y_L}(0)} = \frac{\bar{F}_{\hat{Y}_L}\left(\frac{y}{\alpha}\right)}{\bar{F}_{\hat{Y}_L}(0)}, \quad y \geq 0,$$

and its moments

$$E\left[Y_P^k\right] = \frac{E\left[Y_L^k\right]}{\bar{F}_{\hat{Y}_L}(0)} = \frac{\alpha^k E\left[\hat{Y}_L{}^k\right]}{\bar{F}_{\hat{Y}_L}(0)}.$$

Example 12.1.1

The cdf of a ground-up loss X is given by

$$F_X(x) = 1 - \left(1 - \frac{x}{800}\right)^2, \quad 0 \le x \le 800.$$

Assuming a policy limit of 600, an ordinary deductible of 200, and a coinsurance facotr of 0.8, determine the cdf of Y_L and the expected amount paid per loss $E\left[Y_L\right]$.

Solution

We are given

$$Y_L = 0.8[(X \wedge 600) - 200]_+.$$

The survival function of Y_L is

$$\begin{aligned}
\bar{F}_{Y_L}(y) &= P\left(\max\{0, (X \wedge 600) - 200\} > \frac{y}{0.8}\right) \\
&= \begin{cases} 1 & y < 0 \\ P\left(X \wedge 600 > \frac{5y}{4} + 200\right) & y \ge 0 \end{cases} \\
&= \begin{cases} 1 & y < 0 \\ P\left(X > \frac{5y}{4} + 200\right) & 0 \le y < 0.8(600 - 200) = 320 \\ 0 & y \ge 320 \end{cases} \\
&= \begin{cases} 1 & y < 0 \\ \bar{F}_X\left(\frac{5y}{4} + 200\right) & 0 \le y < 320 \\ 0 & y \ge 320 \end{cases}
\end{aligned}$$

Thus the cdf of Y_L is

$$\bar{F}_{Y_L}(y) = \begin{cases} 0 & y < 0 \\ 1 - \bar{F}_X\left(\frac{5y}{4} + 200\right) & 0 \le y < 320 \\ 1 & y \ge 320 \end{cases}.$$

The expected amount paid per loss is

$$E\left[Y_L\right] = 0.8 \int_{200}^{600} \left(1 - \frac{x}{800}\right)^2 dx = \frac{260}{3}$$

Example 12.1.2

Consider a ground-up loss X with a Franchise deductible d. You are given that

- 15% of the losses are below the Franchise deductible d,
- the mean excess loss $e_X(d) = 50$,
- the expected amount paid per loss is 51.

Determine the value of d.

Solution

We are given

$$\begin{aligned} Y_L &= X \cdot \mathbb{1}_{\{X>d\}} \\ P(X < d) &= 0.15 \end{aligned}$$

Thus $P(X > d) = 0.85$. We have

$$\bar{F}_{Y_L}(y) = \begin{cases} 1 & y < 0 \\ \bar{F}_X(d) & 0 \leq y < d \\ \bar{F}_X(y) & y \geq d \end{cases} = \begin{cases} 1 & y < 0 \\ 0.85 & 0 \leq y < d \\ \bar{F}_X(y) & y \geq d \end{cases}.$$

We are given

$$50 = e_X(d) = \frac{\int_d^\infty \bar{F}_X(y)\,dy}{\bar{F}_X(d)},$$

and so

$$\int_d^\infty \bar{F}_X(y)\,dy = 50 * 0.85 = 42.5.$$

We are given

$$\begin{aligned} 51 = E[Y_L] &= \int_0^\infty \bar{F}_X(y)\,dy = d\bar{F}_X(d) + \int_d^\infty \bar{F}_X(y)\,dy \\ &= 0.85d + 42.5. \end{aligned}$$

Thus

$$d = \frac{8.5}{0.85} = 10.$$

13 Lecture 13 Oct 25th

13.1 Severity Distribution — Policy Adjustments (Continued 3)

By introducing policy adjustments, it is within our interest to determine if the introduced adjustments have helped to eliminate the expected propotion of loss.

Definition 39 (Loss Elimination Ratio)

*The **loss elimination ratio**, denoted as LER, is the ratio of which loss has been mitigated, or eliminated, as a result of policy adjustments, and it is given by*

$$\mathrm{LER} = \frac{E[X - Y_L]}{E[X]} = 1 - \frac{E[Y_L]}{E[X]},$$

where $\frac{E[Y_L]}{E[X]}$ corresponds to the percentage of loss retained by the insurer.

Example 13.1.1

For a policy that has only an ordinary deductible, i.e. $Y_L = [X - d]_+$, we have

$$\mathrm{LER} = 1 - \frac{E([X-d]_+)}{E[X]} = 1 - \frac{E[X] - E[X \wedge d]}{E[X]} = \frac{E[X \wedge d]}{E[X]}.$$

Example 13.1.2

Consider a ground-up loss $X \sim \mathrm{Pareto}(\alpha, \theta)$ with $\alpha = 2$ and $\theta = 1000$.

1. Calculate the LER if an ordinary deductible of 500 is applied.

2. What is the required value of d to eliminate 20% of the loss?

Solution

1. Note that
$$\bar{F}_X(x) = \frac{\theta^\alpha}{(x+\theta)^\alpha}.$$
Now
$$\begin{aligned} E[X] &= \int_0^\infty \bar{F}_X(x)\,dx = \theta^\alpha \int_0^\infty \frac{1}{(x+\theta)^\alpha}\,dx \\ &= \frac{\theta^\alpha}{1-\alpha} \cdot \frac{1}{(x+\theta)^{\alpha-1}}\bigg|_0^\infty = \frac{\theta}{\alpha - 1}. \end{aligned}$$
and
$$\begin{aligned} E[X \wedge d] &= \int_0^d \bar{F}_X(x)\,dx = \frac{\theta^\alpha}{1-\alpha} \cdot \frac{1}{(x+\theta)^{\alpha-1}}\bigg|_0^d \\ &= \frac{\theta^\alpha}{1-\alpha}\left(\frac{1}{(d+\theta)^{\alpha-1}} - \frac{1}{\theta^{\alpha-1}}\right) \\ &= \frac{\theta}{\alpha-1}\left(1 - \left(\frac{\theta}{d+\theta}\right)^{\alpha-1}\right). \end{aligned}$$
Thus
$$\mathrm{LER} = \frac{E[X \wedge d]}{E[X]} = \left(1 - \left(\frac{\theta}{d+\theta}\right)^{\alpha-1}\right) = \frac{1}{3}.$$
In other words, $\frac{1}{3}$ is mitigated by setting an ordinary deductible of 500.

2. In this case, let $\mathrm{LER} = 0.2 = \frac{1}{5}$. Then
$$\frac{1}{5} = 1 - \frac{1000}{d+1000} \iff \frac{1000}{d+1000} = \frac{4}{5} \iff d = 250$$

13.2 Frequency Distributions — Basic Frequency Distributions

Recall from our Collective Risk Model that
$$S = \sum_{i=1}^N X_i,$$
where

$X_i \equiv$ size of the i^{th} claim, modelled by severity distributions

$N \equiv$ a nonnegative integer-valued rv that represents the number of claims, modelled by frequency distributions

Definition 40 (Counting Distributions and RVs)

A nonnegative rv, usually represented by N, is called a ***counting rv*** *and its distribution is called a* ***counting distirbution****.*

Note

For this section, the pgf is important.

Importance of PGF Given $G(t) = E\left[t^N\right] = \sum_{k=0}^{\infty} t^K p_k$, provided that the moments exist, we have

$$
\begin{aligned}
G^{(n)}(t) = \frac{d^n}{dt^n} G(t) &\overset{(*)}{=} E\left[\prod_{i=1}^{n}(N-i+1)t^{N-n}\right] \\
&= \sum_{k=0}^{\infty}\prod_{i=1}^{n}(k-i+1)t^{k-n}p_k \\
&\overset{(**)}{=} \sum_{k=n}^{\infty}\prod_{i=1}^{n}(k-i+1)t^{k-n}p_k
\end{aligned}
$$

where $(*)$ is because the moments exist, and $(**)$ is because for $k = 0, 1, \ldots, n-1$, the product $\prod_{i=1}^{n}(k-i+1) = 0$.

Also for $(*)$: The derivation is

$$
\begin{aligned}
G_N^{(n)}(t) &= \frac{d^n}{dt^n} G_N(t) \\
&= \sum_{k=0}^{\infty} \frac{d^n}{dt^n} t^k p_k \\
&= \sum_{k=0}^{\infty} k(k-1)\ldots(k-n+1)t^{k-n}p_k \\
&= E\left[\prod_{k=1}^{n}(N-k+1)t^{N-n}\right]
\end{aligned}
$$

We can obtain the pmf of N from the pgf by

$$
\begin{aligned}
G^{(n)}(0) &= \sum_{k=n}^{\infty}\prod_{i=1}^{n}(k-i+1)t^{k-n}p_k\Big|_{t=0} \\
&= \prod_{i=1}^{n}(n-i+1)p_n = n!p_n
\end{aligned}
\tag{13.1}
$$

where we notice in Equation (13.1) that only the n^{th} term survives as $t^{n-n} = 1$.

Factorial Moments[1] can be obtained by

$$
G^{(n)}(1) = E\left[\prod_{i=1}^{n}(N-i+1)\right], \quad n = 1, 2, 3, \ldots.
$$

In particular, we have that

$$
G'(1) = E[N] \text{ and } G''(1) = E[N(N-1)] = E(N^2) - E(N)
$$

and so

$$
\operatorname{Var}(N) = G''(1) + G'(1) - G'(1)^2.
$$

[1] **Definition 41 (Factorial Moments from PGF)**
We can obtain the ***factorial moments*** *of an rv X from its pgf. In particular,*

$$
G^{(n)}(1) = E\left[\prod_{i=1}^{n}(X-i+1)\right]
$$

where $n \in \mathbb{N} \setminus \{0\}$.

13.2.1 Frequency Distributions

13.2.1.1 Poisson Distribution

Definition 42 (Poisson Distribution)

*A counting rv N is said to have a **Poisson distribution** with parameter λ, and denote $N \sim \mathrm{Poi}(\lambda)$, if it has the pmf*

$$p_k = P(N = k) = \frac{e^{-\lambda}\lambda^k}{k!}, \quad k = 0, 1, 2, \ldots.$$

Remark

We can easily verify that the Poisson distribution is indeed a probability distribution, by noticing that

$$\sum_{k=0}^{\infty} p_k = \sum_{k=0}^{\infty} \frac{e^{-\lambda}\lambda^k}{k!} = e^{-\lambda}e^{\lambda} = 1$$

*where we used the **Taylor expansion** $e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$.*

Proposition 15 (PGF, Mean, and Variance of Poisson Distribution)

For $N \sim \mathrm{Poi}(\lambda)$, its pgf is

$$G(t) = e^{\lambda(t-1)},$$

and its mean and variance are

$$E(N) = \mathrm{Var}(N) = \lambda.$$

Proof

Notice that

$$G(t) = E\left[t^N\right] = \sum_{k=0}^{\infty} t^k p_k = \sum_{k=0}^{\infty} \frac{e^{-\lambda}(t\lambda)^k}{k!} = e^{-\lambda}e^{t\lambda} = e^{\lambda(t-1)}.$$

Thus

$$E[N] = G'(1) = \lambda \text{ and } G''(1) = \lambda^2,$$

and so

$$\mathrm{Var}(N) = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

□

Proposition 16 (Sum of Independent Poisson RVs)

If $N_1, N_2, \ldots, N_m$ are ***independent*** *Poisson rvs with parameters $\lambda_1, \lambda_2, \ldots, \lambda_m$ respectively, then*

$$N = \sum_{i=1}^{m} N_i \sim \mathrm{Poi}\left(\sum_{i=1}^{m} \lambda_i\right)$$

Proof

Using the pgf method for N, we see that

$$\begin{aligned} G(t) &= e\left[t^N\right] = E\left[t^{\sum_{i=1}^m N_i}\right] = \prod_{i=1}^{m} E\left[t^{N_i}\right] \\ &= \prod_{i=1}^{m} e^{\lambda_i(t-1)} = e^{(t-1)\sum_{i=1}^m \lambda_i} \end{aligned}$$

which is the pgf of $\mathrm{Poi}\left(\sum_{i=1}^m \lambda_i\right)$ as required. □

14 Lecture 14 Oct 30th

14.1 Frequency Distribution — Basic Frequency Distributions (Continued)

14.1.1 Frequency Distributions (Continued)

14.1.1.1 Poisson Distribution (Continued)

Proposition 17 (Splitting a Poisson Distribution)

Suppose that the total number of claim arrivals follows $N \sim \mathrm{Poi}(\lambda)$*. There are* m *distinct types of claims. Given a claim occurs, it is of type* i *with probability* p_i *such that*

$$p_1 + \ldots + p_m = 1.$$

Then, for each fixed $i = 1, \ldots, m$*, the number of claims of type* i*,* $N_i \sim \mathrm{Poi}(\lambda p_i)$*. Furthermore,* $N_1, N_2, \ldots, N_m$ *are independent.*

Note

The above proposition can be visualized using a tree.

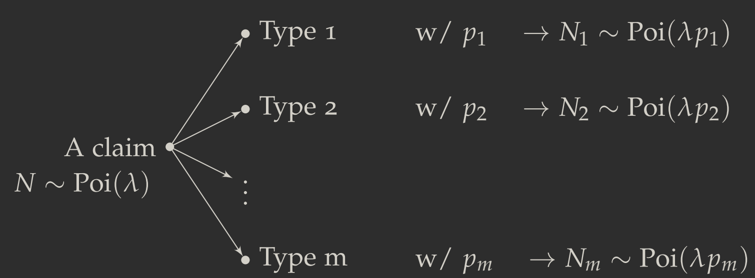


Figure 14.1: Visualization of Proposition 17

Proof

We shall use **mathematical induction** on m, for the statement "$N_1, N_2, \ldots, N_m$ are independent". $N_i \sim \mathrm{Poi}(\lambda p_i)$ will follow from the induction step.

For $m = 1$, there is nothing to prove. It suffices to prove for $m = 2$, since we may think of the problem as

$$\text{Type 1} \quad \underbrace{\text{Type 2} \quad \text{Type 3} \ldots \text{Type m}}_{\text{Type 2}'}.$$

Case $m = 2$ Suppose $N = N_1 + N_2 \sim \mathrm{Poi}(\lambda)$. To show that N_1 and N_2 are independent, a relation which we denote as $N_1 \perp N_2$, we need to show

$$P(N_1 = k_1, N_2 = k_2) = P(N_1 = k_1)P(N_2 = k_2), \tag{14.1}$$

which is a defining property of independence.

Firstly, note that if given sets $A \subset B$, we have

$$P(A) = P(A \cap B).$$

With that,

$$\begin{aligned}
P(N_1 = k_1, N_2 = k_2) &= P(\overbrace{N_1 = k_1, N_2 = k_2}^{A}, \overbrace{N_1 + N_2 = k_1 + k_2}^{B}) \\
&= P(A \mid B)P(B) \\
&= \binom{k_1 + k_2}{k_1} p_1^{k_1} p_2^{k_2} \cdot \frac{e^{-\lambda} \lambda^{k_1 + k_2}}{(k_1 + k_2)!} \\
&= \frac{(k_1 + k_2)!}{k_1! k_2!} p_1^{k_1} p_2^{k_2} \cdot \frac{e^{-\lambda(1)} \lambda^{k_1 + k_2}}{(k_1 + k_2)!} \\
&= e^{-\lambda(p_1 + p_2)} \cdot \frac{(\lambda p_1)^{k_1}}{k_1!} \cdot \frac{(\lambda p_2)^{k_2}}{k_2!} \\
&= \frac{e^{-\lambda p_1} (\lambda p_1)^{k_1}}{k_1!} \cdot \frac{e^{-\lambda p_2} (\lambda p_2)^{k_2}}{k_2!}
\end{aligned} \tag{14.2}$$

Thus, the marginal distribution of N_1

$$\begin{aligned} P(N_1 = k_1) &= \sum_{k_2=0}^{\infty} P(N_1 = k_1, N_2 = k_2) \\ &= \frac{e^{-\lambda p_1}(\lambda p_1)^{k_1}}{k_1!} e^{-\lambda p_2} \sum_{k_2=0}^{\infty} \frac{(\lambda p_2)^{k_2}}{k_2!} \\ &= \frac{e^{-\lambda p_1}(\lambda p_1)^{k_1}}{k_1!} e^{-\lambda p_2} e^{\lambda p_2} \\ &= \frac{e^{-\lambda p_1}(\lambda p_1)^{k_1}}{k_1!} \end{aligned}$$

which is the pmf of $\mathrm{Poi}(\lambda p_1)$. The marginal distribution of N_2 is similar. It is clear from Equation (14.2) that we have Equation (14.1). The result then follows from induction. □

Example 14.1.1

The number of claims of a portfolio follows $\mathrm{Poi}(\lambda)$. The severity of ground-up loss follows $\mathrm{Unif}(0, b)$. The insurer would like to impose an ordinary deductible d and a policy limit u such that

$$0 < d < u < b.$$

What is the frequency distribution of **positive payments**?

Solution

Let Type 1 be the case where $X < d$ and Type 2 be $X > d$. Since the severity of the ground-up loss follows $\mathrm{Unif}(0, b)$, the probability of an occurrence of Type 2 is

$$1 - \frac{d}{b} = \frac{b-d}{b}.$$

By Proposition 17, we have that the frequency distribution of positive payments, i.e. Type 2, follows $\mathrm{Poi}\left(\lambda \frac{b-d}{b}\right)$.

14.1.1.2 *Binomial Distribution*

Definition 43 (Binomial Distribution)

*A counting rv N is said to have a **binomial distribution** with parameters $q \in (0, 1)$ and $m \in \mathbb{N} \setminus \{0\}$, written as $N \sim \mathrm{Bin}(q, m)$, if it has the* *pmf*

$$p_k = P(N = k) = \binom{m}{k} q^k (1-q)^{m-k}, \; k = 0, 1, 2, \ldots.$$

Remark

It is easy to verify that this is a valid probability distribution, since

$$\sum_{k=0}^{m} \binom{m}{k} q^k (1-q)^{m-k} = (q + 1 - q)^m = 1^m = 1$$

by the ***binomial theorem****.*

Note

- *When* $m = 1$*, the distribution is called a* ***Bernoulli*** *rv with mean* q*.*
- *The binomial distribution is a* ***bounded*** *rv, since it has a fixed number of trials.*

Proposition 18 (PGF of Binomial Distribution)

Let $N \sim \mathrm{Bin}(q, m)$*. Its pgf is given by*

$$G(t) = (1 - q + tq)^m.$$

Moreover, its mean and variance are

$$E[N] = mq \text{ and } \mathrm{Var}(N) = mq(1-q)$$

respectively.

Proof

We have

$$G(t) = \sum_{k=0}^{m} \binom{m}{k} (tq)^k (1-q)^{m-k} = (1 - q + tq)^m$$

The mean is, therefore,

$$G'(1) = mq(1 - q + (1)q)^{m-1} = mq,$$

and its variance

$$\begin{aligned}\mathrm{Var}(N) &= G''(1) + G'(1) - G'(1)^2 \\ &= m(m-1)q^2 + mq - m^2q^2 = mq(1-q).\end{aligned}$$

□

Proposition 19 (Sum of Independent Binomial RVs)

If $N_1, \ldots, N_n$ are independent and $N_i \sim \mathrm{Bin}(q, m_i)$ for $i = 1, \ldots, n$, then

$$N = \sum_{i=1}^{n} N_i \sim \mathrm{Bin}\left(q, \sum_{i=1}^{n} m_i\right).$$

Proof

We shall use the pgf to prove this, instead of using the mgf (which is the common approach).

$$\begin{aligned}G_N(t) = E\left[t^N\right] = E\left[t^{\sum_{i=1}^{n} N_i}\right] &\overset{(*)}{=} \prod_{i=1}^{n} E\left[t^{N_i}\right] \\ &= \prod_{i=1}^{n}(1-q+tq)^{m_i} = (1-q+tq)^{\sum\limits_{i=1}^{n} m_i}\end{aligned}$$

Thus $N = \sum_{i=1}^{n} N_i \sim \mathrm{Bin}\left(q, \sum\limits_{i=1}^{n} m_i\right)$. □

Note

As a result of Proposition 19, if we have a sequence of Bernoulli trials, each with the same "success" probability q, call each of them I_i, then

$$N = \sum_{i=1}^{m} I_i \sim \mathrm{Bin}(q, m)$$

Consequently, it becomes rather silly how easy it is we can get the mean *and variance of N:*

$$E[N] = E\left[\sum_{i=1}^{m} I_i\right] = \sum_{i=1}^{m} E[I_i] = mq$$
$$\mathrm{Var}(N) = \mathrm{Var}\left(\sum_{i=1}^{m} I_i\right) = \sum_{i=1}^{m} \mathrm{Var}(I_i) = mq(1-q)$$

15 Lecture 15 Nov 01st

15.1 Frequency Distribution — Basic Frequency Distributions (Continued 2)

15.1.1 Frequency Distributions (Continued 2)

15.1.1.1 Negative Binomial Distribution

Definition 44 (Negative Binomial Distribution)

A counting rv N is said to have a ***negatiev binomial distribution*** *with parameters $\beta > 0$ and $r > 0$, denoted $N \sim \mathrm{NB}(\beta, r)$, if it has the pmf*

$$p_k = P(N = k) = \binom{k+r-1}{k} \left(\frac{1}{1+\beta} \right)^r \left(\frac{\beta}{1+\beta} \right)^k, k = 0, 1, 2, \ldots$$

Remark

Note that

$$\binom{k+r-1}{k} = \frac{\Gamma(k+r)}{k!\Gamma(r)} = \frac{(k+r-1)!}{k!(r-1)!},$$

where the later equality follows if $r \in \mathbb{N} \setminus \{0\}$.

Note

- *When $r = 1$, we can also write the pmf of $\mathrm{NB}(\beta, 1)$ as the pmf of the* ***geometric distribution****:*

$$p_k = \frac{1}{1+\beta} \left(\frac{\beta}{1+\beta} \right)^k, k = 0, 1, 2, \ldots.$$

- *To verify that the negative binomial distribution is a valid probability* *distribution, we need the following identity:*

$$(1-x)^{-r} = \sum_{k=0}^{\infty} \binom{k+r-1}{k} x^k,$$

which is proven as follows:

Exercise 15.1.1
Verify that the negative binomial distribution is a valid probability distribution.

Proof

We shall use the **Taylor expansion** of $(1-x)^{-r}$.

$$\begin{aligned}(1-x)^{-r} &= 1 + (-1)(-r)(1-x)^{-r-1}\Big|_{x=0} x \\ &\quad + \frac{r}{2}(-1)(-r-2)(1-x)^{-r-2}\Big|_{x=0} x^2 + \ldots \\ &= 1 + rx + \frac{r(r+1)}{2}x^2 + \frac{r(r+1)(r+2)}{3!}x^3 + \ldots \\ &= \sum_{k=0}^{\infty} \frac{r(r+1)\ldots(r+k-1)}{k!} x^K \\ &= \sum_{k=0}^{\infty} \binom{k+r-1}{k} x^k.\end{aligned}$$

□

- *The negative binomial distribution is an* ***unbounded*** *rv, and can take all natural numbers sans* 0.

Interpretation Consider an experiment with independent trails, of which each has only two possible outcomes: success with probability $\frac{1}{1+\beta}$, and failure with probability $1 - \frac{1}{1+\beta} = \frac{\beta}{1+\beta}$. Let N denote the number of failures until reaching the r^{th} success.

Proposition 20 (PGF of the Negative Binomial Distribution)

Let $N \sim \mathrm{NB}(\beta, r)$. *Its pgf is thus*

$$G(t) = [1 - \beta(t-1)]^{-r}.$$

Moreover, its mean and variance are

$$E[N] = r\beta \text{ and } \mathrm{Var}(N) = r\beta(1+\beta),$$

respectively.

Note

Note that the proof for getting the pgf is similar to how we can verify that N is a probability (same case as in earlier counting distributions).

Proof

Using the Taylor Expansion $(1-x)^{-r} = \sum_{k=0}^{\infty} \binom{k+r-1}{k} x^k$, we have

$$\begin{aligned} G(t) &= \sum_{k=0}^{\infty} t^k p_k = \left(\frac{1}{1+\beta}\right)^r \sum_{k=0}^{\infty} \binom{k+r-1}{k} \left(\frac{t\beta}{1+\beta}\right)^k \\ &= \left(\frac{1}{1+\beta}\right)^r \left(1 - \frac{t\beta}{1+\beta}\right)^{-r} = [1-\beta(t-1)]^{-r}. \end{aligned}$$

Consequently, the mean is

$$E[N] = G'(1) = -r(-\beta) = r\beta$$

and variance is

$$\begin{aligned} \mathrm{Var}(N) &= G''(1) + G'(1) - G'(1)^2 \\ &= -r(-r-1)\beta^2 + r\beta - r^2\beta^2 = r\beta(1+\beta) \end{aligned}$$

□

Proposition 21 (Negative Binomial from Poisson Conditioned on Gamma)

Let $N \mid \Lambda = \lambda \sim \mathrm{Poi}(\lambda)$ *and* $\Lambda \sim \mathrm{Gam}(\alpha, \theta)$. *Then*

$$N \sim \mathrm{NB}(\theta, \alpha).$$

Note

We may also write $N \mid \Lambda = \lambda \sim \mathrm{Poi}(\lambda)$ *as* $N \mid \Lambda \sim \mathrm{Poi}(\Lambda)$.

Proof

We shall prove this statement by finding the pgf of N, which identifies the distribution. Note that

$$G_N(t) = E\left[t^N\right] \overset{\text{Proposition 9}}{=} E\left[E\left[t^N \mid \Lambda\right]\right] \overset{(*)}{=} E\left[e^{\Lambda(t-1)}\right],$$

where $(*)$ **requires further clarification**. Now since $\Lambda \sim \mathrm{Gam}(\alpha, \theta)$, and $M_\Lambda(t) = E\left[e^{t\Lambda}\right] = (1 - \theta t)^{-\alpha}$, it follows that

$$G_N(t) = [1 - \theta(t-1)]^{-\alpha}.$$

Thus $N \sim \mathrm{NB}(\theta, \alpha)$. □

Proposition 22 (Combining Negative Binomial Distributions)

If $\{N_i\}_{i=1}^n$ is a sequence of independent rvs, and $N_i \sim \mathrm{NB}(\beta, r_i)$. Then

$$N = \sum_{i=1}^n N_i \sim \mathrm{NB}\left(\beta, \sum_{i=1}^n r_i\right).$$

Proof

We shall, again, use the pgf. We have

$$\begin{aligned} G_N(t) = E\left[t^N\right] &\overset{(*)}{=} \prod_{i=1}^n E\left[t^{N_i}\right] = \prod_{i=1}^n G_{N_i}(t) \\ &= \prod_{i=1}^n [1 - \beta(t-1)]^{r_i} = [1 - \beta(t-1)]^{-\sum_{i=1}^n r_i}, \end{aligned}$$

where $(*)$ is by independence of the rvs, and the last equality is thanks to β being fixed for all the rvs. This completes the proof. □

15.1.2 (a, b, n) Classes

15.1.2.1 $(a, b, 0)$ Class

Definition 45 ($(a, b, 0)$ Class)

The $(a, b, 0)$ class is a set of counting rvs with pmf p_k satisfying the recursive formula

$$\frac{p_k}{p_{k-1}} = a + \frac{b}{k}, \quad k \in \mathbb{N} \setminus \{0\}.$$

Remark

An $(a, b, 0)$ distribution is determiend by the parameters a and b.

Note

Observe that

$$\begin{aligned}
\frac{p_1}{p_0} = a + \frac{b}{1} &\iff p_1 = p_0 \left(a + \frac{b}{1} \right) \\
\frac{p_2}{p_2} = a + \frac{b}{2} &\iff p_2 = p_1 \left(a + \frac{b}{2} \right) = p_0 \left(a + \frac{b}{1} \right) \left(a + \frac{b}{2} \right) \\
&\vdots \\
\frac{p_k}{p_{k-1}} = a + \frac{b}{k} &\iff p_k = p_0 \prod_{i=1}^{k} \left(a + \frac{b}{i} \right).
\end{aligned}$$

Thus we see that each of the p_k is completely determined by p_0. In other words, for the distributions of this class, if we can find p_0, then we can get p_k, even if we do not know the actual parameters of the distribution.

In fact, we can solve for p_0, if we already know what a and b are: we need to solve for p_0 in $\sum_{k=0}^{\infty} p_k = 1$. In particular, we need to solve for

$$p_0 \sum_{k=0}^{\infty} \prod_{i=1}^{k} \left(a + \frac{b}{i} \right) = 1.$$

Members of the $(a, b, 0)$ class It can be shown[1] that the Poisson, Binomial, and Negative Binomial distributions are **the only** distributions that belong to this class. We have that

[1] Perhaps this can be shown using `https://www.actuaries.org/ASTIN/Colloquia/Helsinki/Papers/S7_13_Fackler.pdf`.

Table 15.1: The $(a, b, 0)$ distributions

Distribution	a	b	p_0
$\mathrm{Poi}(\lambda)$	0	λ	$e^{-\lambda}$
$\mathrm{Bin}(q, m)$	$-\frac{q}{1-q}$	$(m+1)\frac{q}{1-q}$	$(1-q)^m$
$\mathrm{NB}(\beta, r)$	$\frac{\beta}{1+\beta}$	$(r-1)\frac{\beta}{1+\beta}$	$(1+\beta)^{-r}$

We shall prove for the case of $\mathrm{Poi}(\lambda)$.

Exercise 15.1.2

Find a, b and p_0 for $\mathrm{Bin}(q, m)$ and $\mathrm{NB}(\beta, r)$.

Proof

By the pmf of $\mathrm{Poi}(\lambda)$, it is clear that

$$p_0 = \frac{e^{-\lambda}\lambda^0}{0!} = e^{-\lambda}.$$

Now, since

$$p_1 = \lambda e^{-\lambda} \quad p_2 = \frac{1}{2}\lambda^2 e^{-\lambda}$$

we have the following system of equations:

$$\begin{aligned} \lambda = \frac{p_1}{p_0} = a + b \\ \frac{1}{2}\lambda = \frac{p_2}{p_1} = a + \frac{b}{2} \end{aligned}$$

Thus $b = \lambda$ and $a = 0$. $\square$

Example 15.1.1

Assume that the number of claims in a portfolio N follows $(-0.25, 2.75, 0)$ distribution. Calculate the probability that there is at least one claim.

Solution

Using Table 15.1, we know that $N \sim \mathrm{Bin}(q, m)$, where

$$-\frac{q}{1-q} = -0.25 \text{ and } (m+1)\frac{q}{1-q} = 2.75,$$

which gives $q = 0.2$ and $m = 10$. Thus the desired probability is

$$P(N \geq 1) = 1 - P(N = 0) = 1 - (1 - 0.8)^{10} = 0.8926.$$

16 Lecture 16 Nov 6th

16.1 Frequency Distribution — Basic Frequency Distributioned (Continued 3)

16.1.1 (a, b, n) Classes (Continued)

16.1.1.1 $(a, b, 1)$ Class

Motivation There are times when the $(a, b, 0)$ class of distributions fail to give a complete characterization of certain insurance data with regards to the claim arrival process. This is especially apparent when we notice that we do not have as much freedom in fixing $P(N = 0)$. This provides us with the motivation to define the $(a, b, 1)$ class.

Definition 46 ($(a, b, 1)$ Class)

The $(a, b, 1)$ class is defined as a set of counting rvs with pmf p_k satisfying the recursion

$$\frac{p_k}{p_{k-1}} = a + \frac{b}{k}, \quad k = 2, 3, \ldots.$$

Remark

*Notice that the formula is almost exactly the same as compared to the definition of the $(a, b, 0)$ class, **except** now we have that k starts from 2 instead of 1. This means that this recursive definition will no longer have any control over p_0, which is what we want.*

There are two distributions from the $(a, b, 1)$ class that we shall focus on, namely

- the zero-truncated distribution; and

- the zero-modified distribution.

Zero-Truncated Distribution In this case, we set $p_0 = 0$, i.e. we always expect a claim.

Let p_k, where $k = 0, 1, 2, \ldots$, be the pmf of an $(a, b, 0)$ distribution, of which we label its rv as N. Then, let p_k^T, for $k = 0, 1, 2, \ldots$, be the pmf of the **zero-truncated distribution**, whose rv is denoted by N^T, with $p_0^T = 0$.

We can obtain values for each of the p_k^T's, from $k = 1, 2, \ldots$, from the p_k's: [1]notice that for $k = 2, 3, 4, \ldots$, we have

$$\frac{p_k^T}{p_{k-1}^T} = a + \frac{b}{k} = \frac{p_k}{p_{k-1}} \implies \frac{p_k^T}{p_k} = \frac{p_{k-1}^T}{p_{k-1}}.$$

[1] Perhaps this was implied but it certainly was not explicitly stated: the a, b in the $(a, b, 0)$ can be treated exactly as the a, b from the $(a, b, 0)$ class.

Observe that

$$\begin{aligned}
\text{When } k = 2, \quad & \frac{p_2^T}{p_2} = \frac{p_1^T}{p_1} \\
\text{When } k = 3, \quad & \frac{p_3^T}{p_3} = \frac{p_2^T}{p_2} \\
\text{When } k = 4, \quad & \frac{p_4^T}{p_4} = \frac{p_3^T}{p_3} \\
& \vdots
\end{aligned}$$

Therefore, we have that

$$\frac{p_1^T}{p_1} = \frac{p_2^T}{p_2} = \frac{p_3^T}{p_3} = \ldots =: \beta^T,$$

where we give this value a variable. Consequently, we have

$$p_k^T = \beta^T p_k.$$

Of course, we'd like to know what β^T is. Since $p_0^T = 0$, we have that

$$1 = \sum_{k=1}^{\infty} p_k^T = \beta^T \sum_{k=1}^{\infty} p_k = \beta^T (1 - p_0)$$

and so

$$\beta^T = \frac{1}{1 - p_0}.$$

To store this information, we look to the pgf of N^T: we have

$$G_{N^T}(t) = \sum_{k=0}^{\infty} t^k p_k^T = \sum_{k=0}^{\infty} t^k \beta^T p_k = \frac{1}{1-p_0} \sum_{k=1}^{\infty} t^k p_k = \frac{G_N(t) - p_0}{1-p_0}.$$

Example 16.1.1

Let $N \sim \mathrm{NB}(\beta, r)$. Its zero-truncated version has pmf of the form

$$\begin{aligned} p_K^T = \frac{p_k}{1-p_0} &= \frac{\binom{k+r-1}{k}\left(\frac{1}{1+\beta}\right)^r \left(\frac{\beta}{1+\beta}\right)^k}{1-\left(\frac{1}{1+\beta}\right)^r} \\ &= \frac{\Gamma(k+r)}{k!\Gamma(r)} \frac{\left(\frac{\beta}{+}\right)^k}{(1+\beta)^r - 1} \end{aligned}$$

Zero-Modified Distribution If we choose $P(N=0)$ to be some value that is not any of the p_0's of the distributions from the $(a, b, 0)$ class, then this distribution is called a **zero-modified distribution**.

Let p_k, for $k = 0, 1, 2, \ldots$, be the pmf of an $(a, b, 0)$ distribution, labelled N. Let p_k^M, for $k = 0, 1, 2, \ldots$, be the pmf of the zero-modified distribution, of which we denote by N^M, with p_0^M chosen as described.

We can, again, express p_k^M in terms of p_k: for $k = 1, 2, 3, \ldots$ [2],

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

and so we have, again,

$$p_k^M = \beta^M p_k.$$

So we can solve for β^M:

$$\begin{aligned} 1 = \sum_{k=0}^{\infty} p_k^M &= p_0^M + \sum_{k=1}^{\infty} \beta^M p_k \\ \implies 1 - p_0^M &= \beta^M \sum_{k=1}^{\infty} p_k \\ \implies \beta^M &= \frac{1-p_0^M}{1-p_0}. \end{aligned}$$

[2] Note that in this case, the probabilities after p_0^M is reliant on p_0; of course, since the sum of the probabilities must be 1, within the axioms of probability.

The pgf of a zero-modified distribution is

$$\begin{aligned} G_{N^M}(t) &= \sum_{k=0}^{\infty} t^k p_k^M = p_0^M + \sum_{k=1}^{\infty} t^k \beta^M p_k = p_0^M + \beta^M (G_N(t) - p_0) \\ &= p_0^M + \frac{1 - p_0^M}{1 - p_0} (G_N(t) - p_0) \\ &= \frac{p_0^M - p_0^M p_0}{1 - p_0} + \frac{1 - p_0^M}{1 - p_0} G_N(t) - \frac{p_0 - p_0^M p_0}{1 - p_0} \\ &= \frac{p_0^M - p_0}{1 - p_0} + \frac{1 - p_0^M}{1 - p_0} G_N(t) \end{aligned}$$

Remark

As a consequence of the form of the pgf, if $p_0^M > p_0$, we may interpret N_M as

$$N^M = \begin{cases} 0 & \text{w/ prob } \frac{p_0^M - p_0}{1 - p_0} \\ N & \text{w/ prob } \frac{1 - p_0^M}{1 - p_0} \end{cases},$$

which is a mixture of a degenerated distribution at 0, and the original $(a, b, 0)$ distribution N.

Consequently, we can also solve for $G_{N^M}(t)$ using this notion: let Θ be the indicator-function-like distribution such that

$$P(\Theta = \theta) = \begin{cases} \frac{p_0^M - p_0}{1 - p_0} & \theta = 0 \\ \frac{1 - p_0^M}{1 - p_0} G_N(t) & \theta = 1 \end{cases}.$$

Then

$$\begin{aligned} G_{N^M}(t) &= E\left[t^{N^M}\right] = E\left[E\left[t^{N^M} \mid \Theta\right]\right] \\ &= E\left[t^{N^M} \mid \Theta = 0\right] P(\Theta = 0) + E\left[t^{N^M} \mid \Theta = 1\right] P(\Theta = 1) \\ &= P(\Theta = 0) E\left[t^0 \mid \Theta = 0\right] + P(\Theta = 1) E\left[t^N \mid \Theta = 1\right] \\ &= \frac{p_0^M - p_0}{1 - p_0}(1) + \frac{1 - p_0^M}{1 - p_0} G_N(t), \end{aligned}$$

as what we had.

Proposition 23 (Moments of an $(a, b, 1)$ Distribution)

The moments of an $(a, b, 1)$ distribution can be compited from the original $(a, b, 0)$ distribution as

$$E\left[\left(N^M\right)^k\right] = \frac{1 - p_0^M}{1 - p_0} E\left[N^k\right], \quad k = 1, 2, 3, \ldots$$

where N^M is the rv modified from N from the $(a, b, 0)$ class.

Proof

The derivation is straightforward:

$$\begin{aligned}
E\left[\left(N^M\right)^k\right] &= \sum_{m=0}^{\infty} m^k p_K^M = \sum_{m=1}^{\infty} \frac{1-p_0^M}{1-p_0} m^k p_k \\
&= \frac{1-p_0^M}{1-p_0} \sum_{m=1}^{\infty} m^k p_k = \frac{1-p_0^M}{1-p_0} \sum_{m=0}^{\infty} m^k p_k \\
&= \frac{1-p_0^M}{1-p_0} E[N].
\end{aligned}$$

□

Example 16.1.2

Let $N \sim \text{Poi}(2)$. Find the pmf of a zero-modified version of the Poisson distribution with $p_0^M = 0.3$.

Solution

We are given that the pmf of N is

$$p_k = \frac{e^{-2} 2^k}{k!}, \quad k = 0, 1, 2, \ldots.$$

Thus

$$\begin{aligned}
p_k^M &= \frac{1-p_0^M}{1-p_0} p_k = \frac{1-0.3}{1-e^{-2}} \cdot \frac{e^{-2} 2^k}{k!} \\
&= \frac{0.7\left(2^k e^{-2}\right)}{k!(1-e^{-2})}, \quad k = 1, 2, 3, \ldots.
\end{aligned}$$

Note

It should be noted that a zero-truncated distribution is a special case of the zero-modified distribution.

In addition to the zero-modified $(a, b, 0)$ distributions, there are other members in the $(a, b, 1)$ distributions. One such example is known as the **logarithmic distribution**, which has the pmf

$$p_k = \frac{\beta^k}{k(1+\beta)^k \ln(1+\beta)}, \quad k = 1, 2, 3, \ldots,$$

with parameter $\beta > 0$. Notice that for $k = 2, 3, 4, \ldots$,

$$\begin{aligned}
\frac{p_k}{p_{k-1}} &= \frac{\beta^k}{k(1+\beta)^k \ln(1+\beta)} \cdot \frac{(k-1)(1+\beta)^{k-1}\ln(1+\beta)}{\beta^{k-1}} \\
&= \frac{\beta}{1+\beta} \cdot \frac{k-1}{k} = \frac{\beta}{1+\beta} + \frac{-\frac{\beta}{1+\beta}}{k},
\end{aligned}$$

where we observe that

$$a = \frac{\beta}{1+\beta} \text{ and } b = -\frac{\beta}{1+\beta}.$$

17 Lecture 17 Nov 13th

17.1 Frequency Distributions — Creating New Frequency Distributions and Effect on Frequency

17.1.1 Mixed Frequency Distributions

This is a concept that we have seen before.

Suppose that $N \mid \Theta = \theta$ has conditional pmf $P(N = n \mid \Theta = \theta)$ and the mixing rv Θ is either

- **discrete** with pmf $p_\Theta(\theta_i)$ for $i = 1, \ldots, m$; or
- **continuous** with pdf $f_\Theta(\theta)$.

Both the distributions of $N \mid \Theta = \theta$ and Θ are usually given. The unconditional pmf of N is thus

$$P(N = n) = \begin{cases} \sum_{i=1}^{m} P(N = n \mid \Theta = \theta_i) p_\Theta(\theta_i) & \Theta \text{ is discrete} \\ \int_\Theta P(N = n \mid \Theta = \theta) f_\Theta(\theta)\, d\theta & \Theta \text{ is continuous} \end{cases}.$$

Remark
Due to the context of which we work in, we shall always, perversely so, assume that N is a counting rv that takes on non-negative integers.

Note
Recall Proposition 9, which gives us two concepts that are useful to us in this section: we have

$$E\left[E\left[X \mid \Theta\right]\right] = E[X]$$
$$\mathrm{Var}(X) = \mathrm{Var}(E\left[X \mid \Theta\right]) + E\left[\mathrm{Var}(X \mid \Theta)\right].$$

17.1.1.1 Mixed Poisson Distribution

Definition 47 (Mixed Poisson Distribution)

*If $N \mid \Lambda = \lambda$ follows $\mathrm{Poi}(\lambda)$ for some rv Λ, we say N follows a **mixed Poisson Distribution**.*

Example 17.1.1

Recall Proposition 21. We have that given $N \mid \Lambda = \lambda \sim \mathrm{Poi}(\lambda)$ and $\Lambda \sim \mathrm{Gam}(\alpha, \theta)$, we have that

$$N \sim \mathrm{NB}(\theta, \alpha).$$

Proposition 24 (Mixed Poisson Distribution has a Variance Greater than its Mean)

FOr a mixed Poisson rv N, we have $\mathrm{Var}(N) > E[N]$.

Proof

When we say that N is a mixed Poisson rv, it actually means that the rv that we have is $N \mid \Lambda = \lambda$, not N alone. Now by Proposition 9, since $\Lambda = \lambda$ is the mean of the mixed Poisson rv, we have

$$\begin{aligned}\mathrm{Var}(N) &= \mathrm{Var}(E[N \mid \Lambda]) + E[\mathrm{Var}(N \mid \Lambda)] \\ &= \mathrm{Var}(\Lambda) + E[\Lambda] = \mathrm{Var}(\Lambda) + E[N].\end{aligned}$$

Since $\mathrm{Var}(\Lambda) > 0$, the result follows. □

Example 17.1.2

Suppose that $N \mid \Lambda = \lambda$ follows $\mathrm{Poi}(\lambda)$ and the pdf of Λ is given by

$$f_\Lambda(\lambda) = \frac{\alpha^2}{1+\alpha}(1+\lambda)e^{-\alpha\lambda}, \quad \lambda > 0,$$

where $\alpha > 0$. Show that N is the mixture of two negative binomial distributions.

Solution

We shall show the claim by the pgf of N. We have

$$\begin{aligned} G_N(t) &= E\left[t^N\right] = E\left[E\left[t^N \mid \Lambda\right]\right] = E\left[e^{\Lambda(t-1)}\right] \\ &= \int_0^\infty \frac{\alpha^2}{1+\alpha}(1+\lambda)e^{-\lambda(\alpha+1-t)}\,d\lambda \end{aligned}$$

From here, we can only proceed iff $1+\alpha-t > 0$, i.e. $t < 1+\alpha$; otherwise the integral diverges. Now, using integration by parts

$$\begin{aligned} G_N(t) = \frac{\alpha^2}{1+\alpha}&\left[-\frac{1}{\alpha+1-t}e^{-\lambda(\alpha+1-t)}\Big|_0^\infty + \right. \\ &\left.\left[-\frac{1}{\alpha+1-t}\lambda e^{-\lambda(\alpha+1-t)}\Big|_0^\infty + \frac{1}{(\alpha+1-t)^2}e^{-\lambda(\alpha+1-t)}\Big|_0^\infty\right]\right] \\ = \frac{\alpha^2}{1+\alpha}&\left[\frac{1}{\alpha+1-t}+\frac{1}{(\alpha+1-t)^2}\right] \end{aligned}$$

Now to arrive at a mixture of two negative binomial distributions, we need to know the form of the pgf for a negative binomial distribution. Note that the pgf of $\mathrm{NB}(\beta, r)$ is

$$G(t) = [1-\beta(t-1)]^{-r}.$$

Notice that

$$G_N(t) = \frac{\alpha^2}{1+\alpha}\left[[\alpha-(t-1)]^{-1}+[\alpha-(t-1)]^{-2}\right].$$

If we expand the appropriate power of α into the reciprocals, we can get our desired form:

$$G_N(t) = \frac{\alpha}{1+\alpha}\left[1-\frac{1}{\alpha}(t-1)\right]^{-1} + \frac{1}{1+\alpha}\left[1-\frac{1}{\alpha}(t-1)\right]^{-2}.$$

It is clear that

$$\frac{\alpha}{1+\alpha}+\frac{1}{1+\alpha}=1,$$

and so we have obtained our desired result; that is N is a weighted mixture of two negative binomial distributions. In particular,

$$N = \begin{cases} X_1 \sim \mathrm{NB}\left(\frac{1}{\alpha}, 1\right) & \text{w/ prob } \frac{\alpha}{1+\alpha} \\ X_2 \sim \mathrm{NB}\left(\frac{1}{\alpha}, 2\right) & \text{w/ prob } \frac{1}{1+\alpha} \end{cases}.$$

We can make a similar derivation of mixed distributions for Binomial and Negative Binomial.

17.1.2 Compound Frequency Distributions

Definition 48 (Compound Frequency Distribution)

For two counting rvs N and M, let

$$S = \sum_{i=1}^{N} M_i = \begin{cases} M_1 + \ldots + M_N & N \geq 1 \\ 0 & N = 0 \end{cases},$$

*where $\{M_i\}_{i=1}^{\infty}$ is a sequence of iid rv's distributed as M, and are independent of N. We call S a **compound rv**, N the **primal distribution**, and M the **secondary distribution**.*

Remark

- *Compounding two counting rvs is also an approach to create new frequency distributions.*
- *S is called **compound Poisson, Binomial, or Negative Binomial** if N is a Poisson, a Binomial, or a Negative Binomial rv, respectively.*

Example 17.1.3 (Interpretation)

In an insurance context, compound rvs arise rather naturally. E.g. in the auto insurance context, we could have

- N represents the number of accidents;
- M_i represents the number of claims generated by the i^{th} accident;
- And so in this case S stands for the total number of claims for a portfolio of auto insurance policies over a given time period.

Proposition 25 (Mean and Variance of the Compound RV)

For a compound rv $S = \sum_{i=1}^{N} M_i$, where $M_i \sim M$, we have

$$E[S] = E[N]E[M]$$
$$\operatorname{Var}(S) = \operatorname{Var}(N)E[M]^2 + E[N]\operatorname{Var}(M).$$

Proof

Note that the definition of S relies on N first, since

$$S = \begin{cases} M_1 + \ldots + M_N & N \geq 1 \\ 0 & N = 0 \end{cases},$$

so we shall go down of the route of conditioning S by N. Observe that

$$E[S \mid N] = E\left[\sum_{i=1}^{N} M_i \mid N\right] \overset{(*)}{=} \sum_{i=1}^{N} E[M_i \mid N] \overset{(**)}{=} \sum_{i=1}^{N} E[M] = NE[M]$$

where $(*)$ is by the linearity of the expectation, and $(**)$ is by $M_i \perp N$ for each i and that $M_i \sim M$. The variance of S conditioned on N is

$$\operatorname{Var}(S \mid N) = \operatorname{Var}\left(\sum_{i=1}^{N} M_i \mid N\right) = \sum_{i=1}^{N} \operatorname{Var}(M) = N \operatorname{Var}(M)$$

mostly for the same reason as for the expectation, but the 2nd equality involves independence of the M_i's (otherwise, we would be left with a bunch of covariances).

Then, using Proposition 9, we have

$$E[S] = E[E[S \mid N]] = E[NE[M]] = E[N]E[M]$$

and

$$\begin{aligned} \operatorname{Var}(S) &= \operatorname{Var}(E[S \mid N]) + E[\operatorname{Var}(S \mid N)] \\ &= \operatorname{Var}(NE[M]) + E[N \operatorname{Var}(M)] \\ &= \operatorname{Var}(N)E[M]^2 + E[N] \operatorname{Var}(M) \end{aligned}$$

as required. □

Note (Notation)

Hereafter, we shall use the following notations: notice that each M, N, and S are counting rvs, i.e. they are discrete and are non-negative integers, so let

- *p_k represent the pmf of N, the primal distribution;*
- *f_k represent the pmf of M, the secondary distribution; and*
- *g_k represent the pmf of S, the compound rv.*

In the next lecture, we shall look into how to compuete g_k. Namely, we have the following three methods:

- pgf method;
- pmf method; and
- Panjer's recursion.

A Additional Material

A.1 Individual Risk Model: An Alternate View

This appendix serves to explain why our note of $Z_i = I_i X_i$ is wrong with as mush rigour as we can go for now. There may be hand-wavy parts, but those will be indicated.

We mentioned, as shown by Klugman, Panjer and Willmot (2012)[1], that for the Individual Risk Model, the aggregate claim is modeled by

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

where Z_i is a random variable for the potential loss of the i^{th} insurance policy, while n is fixed. It is claimed that we can also express each Z_i as

$$Z_i = I_i X_i$$

where I_i is an indicator function given by

$$I_i(x) = \begin{cases} 1 & \text{if a claim occurs} \\ 0 & \text{if there are no claims} \end{cases},$$

while X_i is the size of the claim(s) for the i^{th} policy provided that there is a claim.

[1] Klugman, S. A., Panjer, H. H., and Willmot, G. E. (2012). *Loss Models: From Data to Decisions*. John Wiley & Sons, Inc., 4th edition

One problem that arises is: are X_i and I_i independent? They should be if we wish to define Z_i in such a way. In fact, according to Klugman et al. in page 177,

> *Let $X_j = I_j B_j$, where $I_1, \ldots, I_n, B_1, \ldots, B_n$ are independent.*

where X_j is our Z_i, I_j is our I_i, and B_j is our X_i.

§ *Z_i is not well-defined* Let us be explicit about the definitions of I_i and X_i; we have

$$I_i = \mathbb{1}_{\{Z_i > 0\}}$$
$$X_i = Z_i \mid Z_i > 0$$

However, we observe that such a defintion of X_i is undefined on $Z_i = 0$. So the equation

$$Z_i = I_i X_i$$

is note well-defined.

§ *Independence of I_i and X_i* We cannot actually tell if I_i and X_i are independent from each other, as it is equivalent to comparing apples with oranges[2]. Recall from our earlier courses, in particular STAT330, of the following notion:

[2] In fact, I think this analogy fits our case perfectly so.

Definition (Probability Space)

Let Ω be a sample space, and $\mathcal{F}$ a σ-algebra defined on Ω[3]. A ***probability space*** *is the measurable space $(\Omega, \mathcal{F})$ with a probability measure, $f : \mathcal{F} \to [0, 1]$, defined on the space. We denote a probability space as $(\Omega, \mathcal{F}, f)$.*

[3] Note that $(\Omega, \mathcal{F})$ is called a **measurable space**.

As mentioned in an earlier §, X_i is not defined on $Z_i = 0$, while I_i is defined on $Z_i = 0$ [4]. So the sample space for X_i and I_i are not the same, and so their probability measures are not the same as well. Therefore, **it is meaningless to ask if X_i and I_i are independent**.

[4] This statement is hand-wavy.

Our best attempt at fixing this is probably the following: let

$$Z_i = \sum_{i=1}^{I_i} X_i,$$

which we can then have X_i to be independent from I_i. However, interestingly so, this is a similar approach to a Collective Risk Model.

A.2 *Coherent Risk Measure*

An excerpt from Klugman et al. (2012) [5]:

[5] Klugman, S. A., Panjer, H. H., and Willmot, G. E. (2012). *Loss Models: From Data to Decisions*. John Wiley & Sons, Inc., 4th edition

> *The study of risk measures and their properties has been carried out by authors such as Wang. Specific desirable properties of risk measures were proposed as axioms in connection with risk pricing by Wang, Young, and Panjer and more generally in risk measurement by Artzer et al. The Artzner paper introduced the concept of **coherence** and is considered to be the groundbreaking paper in risk measurement.*

Often, we use the function $\rho(X)$ to denote risk measures. One may think of $\rho(X)$ as **the amount of assets required to protect against adverse outcomes of the risk X**.

Definition 49 (Coherent Risk Measure)

*A **coherent risk measure** is a risk measure $\rho(X)$ that has the following four properties for any two loss rvs X and Y:*

1. *(**Subadditivity**)* $\rho(X+Y) \leq \rho(X) + \rho(Y)$.
2. *(**Monotonicity**) If $X \leq Y$ for all possible outcomes, then $\rho(X) \leq \rho(Y)$.*
3. *(**Positive homogeneity**)* $\forall c \in \mathbb{R}_{>0}$, $\rho(cX) = c\rho(X)$.
4. *(**Translation invariance**)* $\forall c \in \mathbb{R}_{>0}$, $\rho(X+c) = \rho(X) + c$

Interpretation of the conditions

- **Subadditivity**
  - the risk measure (and in return, the capital required to cover for it) for two risks combined will not be greater than for the risks to be treated separately;
  - reflects the fact that there shuld be some diversification benefit from combining risks;
  - this requirement is disputed: e.g. the merger of several small companies into a larger one exposes each of the small companies to the **reputational risks** of the others.
- **Monotonicity**

- if one risk always has greater losses than the other under all circumstances[6], then the risk measure of the greater risk should always be greater than the other.

[6] Probabilistically, this means $P(X > Y) = 0$

- **Positive homogeneity**
  - the risk measure is independent of the currency used to measure it;
  - doubling the exposure to a particular risk requires double the capital, which is sensible as doubling provides no diversification.
- **Translation invariance**
  - there is no additional risk for an additional risk which has no additional uncertainty.

Bibliography

Dhaene, J. and Vyncke, D. (2010). The individual risk model. https://www.researchgate.net/publication/228232062_The_Individual_Risk_Model.

Klugman, S. A., Panjer, H. H., and Willmot, G. E. (2012). *Loss Models: From Data to Decisions*. John Wiley & Sons, Inc., 4th edition.

List of Symbols and Abbreviations

crv	continuous random variable
DFR	Decreasing Failure Rate
DMRL	Decreasing Mean Residual Lifetime
drv	discrete random variable
$e_X(d)$	Mean Excess Loss / Mean Residual Lifetime
$G_N(t)$	probability generating function of random variable N
h_X	hazard rate of random variable X
IFR	Increasing Failure Rate
IMRL	Increasing Mean Residual Lifetime
LER	Loss Elimination Ratio
mgf	moment generating function
pf	probability function
pdf	probability density function
pmf	probability mass function
pgf	probability generating function
rv	random variable
$\bar{F}_X$	survival function of random variable X
T_L	Amount Paid per Loss
T_P	Amount Paid Per Payment
TVaR	Tail-Value-at-Risk
VaR	Value-at-Risk

Index