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

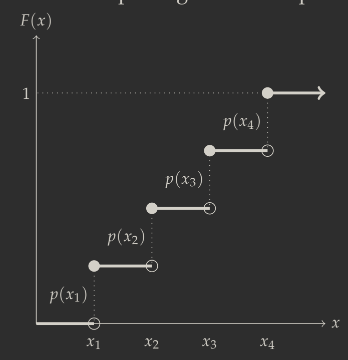
Let X take values from $\{x_1, x_2, x_3, x_4\}$, and

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

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*

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

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

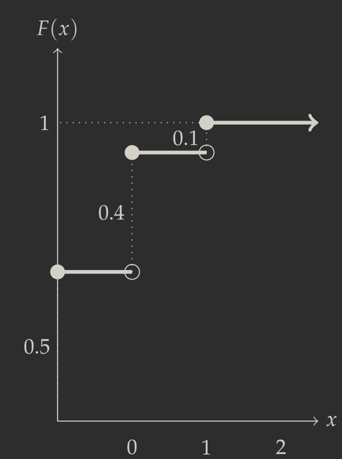
Example 2.1.2 (Lecture Slides: Example 1)

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*

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

where $\bar{F}(x) = 1 - F(x)$ *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*

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

- *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)

Suppose $X \sim \mathrm{Wei}(\theta, \tau)$ *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:

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

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.

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

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 $\{x_1, x_2, ...\}$ 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

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

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

[1] 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.*

Example 3.2.2 (Lecture Notes: Example 1)

Graphical method to interpret this notion will be included.

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***

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

- ***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 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]

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

[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.

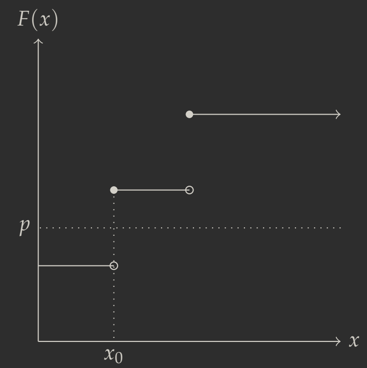


Figure 4.1: Discrete cdf

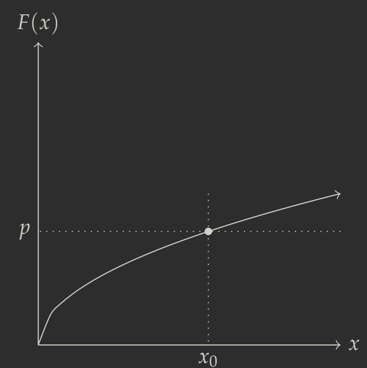


Figure 4.2: Continuous cdf

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 Tail-Value-at-Risk

To compensate for the weakness of VaR at giving us the size of the 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,

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

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

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

□

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,

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

□

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 $\{X_1, ..., X_n\}$ 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:

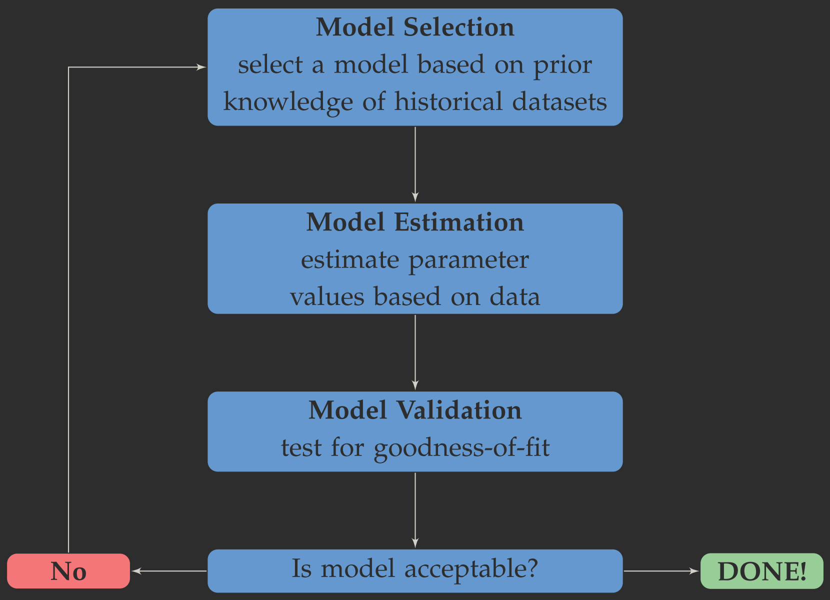


Figure 6.1: Process of a Parametric Approach

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 $X \sim \text{Exp}(\theta)$ 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 $Y \sim \text{Exp}(c\theta)$ 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 $Y = cX \sim \text{Exp}(c\theta)$. Therefore, the parameter θ is a scale parameter.

Example 6.1.3

Determine whether the lognormal distribution $X \sim \text{LogN}(\mu, \sigma^2)$, i.e. $\ln(X) \sim \text{N}(\mu, \sigma^2)$, 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

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

□

Example 6.1.4

Let $X \sim \mathrm{Exp}(\theta)$ *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* $Y \sim \mathrm{Wei}(\alpha, \tau)$*.*

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 $\{\theta_1, \theta_2, ..., \theta_n\}$ with

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

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 $Y_i \sim \text{Exp}(i)$ 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

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

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

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

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

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

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

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

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.

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 very 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 30 (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.

- **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.

[6] Probabilistically, this means $P(X > Y) = 0$

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

rv	random variable
drv	discrete random variable
crv	continuous random variable
pf	probability function
pmf	probability mass function
pdf	probability density function
mgf	moment generating function
pgf	probability generating function

Index