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.**

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

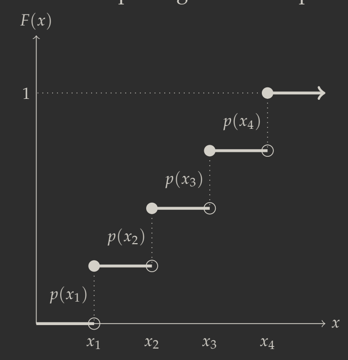
Let X take values from $\{x_1, x_2, x_3, x_4\}$, and

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

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.

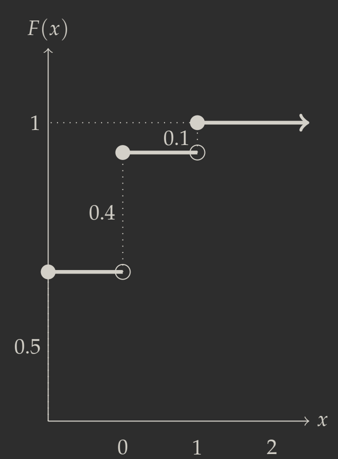
Example 2.1.2 (Lecture Slides: Example 1)

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

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

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

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

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.

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

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 $Y = \begin{cases} Y_1 & U = 1 \\ Y_2 & U = 2 \end{cases}$ [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

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

Therefore

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

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

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

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*

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

A Individual Risk Model: An Alternative 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.

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.

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

§ *Z_i is not well-defined* Let us be explicit about the definitions of I_i and X_i; we have

$$
\begin{aligned}
I_i &= \mathbb{1}_{\{Z_i > 0\}} \\
X_i &= Z_i \mid Z_i > 0
\end{aligned}
$$

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.

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