HKU STAT3906 Study Notes

Chiu Ka Long (Leo)*

Last Updated: 2024-08-11

[Note: Related SOA Exam: FAM (short-term)]

Contents

*email: leockl@connect.hku.hk; personal website: https://leochiukl.github.io

1 Basic Distributional Quantities

1.0.1 Here we will introduce some distributional quantities related to a random variable. (Some of them are discussed in STAT2901.)

1.1 Raw and Central Moments

1.1.1 The **kth raw moment** (or kth moment) of a random variable X, denoted by μ_k', is $\mathbb{E}\left[X^k\right]$.

[Note: The 1st raw moment of a random variable X is the *mean* of X, and is denoted by μ.]

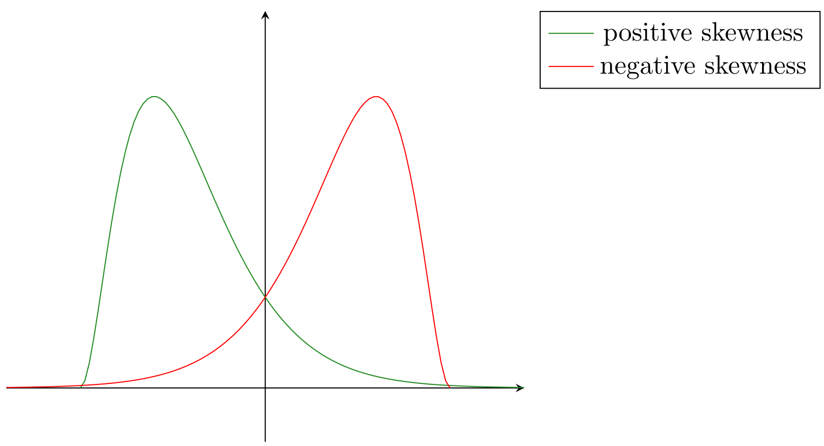
1.1.2 The **kth central moment** of X, denoted by μ_k, is $\mathbb{E}\left[(X-\mu)^k\right]$.

1.1.3 Some quantities of interest related to central moment are as follows.

Quantity	Definition	Notation
variance	μ_2	σ^2
standard deviation	$\sqrt{\mu_2}$	σ
coefficient of variation	σ/μ	$\diagup$
skewness	μ_3/σ^3	γ_1
kurtosis	μ_4/σ^4	γ_2

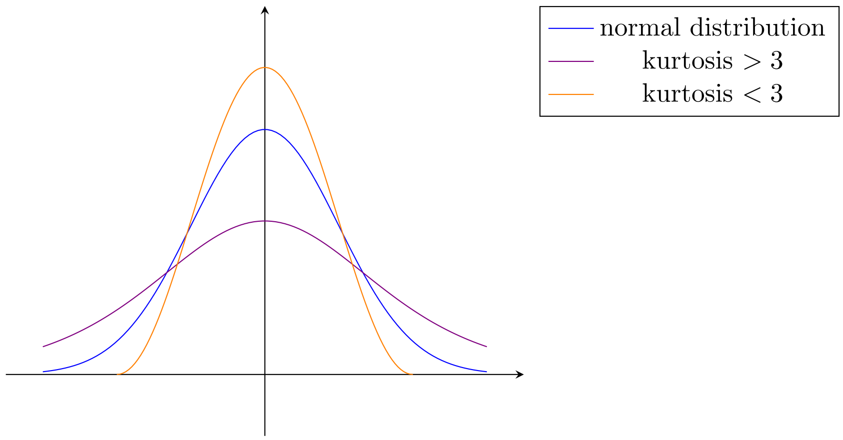
1.1.4 Here we give interpretations of the quantities not covered in STAT2901:

- *Coefficient of variation*: It is "relative" standard deviation (standard deviation per unit mean).
- *Skewness*: It is a measure of *asymmetry*. A symmetric distribution has a skewness of zero. [Warning: This does <u>not</u> mean a distribution with zero skewness is necessarily symmetric!] Positive (negative) skewness indicates that the right (left) tail is *longer* and the mass of distribution is concentrated on the left (right).



  [Intuition: Skewness can be written as $\mathbb{E}\left[(X-\mu)/\sigma^3\right]$. Since the term inside is raised to power 3, long right (left) tail contributes very positively (negatively) to skewness value.]

- *Kurtosis*: It measures "flatness" of the distribution relative to a *normal distribution* (which has a kurtosis of 3). When kurtosis is above (below) 3, it is "flatter" ("less flat") than normal distribution. (Keeping standard deviation constant, more mass of distribution is located away from mean, relative to a normal distribution.)

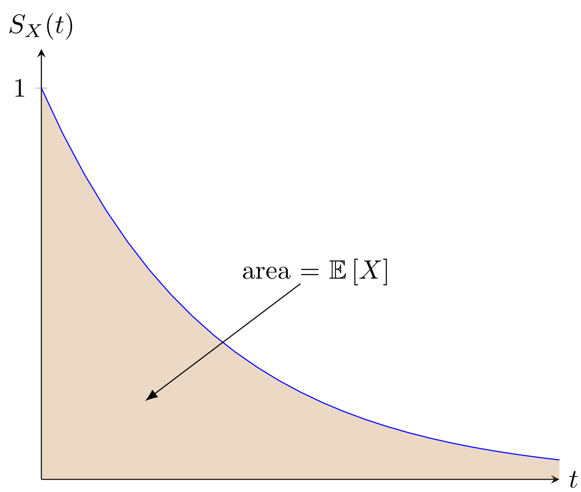


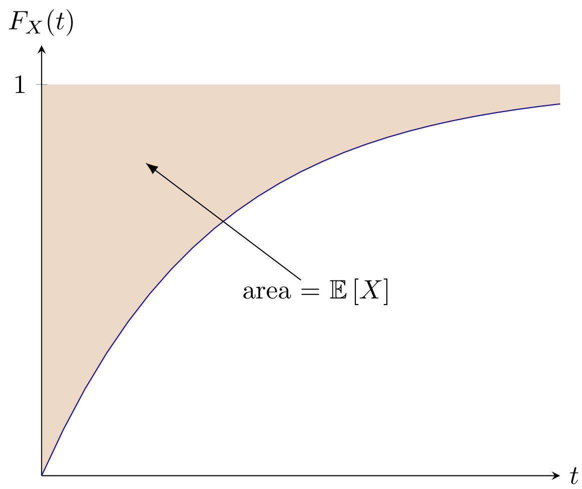
[Intuition 💡: Kurtosis can be written as $\mathbb{E}\left[((X-\mu)/\sigma)^4\right]$. Since the term in the expectation is raised to power 4, mass near mean (with $(X-\mu)/\sigma < 1$) contributes very little to the kurtosis value, while mass away from mean (large $(X-\mu)/\sigma$) contributes a lot to the kurtosis value.]

1.1.5 The following result provides an useful formula for computing mean.

Proposition 1.1.a. Let X be a nonnegative random variable with finite mean (i.e., $\mathbb{E}[X] < \infty$), and let $S_X(x) = \mathbb{P}(X > x)$ be its survival function. Then,

$$\mathbb{E}[X] = \int_0^\infty S_X(t)\,\mathrm{d}t\,.$$





Proof: Since $X \geq 0$, we have

$$X = \int_0^X 1 \,\mathrm{d}t = \int_0^\infty \mathbf{1}_{\{t<X\}} \,\mathrm{d}t$$

[Note: $\mathbf{1}_{\{\cdot\}}$ denotes the indicator function.]

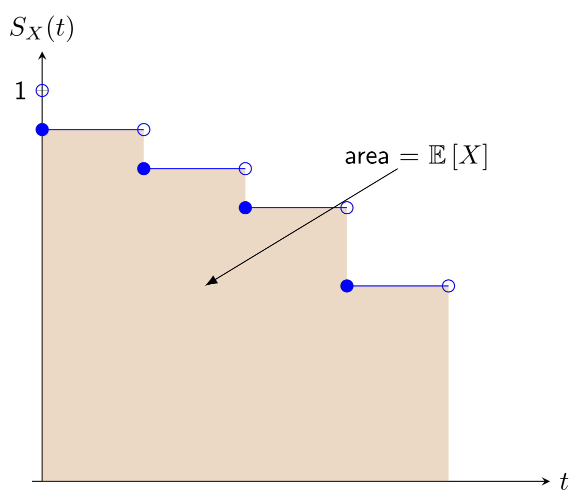
Thus,

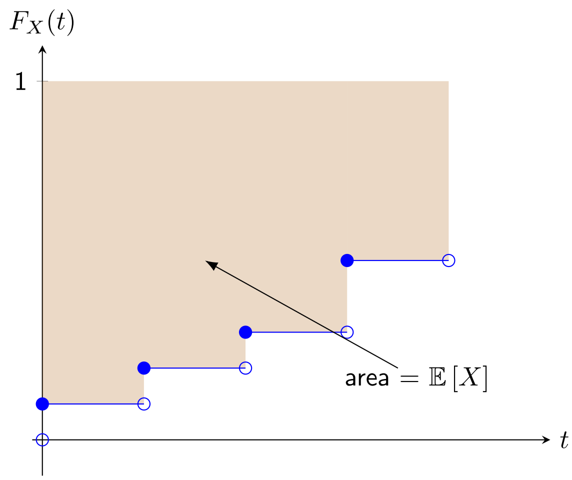
$$\mathbb{E}[X] = \mathbb{E}\left[\int_0^\infty \mathbf{1}_{\{t<X\}} \,\mathrm{d}t\right] = \int_0^\infty \mathbb{E}\left[\mathbf{1}_{\{X>t\}}\right] \mathrm{d}t = \int_0^\infty \mathbb{P}(X>t)\,\mathrm{d}t$$

where the second equality holds by Fubini's theorem. $\square$

Remarks:

- This result holds no matter X is discrete or continuous.
- This result suggests a geometrical interpretation of mean of (nonnegative) X: area under graph of S_X or area between the graph of $y = F_X(t)$ and the line $y = 1$. This still holds when X is discrete!



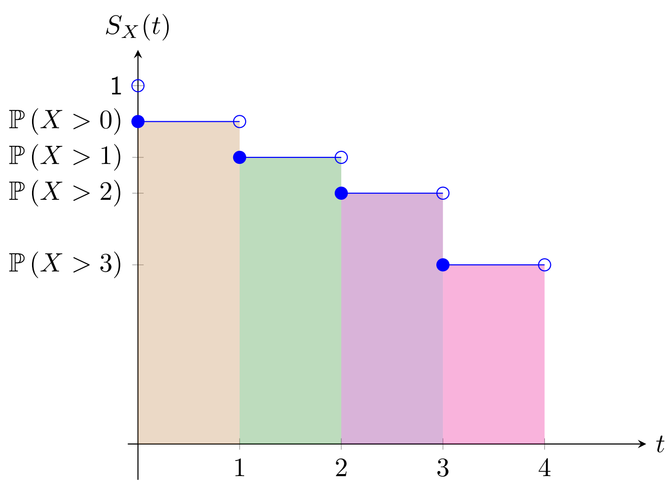


1.1.6 As a corollary, we have the following the result for *discrete* random variable.

Corollary 1.1.b. Let X be a nonnegative discrete random variable. Then,

$$\mathbb{E}[X] = \sum_{n=0}^{\infty} \mathbb{P}(X > n).$$

Proof: $\mathbb{E}[X]$ is the area under S_X:



By summing up the areas of the rectangles, the area under S_X is

$$\mathbb{P}(X > 0) \cdot 1 + \mathbb{P}(X > 1) \cdot 1 + \mathbb{P}(X > 2) \cdot 1 + \mathbb{P}(X > 3) \cdot 1 + \cdots = \sum_{n=0}^{\infty} \mathbb{P}(X > n).$$

□

1.2 Stop Loss Variables

1.2.1 Fix any real number d and consider a loss random variable X (positive value ↔ positive loss). Then, the **stop loss variable** is

$$(X - d)_+ = \begin{cases} X - d & \text{if } X > d; \\ 0 & \text{if } X \le d \end{cases}$$

(where $x_+ = \max\{x, 0\} = x \vee 0$ is the **positive part** of x).

1.2.2 For a practical interpretation of stop loss variable, consider the following. Suppose that the insurer insures a loss X with a **deductible** of d dollars (i.e., the policyholder suffering the loss X is responsible for first d dollars of loss, and insures the remaining portion (if exist)). Then, the stop loss variable represents the payment made by :

- If the loss $X \leq d$, then there is no payment.
- If the loss $X > d$, then the payment amount is $X - d$.

[Note: By having such insurance, the policyholder suffers *at most* d dollars of loss, so the insurance "stops" the loss suffered by (from d dollars onwards) → hence "stop loss".]

1.2.3 If X is continuous with pdf f_X, then

$$\mathbb{E}\left[(X-d)_+\right] = \int_{-\infty}^{\infty} (x-d)_+ f_X(x)\,\mathrm{d}x = \int_{-\infty}^{d} 0\,\mathrm{d}x + \int_{d}^{\infty} (x-d) f_X(x)\,\mathrm{d}x = \boxed{\int_{d}^{\infty} (x-d) f_X(x)\,\mathrm{d}x}.$$

On the other hand, if X is discrete, then

$$\mathbb{E}\left[(X-d)_+\right] = \sum_j (x_j-d)_+ \mathbb{P}\left(X = x_j\right) = \boxed{\sum_{x_j > d} (x_j - d) \cdot p_j}.$$

where $p_j = \mathbb{P}\left(X = x_j\right)$ and the sum is taken over all j where $p_j > 0$ (with $x_j > d$ for the second sum).

1.2.4 We have the following result for the stop loss variable, which is similar to proposition 1.1.a.

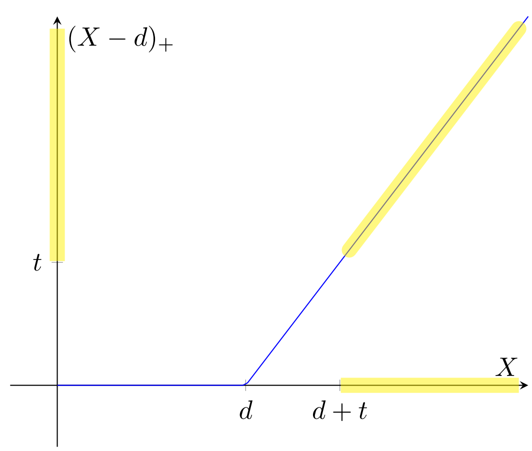
Proposition 1.2.a. Let X be a (nonnegative or not) random variable with finite mean and let $S_X(x)$ be its survival function. Then, for any $d \in \mathbb{R}$,

$$\mathbb{E}\left[(X-d)_+\right] = \int_{d}^{\infty} S_X(x)\,\mathrm{d}x\,.$$

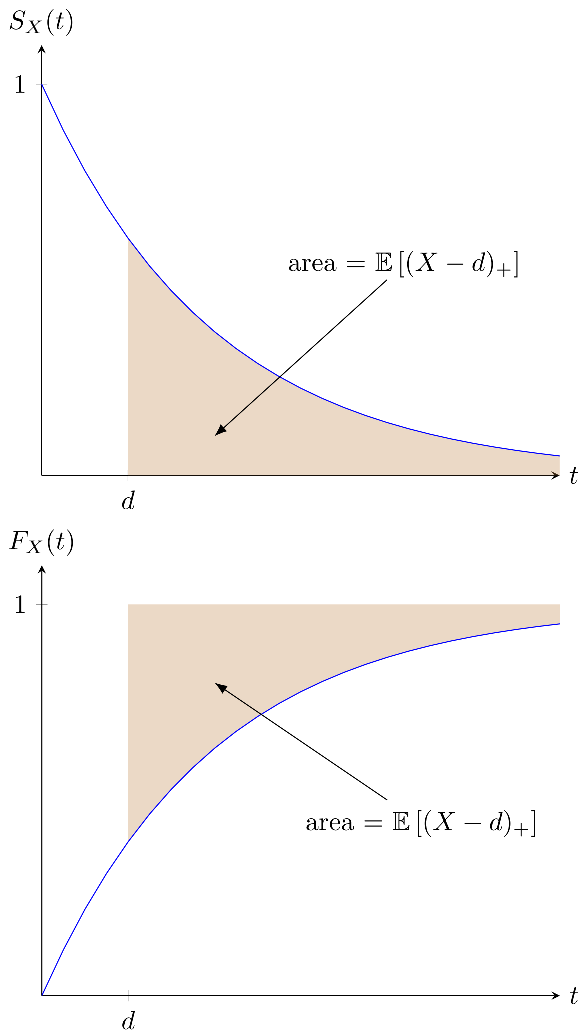
Proof: Since $(X-d)_+ \geq 0$ (with finite mean also), by proposition 1.1.a we have

$$\mathbb{E}\left[(X-d)_+\right] = \int_{0}^{\infty} \mathbb{P}\left((X-d)_+ > t\right)\mathrm{d}t = \int_{0}^{\infty} \mathbb{P}\left(X > d+t\right)\mathrm{d}t \overset{x=t+d}{=} \int_{d}^{\infty} S_X(x)\,\mathrm{d}x$$

where the second equality holds as $(X-d)_+ > t \iff X > d+t$ for any $t \geq 0$.



□



1.3 Excess Loss Variables

1.3.1 Consider again a loss random variable X and fix any $d \in \mathbb{R}$ with $\mathbb{P}(X > d) > 0$. Then, the **excess loss variable** (or **residual lifetime**/"future lifetime random variable for a life aged d" (in life contingencies)) is

$$Y = \begin{cases} \text{undefined} & \text{if } X \leq d; \\ X - d & \text{if } X > d \end{cases} = (X - d | X > d).$$

[Note: The excess loss variable gives the amount of loss in excess of the deductible, *given that* such excess of loss exists. If there is no such excess of loss, we *do not define* excess loss variable.

More precisely, the excess loss variable Y has the same distribution as the *conditional* distribution of $X - d$ given $X > d$.]

1.3.2 We are usually interested in studying the *expected value* of the excess loss variable, which is called the **mean excess loss function** (or **mean residual lifetime** (MRL)/"complete expectation of life" in life contingencies):

$$e_X(d) = \mathbb{E}[Y] = \mathbb{E}[X - d | X > d].$$

[Note: In life contingencies, the actuarial notation is $\mathring{e}_d$.]

1.3.3 If X is continuous with pdf f_X, then

$$e_X(d) = \frac{\mathbb{E}\left[(X-d)\mathbf{1}_{\{X>d\}}\right]}{\mathbb{P}(X>d)} = \boxed{\frac{\int_d^\infty (x-d) f_X(x)\,\mathrm{d}x}{\mathbb{P}(X>d)}}.$$

If X is discrete, then

$$e_X(d) = \boxed{\frac{\sum_{x_j>d}(x_j-d)p_j}{\mathbb{P}(X>d)}}$$

where $p_j = \mathbb{P}(X=x_j)$.

1.3.4 To compute the kth moment of the excess loss variable Y (i.e., $\mathbb{E}\left[Y^k\right]$) (denoted by $e_X^k(d)$), we can use the following formulas.

- X is continuous with pdf f_X:

$$e_X^k(d) = \boxed{\frac{\int_d^\infty (x-d)^k f_X(x)\,\mathrm{d}x}{\mathbb{P}(X>d)}}.$$

- X is discrete:

$$e_X^k(d) = \boxed{\frac{\sum_{x_j>d}(x_j-d)^k p_j}{\mathbb{P}(X>d)}}$$

where $p_j = \mathbb{P}(X=x_j)$.

1.3.5 The MRL $e_X(d)$ and the expected stop loss variable $\mathbb{E}\left[(X-d)_+\right]$ can be related as follows.

$$e_X(d) = \frac{\mathbb{E}\left[(X-d)\mathbf{1}_{\{X>d\}}\right]}{\mathbb{P}(X>d)} = \frac{\mathbb{E}\left[(X-d)\mathbf{1}_{\{X>d\}} + 0\cdot\mathbf{1}_{\{X\le d\}}\right]}{\mathbb{P}(X>d)} = \boxed{\frac{\mathbb{E}\left[(X-d)_+\right]}{\mathbb{P}(X>d)}}.$$

[Note: Using proposition 1.2.a, we can further write

$$e_X(d) = \boxed{\frac{\int_d^\infty S_X(x)\,\mathrm{d}x}{\mathbb{P}(X>d)}}.$$

]

1.4 Limited Loss Variables

1.4.1 Fix any $u \in \mathbb{R}$ and consider a loss random variable X. Then, the **limited loss variable** is

$$X \wedge u = \min\{X, u\} = \begin{cases} X & \text{if } X \le u. \\ u & \text{if } X > u; \end{cases}$$

1.4.2 For a practical interpretation of limited loss variable, consider the following. Suppose that the insurer insures a loss X with a policy limit of u dollars (i.e., the maximum loss insured is u dollars). Then, the limited loss variable represents the payment made by :

- If the loss $X \le u$, then pays the full amount u to the policyholder .

- If the loss $X > u$, then 🏢 only pays u dollars to 👤.

There is a cap of u dollars to the payment made by 🏢.

1.4.3 To compute the kth moment of the limited loss variable $X \wedge u$, we can use the following formulas.

- X is continuous with pdf f_X:

$$\mathbb{E}\left[(X \wedge u)^k\right] = \int_{-\infty}^{\infty} (x \wedge u)^k f_X(x)\,\mathrm{d}x = \int_{-\infty}^{u} x^k f_X(x)\,\mathrm{d}x + \int_{u}^{\infty} u^k f_X(x)\,\mathrm{d}x = \boxed{\int_{-\infty}^{u} x^k f_X(x)\,\mathrm{d}x + u^k \mathbb{P}(X > u)}.$$

- X is discrete:

$$\mathbb{E}\left[(X \wedge u)^k\right] = \sum_j (x_j \wedge u)^k p_j = \sum_{x_j \le u} x_j^k p_j + \sum_{x_j > u} u^k p_j = \boxed{\sum_{x_j \le u} x^k p_j + u^k \mathbb{P}(X > u)}$$

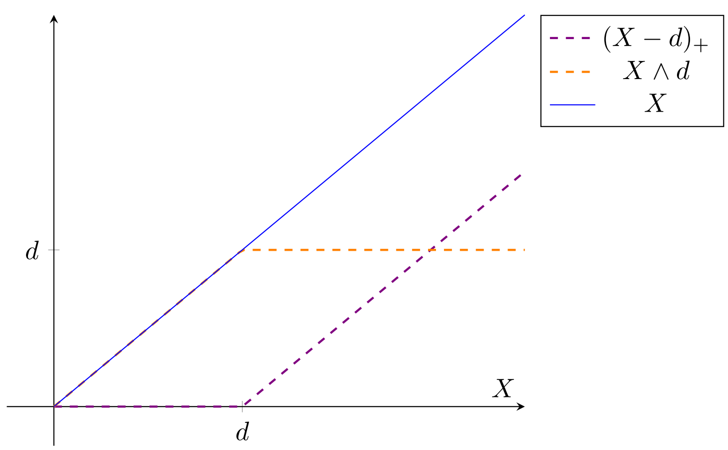
where $p_j = \mathbb{P}(X = x_j)$.

1.4.4 Stop loss and limited loss variables can be related as follows. For any $d \in \mathbb{R}$ and any random variable X,

$$(X - d)_+ + (X \wedge d) = X.$$

Proof: We have

$$(X - d)_+ + (X \wedge d) = \begin{cases} X - d + d & \text{if } X > d; \\ 0 + X & \text{if } X \le d \end{cases} = X.$$



□

[Note: A practical interpretation of this result is that combining an insurance with deductible d and another insurance with policy limit d gives an insurance with full coverage.]

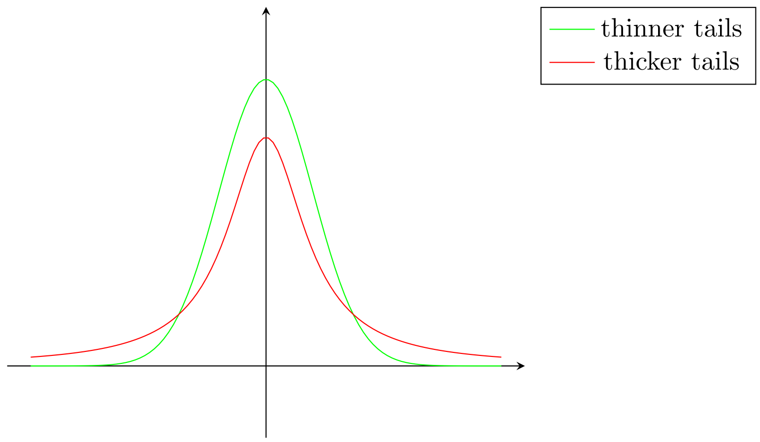
1.4.5 Due to the relationship in [1.4.4], we can obtain the following formula for computing expected limited loss (assuming X has finite mean and is nonnegative), using propositions 1.1.a and 1.2.a:

$$\mathbb{E}[X \wedge u] = \mathbb{E}[X] - \mathbb{E}[(X - u)_+] = \int_0^{\infty} S_X(x)\,\mathrm{d}x - \int_u^{\infty} S_x(x)\,\mathrm{d}x = \boxed{\int_0^{u} S_X(x)\,\mathrm{d}x}.$$

where S_X is the survival function of X.

1.5 Comparing Tail Thickness of Distributions

1.5.1 An important consideration in risk management for an insurer 🏢 is to properly quantify the "thickness" of the right tail of distribution of loss X since it can greatly impact the financial position of 🏢. The higher probability assigned to extremely large values, the "thicker"/"heavier" the right tail.



1.5.2 To *compare* tail thickness, we can consider the following methods:

(a) comparison based on existence and non-existence of moments

(b) comparison based on limit of ratio of survival functions

(c) comparison based on hazard rate function ("force of mortality" in life contingencies)

(d) comparison based on MRL (or mean excess loss function)

[Note: Here we focus on continuous random variables. But the methods can also be used for discrete random variables in a similar manner.]

1.5.3 First we consider comparison based on moments. Here we focus on a nonnegative loss X. Recall that the kth raw moment of X with pdf f_X is

$$\mu_k' = \mathbb{E}\left[X^k\right] = \int_{-\infty}^{\infty} x^k f_X(x)\,\mathrm{d}x$$

If $f_X(x)$ is relatively large for large x, it indicates that X has a relatively thick right tail. Hence, a method for assessing tail thickness is to check the *speed* of $f_X(x) \to 0$ as x rises.

When the kth moment $\mathbb{E}\left[X^k\right]$ exists/is finite (i.e., $\mathbb{E}\left[X^k\right] < \infty$), it suggests that $f_X(x) \to 0$ *much faster* than the speed at which $x^k \to \infty$ ➔ thinner right tail.

On the other hand, if it is infinite (i.e., $\mathbb{E}\left[X^k\right] = \infty$), then it indicates that $f_X(x) \to 0$ *much slower* than the speed at which $x^k \to \infty$ ➔ thicker right tail.

1.5.4 Regarding the existence/non-existence of moments, we have the following result.

Proposition 1.5.a. Suppose that X is a nonnegative random variable. Fix any $r, k > 0$ with $0 < r < k$. Then,

$$\mathbb{E}\left[X^k\right] < \infty \implies \mathbb{E}\left[X^r\right] < \infty.$$

(So, existence of kth moment implies existence of all smaller positive moments.)

[Note: Equivalently,

$$\mathbb{E}\left[X^r\right] = \infty \implies \mathbb{E}\left[X^k\right] = \infty.$$

(So, non-existence of rth moment implies non-existence of all larger positive moments.)]

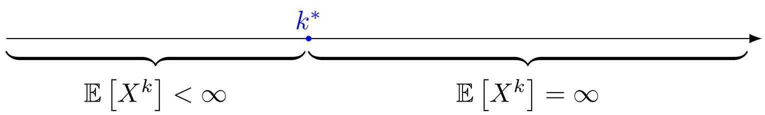
Proof: Assume that $\mathbb{E}\left[X^k\right] < \infty$. Then,

$$\begin{aligned}
\mathbb{E}\left[X^r\right] &= \mathbb{E}\left[X^r \mathbf{1}_{\{0 \le X \le 1\}}\right] + \mathbb{E}\left[X^r \mathbf{1}_{\{X > 1\}}\right] \\
&\le \mathbb{E}\left[\mathbf{1}_{\{0 \le X \le 1\}}\right] + \mathbb{E}\left[X^k \mathbf{1}_{\{X > 1\}}\right] && (X^r \le 1 \text{ when } 0 \le X \le 1, \text{ and } X^r \le X^k \text{ when } X > 1) \\
&\le \mathbb{E}\left[\mathbf{1}_{\{0 \le X \le 1\}}\right] + \mathbb{E}\left[X^k\right] && (\mathbf{1}_{\{X>1\}} \le 1 \text{ and } X^k \ge 0) \\
&= \mathbb{P}(0 \le X \le 1) + \mathbb{E}\left[X^k\right] \\
&< \infty.
\end{aligned}$$

□

Hence, if it is not the case that X has finite kth moment for any $k > 0$, then we can find a "turning point" $k^* > 0$ at which the moment changes from finite to infinite, i.e.,

$$\mathbb{E}\left[X^k\right] < \infty \ \forall 0 < k < k^* \quad \text{and} \quad \mathbb{E}\left[X^k\right] = \infty \ \forall k \ge k^*.$$



1.5.5 Thus, we have the following indicators for tail thickness:

- Existence of kth moment for any $k > 0$ ➔ "rather thin" right tail.
- Between two nonnegative losses X and Y, if the "turning point" of X is smaller than that of Y, then X has a *thicker* right tail than Y. [Note: To see this, note that in this case we can find a $k > 0$ such that $\mathbb{E}\left[X^k\right] = \infty$ and $\mathbb{E}\left[Y^k\right] < \infty$, which means that $f_X(x) \to 0$ *much slower* than $x^k \to \infty$, while $f_Y(x) \to 0$ *much faster* than $x^k \to \infty$ ➔ $f_X(x) \to 0$ *much slower* than $f_Y(x) \to 0$ ➔ X has a *thicker* right tail than Y.]

1.5.6 Now we consider comparison based on limit of ratio of survival functions. To compare the tail thickness between X and Y, we consider the limit

$$\lim_{t\to\infty} \frac{S_X(t)}{S_Y(t)}$$

where S_X and S_Y are survival functions of X and Y respectively.

Suppose the limit is c (which may be ∞). Then, we can compare the tail thickness based on c:

- $c = 0$: $S_X(t) \to 0$ *much faster* than $S_Y(t) \to 0$ as $t \to \infty$ ➔ Y has a *thicker* right tail than X
- $0 < c < \infty$: $S_X(t) \to 0$ at a "similar" speed to $S_Y(t) \to 0$ as $t \to \infty$ ➔ X and Y have "similar" right tail thickness
- $c = \infty$: $S_X(t) \to 0$ *much slower* than $S_Y(t) \to 0$ as $t \to \infty$ ➔ X has a *thicker* right tail than Y

[Note: When $S_X(t)$ is relatively large for large t, it suggests that relatively high probability is assigned to large values ➔ relatively thick right tail.]

1.5.7 To compute the limit $\lim_{t\to\infty} \frac{S_X(t)}{S_Y(t)}$, we can apply the L'Hôpital's rule:

$$\lim_{t\to\infty} \frac{S_X(t)}{S_Y(t)} = \lim_{t\to\infty} \frac{S'_X(t)}{S'_Y(t)} = \boxed{\lim_{t\to\infty} \frac{f_X(t)}{f_Y(t)}}.$$

where f_X and f_Y are pdfs of X and Y respectively.

1.5.8 Next, we consider the comparison based on hazard rate function. The **hazard rate function** of a random variable X is

$$h_X(x) = \frac{f_X(x)}{S_X(x)}$$

where f_X and S_X are pdf and survival function of X respectively.

[Note: It is "force of mortality" in life contingencies.]

To interpret the hazard rate, note that

$$\mathbb{P}(x < X \leq x + h | X > x) = \frac{\mathbb{P}(x < X \leq x + h)}{S_X(x)} \approx \frac{f(x)h}{S_X(x)} = h_X(x)h$$

for small h. Thus, in infinitesimal, $h_X(x)\,\mathrm{d}x$ gives the probability that $X \in (x, x + \mathrm{d}x]$ ("failing instantaneously") given $X > x$ ("surviving"). Hence, the hazard rate can be regarded as *instantaneous failure rate*.

1.5.9 To assess the tail thickness based on hazard rate function, consider the following.

- If $h_X(x)$ is a decreasing (or non-increasing) function, then for larger x, given $X > x$, it is *less* likely for $X \in (x, x+\mathrm{d}x]$ and hence *more* likely for $X > x+\mathrm{d}x$ (higher probability for *even larger* x) ➔ X has a thick right tail.
- If $h_X(x)$ is an increasing (or non-decreasing) function, then for larger x, given $X > x$, it is *more* likely for $X \in (x, x + \mathrm{d}x]$ and hence *less* likely for $X > x + \mathrm{d}x$ (lower probability for *even larger* x) ➔ X has a thin right tail.

[Note: Here "less" and "more" likely are in a non-strict sense: They also include the case "as likely".]

1.5.10 A random variable X has **decreasing failure rate** (DFR) if $h_X(x)$ is decreasing in x, and has **increasing failure rate** (IFR) if $h_X(x)$ is increasing in x.

Hence, X has thick (thin) right tail if X has DFR (IFR).

1.5.11 Lastly, we consider the comparison based on mean residual lifetime (MRL).

- If $e_X(d)$ is decreasing in d, then it suggests that there is not much excess loss (on average) once d is large ➔ not much probability is assigned to large x ➔ X has a thin right tail.
- If $e_X(d)$ is increasing in d, then it suggests that there is still more excess loss (on average) when d gets larger ➔ quite a lot of probability is assigned to large x ➔ X has a thick right tail.

1.5.12 A random variable X has **decreasing mean residual lifetime** (DMRL) if $e_X(d)$ is decreasing in d, and has **increasing mean residual lifetime** (IMRL) if $e_X(d)$ is increasing in d.

Hence, X has thick (thin) right tail if X has IMRL (DMRL).

1.5.13 Naturally, one would then be interested in the relationship between DFR/IFR and DMRL/IMRL. This is given as follows.

Proposition 1.5.b. Let X be a random variable. Then,

$$X \text{ has DFR} \implies X \text{ has IMRL} \quad \text{and} \quad X \text{ has IFR} \implies X \text{ has DMRL}.$$

Proof: Firstly, since

$$h_X(x) = \frac{f_X(x)}{S_X(x)} = -\frac{\mathrm{d}}{\mathrm{d}x} \ln S_X(x),$$

we have

$$S_X(x) = \exp\left(-\int_{-\infty}^{x} h_X(y)\,\mathrm{d}y\right),$$

so

$$\frac{S_X(x+t)}{S_X(x)} = \exp\left(-\int_x^{x+t} h_X(y)\,\mathrm{d}y\right) = \exp\left(-\int_0^t h_X(x+y)\,\mathrm{d}y\right).$$

[Note: The formula corresponding to this in life contingencies is $S_x(t) = \exp\left(-\int_0^t \mu_{x+s}\,\mathrm{d}s\right)$.]

Now, assume X has DFR and fix any $t \geq 0$. Then, from the equation above we know $\dfrac{S_X(x+t)}{S_X(x)}$ is increasing in x. Thus, by [1.3.5], for any $d_1 \leq d_2$,

$$e_X(d_1) = \frac{\int_0^\infty S_X(d_1+t)\,\mathrm{d}t}{S_X(d_1)} = \int_0^\infty \frac{S_X(d_1+t)}{S_X(d_1)}\,\mathrm{d}t \leq \int_0^\infty \frac{S_X(d_2+t)}{S_X(d_2)}\,\mathrm{d}t = e_X(d_2),$$

meaning that $e_X(d)$ is increasing in d, so X is IMRL.

Proof of another implication is similar. □

[⚠ Warning: The converse implications are <u>not</u> true. That is,

$$X \text{ has IMRL } \not\Rightarrow X \text{ has DFR} \quad \text{and} \quad X \text{ has DMRL } \not\Rightarrow X \text{ has IFR}.$$

]

2 Mixing and Conditional Expectation

2.1 Mixing

2.1.1 Consider n random variables $X_1, \dots, X_n$ which are all continuous (all discrete). We can create a new distribution (random variable X) by *mixing* them, whose pdf (pmf) is

$$f_X(x) = p_1 f_{X_1}(x) + \cdots + p_n f_{X_n}(x)$$

where f_{X_i} is the pdf (pmf) of X_i ($i = 1, \dots, n$), $p_1, \dots, p_n \geq 0$, and $p_1 + \cdots + p_n = 1$. We can interpret the distribution resulting from mixing by:

$$X = \begin{cases} X_1 & \text{with probability } p_1; \\ \vdots & \vdots \\ X_n & \text{with probability } p_n. \end{cases}$$

2.1.2 By introducing another discrete random variable Λ with support $\{\lambda_1, \dots, \lambda_n\}$ such that

$$X_i \stackrel{d}{=} (X|\Lambda = \lambda_i) \quad \text{for any } i = 1, \dots, n,$$

we can associate $f_{X_i}(x)$ with $f_{X|\Lambda}(x|\lambda_i)$ and p_i with $\mathbb{P}(\Lambda = \lambda_i)$:

$$f_X(x) = \sum_{i=1}^{n} f_{X|\Lambda}(x|\lambda_i)\mathbb{P}(\Lambda = \lambda_i) = \boxed{\sum_{i=1}^{n} f_{X|\Lambda}(x|\lambda_i)p_i}.$$

[Note: $f_{X|\Lambda}(x|\lambda_i)$ is the conditional probability function of X given $\Lambda = \lambda_i$.]

(The equality holds since each summand is the joint probability function $f_{X,\Lambda}(x, \lambda_i)$i, so summing them up gives the marginal probability function $f_{X_i}(x)$.)

This induces mixing of finitely many (all continuous/all discrete) random variables using (continuous or discrete) X and discrete Λ.

[Note: Here the notion of conditional distribution includes also the case where one of the two random variables involved is discrete while another is continuous. (This may not be defined in elementary probability theory, but this can be allowed through more advanced probability theory. So one may take it as given.)]

2.1.3 To induce mixing of countably infinitely many random variables, we can introduce a discrete random variable Λ with support $\{\lambda_1, \lambda_2, \dots\}$. Then, the pdf/pmf of X can similarly be written as

$$f_X(x) = \boxed{\sum_{i=1}^{\infty} f_{X|\Lambda}(x|\lambda_i)p_i}$$

where $p_i = \mathbb{P}(\Lambda = \lambda_i)$ for any $i \in \mathbb{N}$.

2.1.4 To induce mixing of uncountably infinitely many random variables, we can introduce a continuous random variable Λ. Then, the pdf/pmf of X can be written as

$$f_X(x) = \boxed{\int_{-\infty}^{\infty} f_{X|\Lambda}(x|\lambda) f_\Lambda(\lambda)\,\mathrm{d}\lambda}.$$

(Here the integrand is the joint probability function $f_{X,\Lambda}(x, \lambda)$, so integrating it gives the marginal probability function $f_X(x)$.)

2.1.5 We can also express the mixing using cdf as suggested by the following result.

Proposition 2.1.a. Let X and Λ be random variables. Then,

- (Λ is continuous)
$$F_X(x) = \int_{-\infty}^{\infty} F_{X|\Lambda}(x|\lambda) f_\Lambda(\lambda)\,\mathrm{d}\lambda\,.$$
- (Λ is discrete)
$$F_X(x) = \sum_i F_{X|\Lambda}(x|\lambda_i) p_i$$
where $p_i = \mathbb{P}(\Lambda = \lambda_i)$ and the sum is taken over all i where $p_i > 0$.

[Note: F_X is the cdf of X and $F_{X|\Lambda}(x|\lambda)$ is the conditional cdf of X given $\Lambda = \lambda$.]

Proof: We prove only the continuous case. The discrete case can be proven similarly. Note that

$$\begin{aligned}
F_X(x) &= \int_{-\infty}^{x} f_X(t)\,\mathrm{d}t \\
&= \int_{-\infty}^{x}\int_{-\infty}^{\infty} f_{X|\Lambda}(t|\lambda) f_\Lambda(\lambda)\,\mathrm{d}\lambda\,\mathrm{d}t \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{x} f_{X|\Lambda}(t|\lambda) f_\Lambda(\lambda)\,\mathrm{d}t\,\mathrm{d}\lambda \\
&= \int_{-\infty}^{\infty} f_\Lambda(\lambda) \int_{-\infty}^{x} f_{X|\Lambda}(t|\lambda)\,\mathrm{d}t\,\mathrm{d}\lambda \\
&= \int_{-\infty}^{\infty} f_\Lambda(\lambda) F_{X|\Lambda}(x|\lambda)\,\mathrm{d}\lambda\,.
\end{aligned}$$

□

2.2 Conditional Expectation

2.2.1 This sections serves as a review on the concept of conditional expectation, which is covered in STAT2901.

2.2.2 Consider again the random variables X and Λ. Recall that the conditional expectation of $g(X)$ given $\Lambda = \lambda$ is

$$\mathbb{E}[g(X)|\Lambda = \lambda] = \begin{cases} \displaystyle\int_{-\infty}^{\infty} g(x) f_{X|\Lambda}(x|\lambda)\,\mathrm{d}x & \text{if } X \text{ and } \Lambda \text{ are both continuous;} \\ \displaystyle\sum_i g(x_i) f_{X|\Lambda}(x_i|\lambda) & \text{if } X \text{ and } \Lambda \text{ are both discrete} \end{cases}$$

where the sum is taken over all i where $f_{X|\Lambda}(x_i|\lambda) > 0$.

Write $h(\lambda) = \mathbb{E}[g(X)|\Lambda = \lambda]$. Then, the conditional expectation of $g(X)$ given Λ is denoted by $\mathbb{E}[g(X)|\Lambda]$ and is a *random variable* (as a function of Λ):

$$h(\Lambda) = \mathbb{E}[g(X)|\Lambda]$$

which takes the value $h(\lambda)$ when $\Lambda = \lambda$.

[Note: Practically, to obtain an expression of $\mathbb{E}[g(X)|\Lambda]$, we can first find an expression for $\mathbb{E}[g(X)|\Lambda = \lambda]$ and replace every λ by Λ.]

2.2.3 Next, recall that the conditional variance of $g(X)$ given $\Lambda = \lambda$ is

$$\mathrm{Var}(g(X)|\Lambda = \lambda) = \mathbb{E}\left[(g(X) - (\mathbb{E}[g(X)|\Lambda = \lambda])^2|\Lambda = \lambda\right] = \mathbb{E}\left[g(X)^2|\Lambda = \lambda\right] - (\mathbb{E}[g(X)|\Lambda = \lambda])^2.$$

[Note: Similarly, we can write $h(\lambda) = \mathrm{Var}(g(X)|\Lambda = \lambda)$, and $\mathrm{Var}(g(X)|\Lambda)$ is the random variable $h(\Lambda)$. Thus,

$$\mathrm{Var}(g(X)|\Lambda) = \mathbb{E}\left[(g(X))^2|\Lambda\right] - (\mathbb{E}[g(X)|\Lambda])^2.$$

]

2.2.4 Two remarkable results related to conditional expectation and conditional variance are *law of total expectation* and *law of total variance*.

2.2.5 Law of total expectation relates unconditional and conditional expectations.

Theorem 2.2.a (Law of total expectation)**.** For any function g and any random variables X and Λ (where $\mathbb{E}[g(X)]$ is finite),

$$\mathbb{E}[\mathbb{E}[g(X)|\Lambda]] = \mathbb{E}[g(X)].$$

Proof: We only prove for the case where X and Λ are both continuous. (The case where both are discrete can be proved similarly.[1]) Let $h(\lambda) = \mathbb{E}[g(X)|\Lambda = \lambda]$. Consider:

$$\begin{aligned}
\mathbb{E}[g(X)] &= \int_{-\infty}^{\infty} g(x) f_X(x)\,\mathrm{d}x \\
&= \int_{-\infty}^{\infty} g(x) \int_{-\infty}^{\infty} f_{X|\Lambda}(x|\lambda) f_\Lambda(\lambda)\,\mathrm{d}\lambda\,\mathrm{d}x \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x) f_{X|\Lambda}(x|\lambda) f_\Lambda(\lambda)\,\mathrm{d}\lambda\,\mathrm{d}x \\
&= \int_{-\infty}^{\infty} f_\Lambda(\lambda) \int_{-\infty}^{\infty} g(x) f_{X|\Lambda}(x|\lambda)\,\mathrm{d}x\,\mathrm{d}\lambda \\
&= \int_{-\infty}^{\infty} h(\lambda) f_\Lambda(\lambda)\,\mathrm{d}\lambda \\
&= \mathbb{E}[h(\Lambda)] \\
&= \mathbb{E}[\mathbb{E}[g(X)|\Lambda]].
\end{aligned}$$

□

2.2.6 Law of total variance relates unconditional variance and conditional mean & variance:

Proposition 2.2.b (Law of total variance)**.** For any random variables X (with finite variance) and Λ,

$$\mathrm{Var}(X) = \mathbb{E}[\mathrm{Var}(X|\Lambda)] + \mathrm{Var}(\mathbb{E}[X|\Lambda]).$$

Proof: Note that

$$\begin{aligned}
\mathrm{Var}(X) &= \mathbb{E}\left[X^2\right] - (\mathbb{E}[X])^2 \\
&= \mathbb{E}[\mathbb{E}\left[X^2|\Lambda\right]] - (\mathbb{E}[\mathbb{E}[X|\Lambda]])^2 \\
&= \mathbb{E}[\mathrm{Var}(X|\Lambda) + (\mathbb{E}[X|\Lambda])^2] - (\mathbb{E}[\mathbb{E}[X|\Lambda]])^2 \\
&= \mathbb{E}[\mathrm{Var}(X|\Lambda)] + \mathbb{E}\left[(\mathbb{E}[X|\Lambda])^2\right] - (\mathbb{E}[\mathbb{E}[X|\Lambda]])^2 \\
&= \mathbb{E}[\mathrm{Var}(X|\Lambda)] + \mathrm{Var}(\mathbb{E}[X|\Lambda]).
\end{aligned}$$

□

[1] A general proof where both random variables can be arbitrary (as long as $\mathbb{E}[g(X)]$ is finite) is out of scope.

3 Basic Frequency Models

3.1 Preliminaries

3.1.1 Suppose that an insurer sells a type of insurance to policyholders which provides some payment on each *claim* (of suffering some loss protected by the insurance) filed by policyholders.

[Note: A single policyholder may file *multiple* claims → multiple payments are possible and the *number* of claims is random.]

3.1.2 Let N be the random variable representing the number of *claims* filed by a certain policyholder in a certain year (*claim frequency*). Then, N is a nonnegative discrete random variable. (It is possible for to file no claim in a certain year.)

3.1.3 Let $p_k = \mathbb{P}(N = k)$ ($k \in \mathbb{N}_0$[2]) be the pmf of N. Then, the **probability generating function** (pgf) of nonnegative discrete N is

$$P_N(t) = \mathbb{E}\left[t^N\right] = \sum_{k=0}^{\infty} t^k p_k.$$

Remarks:

- Note that $P_N(1) = \mathbb{E}[1] = 1$ always.
- In this context, we set $0^0 = 1$. So, $P_N(0) = p_0$.
- Pgf is only defined for nonnegative discrete random variable.

3.1.4 The following result suggests the "probability generating" property of pgf:

Theorem 3.1.a. Let P_N be the pgf of a nonnegative discrete random variable N. Then, the pmf of N is given by

$$p_m = \frac{P_N^{(m)}(0)}{m!}$$

for any $m \in \mathbb{N}_0$ (where $P_N^{(m)}$ denote the mth derivative of P_N with $P_N^{(0)} = P_N$).

Proof: Fix any $m \in \mathbb{N}_0$. Firstly, if $m = 0$, we have $p_0 = P_N(0) = \dfrac{P_N(0)}{0!}$. So, henceforth consider the case where $m \in \mathbb{N}$. Now, note that for any $k \in \mathbb{N}_0$,

$$\frac{\mathrm{d}^m}{\mathrm{d}t^m} t^k p_k = \begin{cases} k(k-1)\cdots(k-m+1)t^{k-m} & \text{if } k \geq m \\ 0 & \text{if } k < m. \end{cases}$$

Thus,

$$P_N^{(m)}(t) = \frac{\mathrm{d}^m}{\mathrm{d}t^m} \sum_{k=0}^{\infty} t^k p_k = \sum_{k=0}^{\infty} \frac{\mathrm{d}^m}{\mathrm{d}t^m} t^k p_k = \sum_{k=m}^{\infty} [k(k-1)\cdots(k-m+1)t^{k-m} p_k].$$

Hence,

$$P_N^{(m)}(0) = m(m-1)\cdots(m-m+1)(1)p_m = m! p_m \implies p_m = \frac{P_N^{(m)}(0)}{m!}.$$

[Note: Recall that $0^0 = 1$.] □

3.1.5 As a corollary of theorem 3.1.a, pgf gives a sufficient condition for equality in distribution:

Corollary 3.1.b. Let M and N be two nonnegative discrete random variables with pgf P_M and P_N. If they have the same pgf, then they have the same distribution.

[2] $\mathbb{N}_0 = \{0, 1, \dots\}$

Proof: By assumption, we have $P_M^{(m)}(0) = P_N^{(m)}(0)$ for any $m \in \mathbb{N}_0$. Thus, by theorem 3.1.a, for any $m \in \mathbb{N}_0$,

$$\mathbb{P}(M = m) = \mathbb{P}(N = m),$$

which means that M and N have the same distribution. $\square$

3.1.6 Pgf also has a (partial) "moment generating" property as follows.

Proposition 3.1.c. Let P_N be the pgf of a nonnegative discrete random variable N. Then,

$$P_N'(1) = \mathbb{E}[N] \quad \text{and} \quad P_N''(1) = \mathbb{E}[N(N-1)].$$

Proof: Firstly, we have

$$P_N'(t) = p_1 + 2p_2 t + 3p_3 t^2 + \cdots,$$

which implies

$$P_N'(1) = p_1 + 2p_2 + 3p_3 + \cdots = \mathbb{E}[N].$$

Next,

$$P_N''(t) = 2p_2 + 3(2)p_3 t + 4(3)p_4 t^2 + \cdots,$$

implying that

$$P_N''(1) = 1(0)p_1 + 2(1)p_2 + 3(2)p_3 + 4(3)p_4 + \cdots = \mathbb{E}[N(N-1)].$$

$\square$

3.1.7 The pgf of a sum of independent nonnegative discrete random variables can be obtained by the following formula.

Proposition 3.1.d. Let $N_1, \ldots, N_m$ be independent nonnegative discrete random variables, and let $S = N_1 + \cdots + N_m$. Then,

$$P_S(t) = P_{N_1}(t) \cdots P_{N_m}(t)$$

for any t (at which all terms exist), where P_N denotes the pgf of N.

Proof: First note that $t^{N_1}, \ldots, t^{N_m}$ are also independent. Thus,

$$P_S(t) = \mathbb{E}\left[t^S\right] = \mathbb{E}\left[t^{N_1 + \cdots + N_m}\right] = \mathbb{E}\left[t^{N_1} \cdots t^{N_m}\right] = \mathbb{E}\left[t^{N_1}\right] \cdots \mathbb{E}\left[t^{N_m}\right] = P_{N_1}(t) \cdots P_{N_m}(t).$$

$\square$

3.1.8 Starting from here, we will discuss several kinds of probability distributions for modelling the number of claims N (*frequency models*):

(a) Poisson distribution
(b) mixed Poisson distribution
(c) negative binomial distribution
(d) geometric distribution
(e) binomial distribution

3.2 The Poisson Distribution

3.2.1 A random variable N follows the **Poisson distribution** with parameter $\lambda > 0$ (denoted by $N \sim \text{Poi}(\lambda)$) if its pmf is

$$p_k = \mathbb{P}(N = k) = \frac{e^{-\lambda}\lambda^k}{k!}$$

for any $k \in \mathbb{N}_0$.

3.2.2 The pgf of $N \sim \text{Poi}(\lambda)$ is

$$P_N(t) = \mathbb{E}\left[t^N\right] = \boxed{e^{\lambda(t-1)}}.$$

Proof: Note that

$$P_N(t) = \sum_{k=0}^{\infty} t^k p_k = \sum_{k=0}^{\infty} t^k \frac{e^{-\lambda}\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda t)^k}{k!} = e^{-\lambda} e^{\lambda t}.$$

(Recall that $e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$.) □

3.2.3 We can then deduce its mean and variance based on pgf:

$$\mathbb{E}[N] = \text{Var}(N) = \boxed{\lambda}.$$

Proof: Since $P_N(t) = e^{\lambda(t-1)}$, we have

$$P_N'(t) = \lambda e^{\lambda(t-1)} \quad \text{and} \quad P_N''(t) = \lambda^2 e^{-\lambda(t-1)}.$$

Therefore, by proposition 3.1.c,

$$\mathbb{E}[N] = P_N'(1) = \lambda \quad \text{and} \quad \mathbb{E}[N(N-1)] = P_N''(1) = \lambda^2.$$

Since $\mathbb{E}[N(N-1)] = \mathbb{E}\left[N^2\right] - \mathbb{E}[N] = \mathbb{E}\left[N^2\right] - \lambda$, it follows that

$$\text{Var}(N) = \mathbb{E}\left[N^2\right] - (\mathbb{E}[N])^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

□

3.2.4 Here we introduce two remarkable results for a Poisson distribution: *convolution* and *thinning/decomposition*.

3.2.5 The convolution result concerns *sum* of independent Poisson random variables. It turns out that such sum is *also* Poisson distributed.

Theorem 3.2.a. Let $N_1, \dots, N_k$ be k independent Poisson random variables with parameters $\lambda_1, \dots, \lambda_k$ respectively. Then, the sum $S = N_1 + \cdots + N_k$ follows the Poisson distribution with parameter $\lambda_1 + \cdots + \lambda_k$.

Proof: We prove this using pgf. We denote pgf of N by $P_N(t)$. By proposition 3.1.d, we have

$$P_S(t) = P_{N_1}(t) \cdots P_{N_k}(t).$$

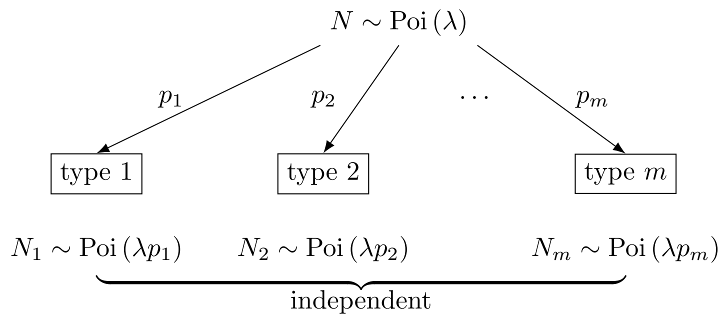
Now using [3.2.2], we can further write

$$P_S(t) = e^{\lambda_1(t-1)} \cdots e^{\lambda_k(t-1)} = e^{(\lambda_1 + \cdots + \lambda_k)(t-1)}.$$

But note that this is exactly the same as the pgf of (a random variable following) $\text{Poi}(\lambda_1 + \cdots + \lambda_k)$ distribution. Hence, by corollary 3.1.b, we conclude that $S \sim \text{Poi}(\lambda_1 \cdots + \lambda_k)$. □

3.2.6 The thinning/decomposition result concerns the "finer pieces" obtained from "slicing"/"decomposing" a Poisson random variable ("thinning"). It turns out that by performing the "slicing" in a certain way, the resulting "finer pieces" of random variables are also Poisson distributed.

Theorem 3.2.b. Let $N \sim \text{Poi}(\lambda)$ be the number of claims. Suppose that each claim is independently classified into exactly one of type $1, \dots, m$ with probabilities $p_1, \dots, p_m$ respectively (where $p_1 + \cdots + p_m = 1$). Let $N_1, \dots, N_m$ be the numbers of claims with types $1, \dots, m$ respectively. Then, $N_1, \dots, N_m$ are independent Poisson random variables with parameters $\lambda p_1, \dots, \lambda p_m$ respectively.



3.3 The Mixed Poisson Distribution

3.3.1 For the mixed Poisson distribution, as its name suggests, it involves *mixing* of Poisson random variables. In this case, we consider mixing of uncountably infinitely many Poisson random variables, so an auxiliary *continuous* random variable is introduced.

3.3.2 A random variable N follows a **mixed Poisson distribution** if

$$N|\Theta \sim \text{Poi}(g(\Theta))$$

where Θ is a continuous random variable (with pdf f_Θ).

[Note: This notation means that the conditional distribution of N given $\Theta = \theta$ is Poisson distribution with parameter $g(\theta)$, i.e.,

$$(N|\Theta = \theta) \sim \text{Poi}(g(\theta)).$$

]

3.3.3 Based on this definition, we know

$$\mathbb{E}[N|\Theta = \theta] = \text{Var}(N|\Theta = \theta) = g(\theta) \quad \text{or} \quad \mathbb{E}[N|\Theta] = \text{Var}(N|\Theta) = g(\Theta).$$

Hence, we can obtain the following formulas for $\mathbb{E}[N]$ and $\text{Var}(N)$ where N is a mixed Poisson random variable:

- By law of total expectation, $\mathbb{E}[N] = \mathbb{E}[\mathbb{E}[N|\Theta]] = \boxed{\mathbb{E}[g(\Theta)]} = \int_{-\infty}^{\infty} g(\theta) f_\Theta(\theta)\,d\theta.$
- By law of total variance, $\text{Var}(N) = \mathbb{E}[\text{Var}(N|\Theta)] + \text{Var}(\mathbb{E}[N|\Theta]) = \boxed{\mathbb{E}[g(\Theta)] + \text{Var}(g(\Theta))}.$

[Note: Assuming $g(\Theta)$ is random (which is almost always the case), the variance $\text{Var}(g(\Theta)) > 0$. Thus, in this case we have

$$\text{Var}(N) > \mathbb{E}[N]$$

(unlike the case where N is Poisson distributed). A practical implication of this result is that mixed Poisson distribution can better model claim frequency that is "relatively more disperse" (variance higher than mean), than a Poisson distribution (which forces equality of variance and mean).]

3.3.4 An important special case is when $g(\theta) = \lambda\theta$ for some *scale parameter* $\lambda > 0$. In this case, we have

$$N|\Theta \sim \text{Poi}(\lambda\Theta).$$

We can then derive the following result.

Proposition 3.3.a. In this case, the pgf of N is given by

$$P_N(t) = M_\Theta(\lambda(t-1))$$

where $M_\Theta(s) = \mathbb{E}\left[e^{s\Theta}\right]$ denotes the moment generating function of Θ.

Proof: By law of total expectation,

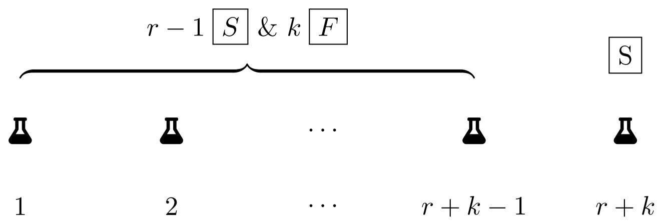
$$P_N(t) = \mathbb{E}\left[t^N\right] = \mathbb{E}\left[\mathbb{E}\left[t^N \middle| \Theta\right]\right] = \mathbb{E}\left[e^{\lambda\Theta(t-1)}\right] = M_\Theta(\lambda(t-1)).$$

□

3.4 The Negative Binomial Distribution

3.4.1 In elementary probability course, a negative binomial random variable $N \sim \mathrm{NB}(r, p)$ is defined as the number of failures before rth success in a sequence of independent Bernoulli trials[3] with success probability $p \in (0, 1)$ (neither 0 nor 1 so that there is randomness). Under this definition, the pmf of N is

$$p_k = \mathbb{P}(N = k) = \binom{r+k-1}{k}(1-p)^k p^r.$$



Here the conditions on the parameters r and p are $r \in \mathbb{N}$ and $0 < p < 1$.

3.4.2 Here we consider a more general definition of negative binomial distribution. A random variable N follows the **negative binomial distribution** with parameters $r > 0$ and $\beta > 0$ (denoted by $N \sim \mathrm{NB}(r, \beta)$) if its pmf is given by

$$p_k = \mathbb{P}(N = k) = \binom{k+r-1}{k}\left(\frac{\beta}{1+\beta}\right)^k \left(\frac{1}{1+\beta}\right)^r \tag{1}$$

for any $k \in \mathbb{N}_0$.

Remarks:

- Here, when we use the notation $\mathrm{NB}(\cdot, \cdot)$, it carries the meaning of $\mathrm{NB}(r, \beta)$ here instead of $\mathrm{NB}(r, p)$ from the definition in [3.4.1].
- To move from the definition in [3.4.1] to the definition here, we allow r to be any positive*real* number and reparametrize $p \in (0, 1)$ by $\dfrac{1}{1+\beta}$ where $\beta > 0$.
- Here the binomial coefficient $\dbinom{x}{k}$ has the *general* definition (which permits x to be any real number and k be any nonnegative integer):

$$\binom{x}{k} = \begin{cases} \dfrac{x(x-1)\cdots(x-k+1)}{k!} & \text{if } k \neq 0; \\ 1 & \text{if } k = 0. \end{cases}$$

3.4.3 We can express eq. (1) in an alternative way using *gamma function*. The gamma function Γ is defined by

$$\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}\,\mathrm{d}t$$

which notably satisfies the recursive relationship

$$\Gamma(x+1) = x\Gamma(x)$$

for any $x > 0$. [Note: It serves as a generalization to factorial. (We have $\Gamma(1) = \int_0^\infty e^{-t}\,\mathrm{d}t = 1$, so by the recursive relationship we have $\Gamma(n) = (n-1)!$ for any $n \in \mathbb{N}$.)]

[3] i.e., experiments with two possible outcomes: "success" and "failures"

Then, note that

$$\binom{k+r-1}{k} = \frac{(k+r-1)(k+r-2)\cdots r}{k!} = \frac{(k+r-1)(k+r-2)\cdots r\Gamma(r)}{\Gamma(r)k!} = \frac{\Gamma(k+r)}{\Gamma(r)k!}.$$

[Note: The equality also holds when $k=0$ as $\dfrac{\Gamma(0+r)}{\Gamma(r)0!} = 1$.]

Thus, we can write eq. (1) as

$$p_k = \frac{\Gamma(k+r)}{\Gamma(r)k!}\left(\frac{\beta}{1+\beta}\right)^k\left(\frac{1}{1+\beta}\right)^r.$$

3.4.4 The pgf of $N \sim \mathrm{NB}(r,\beta)$ is given by

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

Proof: Note that

$$\begin{aligned}
P_N(t) &= \mathbb{E}\left[t^N\right] \\
&= \sum_{k=0}^{\infty} t^k \binom{k+r-1}{k}\left(\frac{\beta}{1+\beta}\right)^k\left(\frac{1}{1+\beta}\right)^r \\
&= \left(\frac{1}{1+\beta}\right)^r \sum_{k=0}^{\infty}\binom{k+r-1}{k}\left(\frac{t\beta}{1+\beta}\right)^k \\
&= \left(\frac{1}{1+\beta}\right)^r\left(1-\frac{t\beta}{1+\beta}\right)^{-r} && \text{(negative binomial series formula)} \\
&= \left(\frac{1}{1+\beta}\right)^r\left(\frac{1+\beta}{1-(t-1)\beta}\right)^r \\
&= [1-\beta(t-1)]^{-r}.
\end{aligned}$$

□

3.4.5 Based on the pgf, we know

- $P_N'(t) = r\beta[1-\beta(t-1)]^{-r-1}$.
- $P_N''(t) = r(r+1)\beta^2[1-\beta(t-1)]^{-r-2}$.

We can thus obtain the mean and variance of $N \sim \mathrm{NB}(r,\beta)$ as follows.

- $\mathbb{E}[N] = P_N'(1) = \boxed{r\beta}$.
- $\mathbb{E}[N(N-1)] = P_N''(1) = r(r+1)\beta^2 \implies \mathrm{Var}(N) = \mathbb{E}\left[N^2\right] - (\mathbb{E}[N])^2 = r(r+1)\beta^2 + r\beta - (r\beta)^2 = \boxed{r\beta(1+\beta)}$.

[Note: Since $r\beta^2 > 0$, we have again $\mathrm{Var}(N) > \mathbb{E}[N]$ in this case. So, negative binomial distribution can also be used to model "relatively more disperse" claim frequency.]

3.4.6 It turns out that a negative binomial random variable is actually a special case of mixed Poisson random variable, as suggested below.

Proposition 3.4.a. Let N and Λ be two random variables where $N|\Lambda \sim \mathrm{Poi}(\Lambda)$ and $\Lambda \sim \mathrm{Gamma}(\alpha,\theta)$. Then,

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

Remarks:

- $\text{Gamma}(\alpha, \theta)$ denotes the gamma distribution with shape parameter α and scale parameter θ. The pdf of $\Lambda \sim \text{Gamma}(\alpha, \theta)$ is given by

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

(The pdf is zero elsewhere).

- This suggests that a mixed Poisson random variable with mixing random variable following gamma distribution is negative binomial distributed.

Proof: For any $k \in \mathbb{N}_0$, by [2.1.4],

$$\begin{aligned}
\mathbb{P}(N = k) &= \int_0^\infty f_{N|\Lambda}(k|\lambda) f_\Lambda(\lambda)\,\mathrm{d}\lambda \\
&= \int_0^\infty \frac{e^{-\lambda}\lambda^k}{k!} \cdot \frac{\lambda^{\alpha-1}e^{-\lambda/\theta}}{\Gamma(\alpha)\theta^\alpha}\,\mathrm{d}\lambda \\
&= \frac{\Gamma(k+\alpha)}{\Gamma(\alpha)k!}\left(\frac{\theta}{1+\theta}\right)^k \left(\frac{1}{1+\theta}\right)^\alpha \int_0^\infty \underbrace{\frac{\lambda^{k+\alpha-1}e^{-\lambda(1+\theta)/\theta}}{\Gamma(k+\alpha)[\theta/(1+\theta)]^{k+\alpha}}}_{\text{pdf of Gamma}\,(k+\alpha,\,\theta/(1+\theta))}\,\mathrm{d}\lambda \\
&= \frac{\Gamma(k+\alpha)}{\Gamma(\alpha)k!}\left(\frac{\theta}{1+\theta}\right)^k \left(\frac{1}{1+\theta}\right)^\alpha,
\end{aligned}$$

which implies that $N \sim \text{NB}(\alpha, \theta)$. □

3.5 The Geometric Distribution

3.5.1 The geometric distribution is the special case of the negative binomial distribution when $r = 1$. More explicitly, a random variable N follows the **geometric distribution** with parameter $\beta > 0$ (denoted by $N \sim \text{Geom}(\beta)$) if its pmf is given by

$$p_k = \mathbb{P}(N = k) = \left(\frac{\beta}{1+\beta}\right)^k \left(\frac{1}{1+\beta}\right)$$

for any $k \in \mathbb{N}_0$.

[Note: As a special case of negative binomial distribution, the random variable $N \sim \text{Geom}(\beta)$ can be interpreted as the number of failures in a sequence of independent Bernoulli trials before the *first* success (with success probability being $1/(1+\beta)$).]

3.5.2 By [3.4.4], the pgf of $N \sim \text{Geom}(\beta)$ is given by

$$P_N(t) = \boxed{[1 - \beta(t-1)]^{-1}}.$$

3.5.3 By [3.4.5], the mean and variance of $N \sim \text{Geom}(\beta)$ are given by:

- $\mathbb{E}[N] = \boxed{\beta}$.
- $\text{Var}(N) = \boxed{\beta(1+\beta)}$.

3.5.4 By proposition 3.4.a, a geometric random variable can be regarded as a mixed Poisson random variable with the mixing variable following *exponential* distribution:

$$N|\Lambda \sim \text{Poi}(\Lambda) \quad \text{and} \quad \Lambda \sim \text{Exp}(\theta) \implies N \sim \text{Geom}(\theta).$$

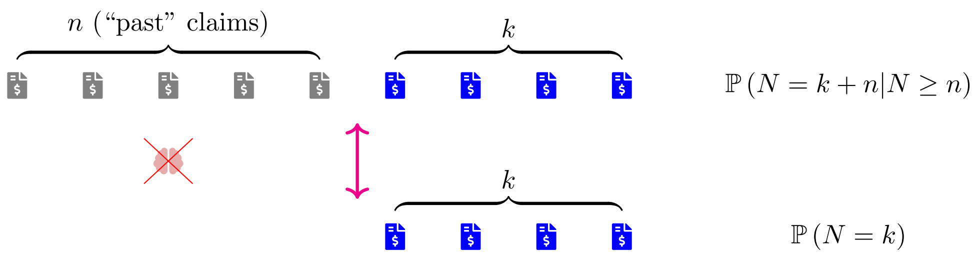
[Note: We have $\text{Gamma}(1, \theta) \equiv \text{Exp}(\theta)$.]

3.5.5 A remarkable property of geometric distribution is the *memoryless property.*

Proposition 3.5.a. If $N \sim \text{Geom}(\beta)$, then

$$\mathbb{P}(N = k + n | N \geq n) = \mathbb{P}(N = k)$$

for any $k, n \in \mathbb{N}_0$.



Proof: First denote the success probability by $p = 1/(1+\beta)$. Then, $p_k = (1-p)^k p$. Thus,

$$\begin{aligned}
\mathbb{P}(N = k + n | N \geq n) &= \frac{\mathbb{P}(N = k + n \cap N \geq n)}{\mathbb{P}(N \geq n)} \\
&= \frac{\mathbb{P}(N = k + n)}{\mathbb{P}(N \geq n)} \\
&= \frac{(1-p)^{k+n} \cancel{p}}{\sum_{i=n}^{\infty} (1-p)^i \cancel{p}} \\
&= \frac{(1-p)^{k+n}}{(1-p)^n / (1 - (1-p))} \\
&= (1-p)^k p \\
&= \mathbb{P}(N = k).
\end{aligned}$$

□

3.6 The Binomial Distribution

3.6.1 Consider m independent Bernoulli trials with the same success probability q. Then, in the m trials, the number of successes N (random variable) follows the **binomial distribution** with parameters m and q (denoted by $N \sim \text{Bin}(m, q)$).

The pmf of $N \sim \text{Bin}(m, q)$ is given by

$$p_k = \mathbb{P}(N = k) = \binom{m}{k} q^k (1-q)^{n-k}$$

for any $k = 0, 1, \ldots, n$.

Remarks:

- $N \sim \text{Bin}(m, q)$ can be more practically interpreted as the number of claims made when there are m moments (in a certain year) at which a claim (of suffering loss) is possible, independently with the same claim probability q.
- This also explains the choice of notation q here. For life insurance, a claim is made when the insured dies (where the letter q is used in the notation of death/claim probability). So, to "match" with this case, we use q to denote the "success"/claim probability here.

3.6.2 Since $N \sim \text{Bin}(m, q)$ has the same distribution as a sum of n independent Bernoulli random variables with $\text{Ber}(q) \equiv \text{Bin}(1, q)$ distribution, and the pgf of a random variable with $\text{Ber}(q)$ distribution is $t^0(1-q) + t^1 q = 1 + q(t-1)$, it follows that the pgf of $N \sim \text{Bin}(m, q)$ is

$$P_N(t) = \boxed{[1 + q(t-1)]^m}$$

by proposition 3.1.d.

3.6.3 Based on the pgf, we know

- $P_N'(t) = mq[1 + q(t-1)]^{m-1}$.
- $P_N''(t) = m(m-1)q^2[1 + q(t-1)]^{m-2}$.

Thus, we can obtain the mean and variance of $N \sim \text{Bin}(m, q)$ as follows.

- $\mathbb{E}[N] = P_N'(1) = \boxed{mq}$.
- $\mathbb{E}[N(N-1)] = P_N''(1) = m(m-1)q^2 \implies \text{Var}(N) = \mathbb{E}\left[N^2\right] - (\mathbb{E}[N])^2 = m(m-1)q^2 + nq - (mq)^2 = \boxed{mq(1-q)}$.

4 The $(a, b, 0)$ and $(a, b, 1)$ Classes

4.0.1 In this section, we discuss two *classes* of probability distributions: $(a, b, 0)$ and $(a, b, 1)$ classes, which contain many useful distributions for modelling claim frequency N.

4.1 The $(a, b, 0)$ Class

4.1.1 Let N be a nonnegative discrete random variable, and let $p_k = \mathbb{P}(N = k)$ for any $k \in \mathbb{N}_0$. Then, (the distribution of) N is in the **$(a, b, 0)$ class** if there exist constants a and b such that

$$\frac{p_k}{p_{k-1}} = a + \frac{b}{k}, \quad \text{for any } k \geq 1 \text{ (in the support of } N\text{)}.$$

[Note: The relationship above starts with the pair p_0 and p_1, so the class is called $(a, b, 0)$.]

4.1.2 For a random variable N in the $(a, b, 0)$ class, once a and b are fixed, the probabilities $p_1, p_2, \ldots$ can all be deduced from p_0. Furthermore, we can deduce the value of p_0 by the fact that $p_0 + p_1 + \cdots = 1$. Thus, all the probabilities $p_0, p_1, \ldots$ are fixed after a and b are fixed.

Hence, the values of a and b can be used to uniquely characterize a distribution (provided that it is in the $(a, b, 0)$ class).

4.1.3 An important result regarding the $(a, b, 0)$ class is as follows.

Theorem 4.1.a. Poisson, negative binomial, and binomial distributions are the *only* (non-degenerate[4]) distributions in the $(a, b, 0)$ class.

Proof: Omitted. (See, e.g., Sundt and Jewell (1981).) □

4.1.4 Although here we do not show Poisson, negative binomial, and binomial distributions are the *only* distributions in the $(a, b, 0)$ class, we will show that they *are* (without the "only" part) distributions in the $(a, b, 0)$ class in the following.

4.1.5 For $N \sim \text{Poi}(\lambda)$, for any $k \in \mathbb{N}$,

$$\frac{p_k}{p_{k-1}} = \frac{e^{-\lambda}\lambda^k/k!}{e^{-\lambda}\lambda^{k-1}/(k-1)!} = 0 + \frac{\lambda}{k}.$$

Hence, $N \sim \text{Poi}(\lambda)$ is in the $(a, b, 0)$ class with $a = 0$ and $b = \lambda$.

4.1.6 For $N \sim \text{NB}(r, \beta)$, for any $k \in \mathbb{N}$,

$$\begin{aligned}
\frac{p_k}{p_{k-1}} &= \frac{\binom{r+k-1}{k}\left(\frac{1}{1+\beta}\right)\left(\frac{\beta}{1+\beta}\right)^k}{\binom{r+k-1-1}{k-1}\left(\frac{1}{1+\beta}\right)\left(\frac{\beta}{1+\beta}\right)^{k-1}} \\
&= \frac{\beta}{1+\beta}\frac{(r+k-1)(r+k-2)(r+k-3)\cdots r/k!}{(r+k-2)(r+k-3)\cdots r/(k-1)!} \\
&= \frac{\beta}{1+\beta} \cdot \underbrace{\frac{r+k-1}{k}}_{1+\frac{r-1}{k}} \\
&= \frac{\beta}{1+\beta} + \frac{(r-1)\frac{\beta}{1+\beta}}{k}.
\end{aligned}$$

Hence, $N \sim \text{NB}(r, \beta)$ is in the $(a, b, 0)$ class with $a = \dfrac{\beta}{1+\beta} > 0$ and $b = (r-1)\dfrac{\beta}{1+\beta}$.

[4] This means that a random variable following the distribution is not non-random.

4.1.7 For $N \sim \text{Bin}(m, q)$, for any $k = 1, 2, \ldots, n$,

$$\begin{aligned}\frac{p_k}{p_{k-1}} &= \frac{\binom{m}{k} q^k (1-q)^{n-k}}{\binom{m}{k-1} q^{k-1} (1-q)^{n-k+1}} \\ &= \frac{q}{1-q} \cdot \frac{m(m-1)\cdots(m-k+2)(m-k+1)/k!}{m(m-1)\cdots(m-k+2)/(k-1)!} \\ &= \frac{q}{1-q} \cdot \frac{m-k+1}{k} \\ &= -\frac{q}{1-q} + \frac{(m+1)\frac{q}{1-q}}{k}.\end{aligned}$$

Hence, $N \sim \text{Bin}(m, q)$ is in the $(a, b, 0)$ class with $a = -\frac{q}{1-q} < 0$ and $b = (m+1)\frac{q}{1-q}$.

4.1.8 By theorem 4.1.a, [4.1.5] to [4.1.7] include all possibilities for distributions in the $(a, b, 0)$ class. Notably, if a distribution in the $(a, b, 0)$ class has a *positive* (*negative*) a, it must be *negative binomial* (*binomial*) distributed. (Of course, if $a = 0$, it must be Poisson distributed.) [⚠ Warning: We do not have $a < 0 \implies$ negative binomial distributed!]

4.1.9 We can rewrite the equation in the definition of $(a, b, 0)$ class to

$$k \cdot \frac{p_k}{p_{k-1}} = ka + b \quad \text{for any } k \geq 1,$$

so $k \cdot \frac{p_k}{p_{k-1}}$ is linearly related to k.

[Note: Practically, this is helpful for having a quick check on whether the actual claim frequency in practice is in the $(a, b, 0)$ class as follows. First note the following approximated relationship:

$$k \cdot \frac{\widehat{p}_k}{\widehat{p}_{k-1}} \approx ka + b$$

where $\widehat{p}_k = \frac{\text{no. of policies with } k \text{ claims}}{\text{total no. of policies considered}}$ for any $k \in \mathbb{N}_0$.

Then, by plotting $k \cdot \frac{\widehat{p}_k}{\widehat{p}_{k-1}}$ against k (or other means), we can examine whether a linear relationship is plausible. If such linear relationship is deemed plausible, we can then use some line fitting technique to obtain estimated value of a and b.]

4.2 The $(a, b, 1)$ Class

4.2.1 One main issue of using distribution in the $(a, b, 0)$ class (Poisson/negative binomial/binomial) to model the number of claims N is that the probability $p_0 = \mathbb{P}(N = 0)$ is usually unreasonably low. In practice, especially when the insurance covers some "rare" loss, the probability p_0 is often quite large.

4.2.2 To deal with this issue, a simple way is to add a flexibility on the choice of p_0 on distribution in the $(a, b, 0)$ class → resulting in the $(a, b, 1)$ *class*.

4.2.3 Let N be a nonnegative discrete random variable with pmf $p_k = \mathbb{P}(N = k)$. Then, (the distribution of) N is in the $(a, b, 1)$ **class** if there exist constants a and b such that

$$\frac{p_k}{p_{k-1}} = a + \frac{b}{k} \quad \text{for any } k \geq 2,$$

while p_0 can take any value (as long as none of the probability axioms is violated).

[Note: The relationship above starts with the pair p_1 and p_2, so the class is called $(a, b, 1)$.]

4.2.4 Like the $(a, b, 0)$ class, all the probabilities $p_0, p_1, \ldots$ are fixed after a, b, and p_0 are fixed.

Hence, the values of a, b, and p_0 can be used to uniquely characterize a distribution (provided that it is in the $(a, b, 1)$ class).

4.3 The Zero-Modified and Zero-Truncated $(a, b, 1)$ Classes

4.3.1 Another way to make distribution in the $(a, b, 0)$ class better model the number of claims N is to apply *zero-modification* to the distribution, which yields a distribution in the *zero-modified* $(a, b, 1)$ *class*.

4.3.2 More specifically, the **zero-modification** process is as follows.

(a) Pick any distribution in the $(a, b, 0)$ class. Then, we know

$$\frac{p_k}{p_{k-1}} = a + \frac{b}{k} \quad \text{for any } k \geq 1.$$

(b) Modify p_0 to an arbitrary number $p_0^M \in (0, 1)$ (we exclude 0 and 1 to avoid making the distribution degenerate).

(c) Modify the probabilities (in the support) p_1, p_2, ... to $p_1^M = cp_1$, $p_2^M = cp_2$,... respectively, for some constant c.

Then, $p_0^M, p_1^M, \ldots$ form a distribution in the **zero-modified $(a, b, 1)$ class**.

[Note: For any $k \geq 2$, we have

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

Hence, a distribution in the zero-modified $(a, b, 1)$ class is also in the $(a, b, 1)$ class.]

4.3.3 The constant c in [4.3.2] is indeed uniquely determined, due to the constraint that $p_0^M + p_1^M + \cdots = 1$:

$$p_0^M + p_1^M + p_2^M + \cdots = 1 \implies p_0^M + c(p_1 + p_2 + \cdots) = 1 \implies p_0^M + c(1 - p_0) = 1 \implies c = \boxed{\frac{1 - p_0^M}{1 - p_0}}.$$

4.3.4 Let N be a random variable in the $(a, b, 1)$ class, and let N^M be the random variable in the zero-modified $(a, b, 1)$ class obtained by applying zero-modification on the distribution of N. Then, for any $k \geq 1$, the kth moment of N^M is

$$\mathbb{E}\left[\left(N^M\right)^k\right] = \sum_{j=1}^{\infty} j^k p_j^M = \sum_{j=1}^{\infty} j^k c p_j = c \sum_{j=1}^{\infty} j^k p_j = \boxed{c\mathbb{E}\left[N^k\right]}$$

where $c = \dfrac{1 - p_0^M}{1 - p_0} = \dfrac{1 - \mathbb{P}(M = 0)}{1 - \mathbb{P}(N = 0)}$.

4.3.5 Consider the same setting as [4.3.4]. Then, the pgf of N^M is

$$P_{N^M}(t) = p_0^M + tp_1^M + t^2 p_2^M = p_0^M + c(tp_1 + t^2 p_2 + \cdots) = \boxed{p_0^M + c(P_N(t) - p_0)}$$

where $P_N(t)$ is the pgf of N.

We can further write

$$p_0^M + c(P_N(t) - p_0) = \underbrace{p_0^M - cp_0}_{1-c} + cP_N(t) = \boxed{1 - c + cP_N(t)}$$

which only involves c and $P_N(t)$.

4.3.6 If we modify p_0 to $p_0^M = 0$ in the zero modification process, we call the process as **zero-truncation**, and the resulting distribution is in the **zero-truncated $(a, b, 1)$ class**. Furthermore, we usually denote p_k^M by p_k^T instead for any $k \geq 0$.

4.3.7 Since zero-truncated $(a, b, 1)$ class is essentially a special case of zero-modified $(a, b, 1)$ class, previous formulas in [4.3.3] to [4.3.5] also apply, by setting $p_0^M = 0$. Particularly:

- $c = \dfrac{1}{1 - p_0}$ (where $p_k^T = cp_k$ for any $k \geq 1$).
- The pgf of N^T (obtained by zero-truncating N) is

$$P_{N^T}(t) = (1 - c) + cP_N(t) = \left(1 - \frac{1}{1 - p_0}\right) + \frac{1}{1 - p_0} P_N(t),$$

which only involves p_0 and $P_N(t)$.

4.4 Extended-Truncated Negative Binomial Distribution

4.4.1 In this section we consider a special way to modify a *negative binomial distribution* which involves *extension* and *truncation*.

4.4.2 To motivate this way of modification, consider a random variable $N \sim \mathrm{NB}(r, \beta)$. Its pmf is given by

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

for any $k \in \mathbb{N}_0$. Recall that the constraints on the parameters are $r > 0$ and $\beta > 0$.

Now, suppose we (improperly) set $r \in (-1, 0)$ (while $\beta > 0$ still). A consequence is that

$$p_0 = (1 + \beta)^{-r} > 1 \quad \text{and} \quad p_k < 0 \text{ for any } k \in \mathbb{N},$$

violating the probability axioms.

4.4.3 Nevertheless, in this case, we can observe that $p_0 + p_1 + \cdots = 1$ still holds, and the recursive relation for the $(a, b, 0)$ class is still satisfied:

$$p_k = \left(a + \frac{b}{k}\right) p_{k-1} \quad \text{for any } k \in \mathbb{N},$$

where $a = \dfrac{\beta}{1 + \beta}$ and $b = (r - 1)\dfrac{\beta}{1 + \beta}$ (same as [4.1.6]).

So, allowing $r \in (-1, 0)$ is indeed "mostly" appropriate with just "minor" issues. Hence, we are interested in finding a modification on the $\mathrm{NB}(r, \beta)$ distribution to permit $r \in (-1, 0)$ without violating the probability axioms.

4.4.4 This modification contains two elements: *zero-truncation* ("*truncated*") and *extending* the possible range of r ("*extended*"). We perform zero-truncation on $N \sim \mathrm{NB}(r, \beta)$ (where $r \in (-1, 0)$ and $\beta > 0$), which gives

$$p_0^T = 0 \quad \text{and} \quad p_k^T = cp_k \quad \text{for any } k \in \mathbb{N}$$

where $c = \dfrac{1}{1 - p_0} < 0$ (as $p_0 > 1$).

[Note: By construction of zero-truncation, we always have $p_0^T + p_1^T + p_2^T + \cdots + = 1$. Furthermore, p_k^T is nonnegative for any $k \in \mathbb{N}_0$ (since $p_k < 0$ for any $k \in \mathbb{N}$ and $c < 0$). Thus, the probability axioms are not violated.]

Then, the probabilities $p_0^T, p_1^T, \ldots$ form the **extended-truncated negative binomial distribution** with parameters $r \in (-1, 0)$ and $\beta > 0$ (denoted by $\mathrm{ETNB}(r, \beta)$).

4.4.5 Since ETNB (r, β) is in the zero-truncated $(a, b, 1)$ class, previous results for this class apply. Particularly, we have the following recursive formula for the $(a, b, 1)$ class:

$$p_k^T = \left(a + \frac{b}{k}\right) p_{k-1}^T \quad \text{for any } k \geq 2$$

where $a = \dfrac{\beta}{1+\beta}$ and $b = (r-1)\dfrac{\beta}{1+\beta}$ (same as [4.1.6]).

4.4.6 As a practical note, to compute probabilities for ETNB (r, β) where $r \in (-1, 0)$ and $\beta > 0$, we carry out the following steps:

(a) Treat it as if it were an ordinary NB (r, β) distribution and use the pmf formula to compute

$$p_k = \binom{k+r-1}{k} \left(\frac{\beta}{1+\beta}\right)^k \left(\frac{1}{1+\beta}\right)^r$$

for all k needed (possibly depending on the method chosen in the second step below). Particularly:

- $p_0 = \left(\dfrac{1}{1+\beta}\right)^r > 1.$
- $p_1 = r\left(\dfrac{1}{1+\beta}\right)^r \left(\dfrac{\beta}{1+\beta}\right) < 0.$

(b) The desired ETNB (r, β) probabilities can then be obtained by one of the following methods.

i. Directly use the following:

$$p_0^T = 0 \quad \text{and} \quad p_k^T = c p_k \quad \text{for any } k \in \mathbb{N}$$

where $c = \dfrac{1}{1-p_0}$.

ii. First compute $p_1^T = c p_1$, and then use the following recursive formula:

$$p_k^T = \left(a + \frac{b}{k}\right) p_{k-1}^T \quad \text{for any } k = 2, 3, \ldots$$

where $a = \dfrac{\beta}{1+\beta}$ and $b = (r-1)\dfrac{\beta}{1+\beta}$ (same as [4.1.6]).

5 Compound Frequency Models

5.1 Compound Frequency Distributions

5.1.1 Apart from mixing discussed in section 2, another method to create a new discrete distribution is to *compound* ($\neq$ mix!) two discrete distributions.

5.1.2 Let P_N and P_M be the pgf of the discrete random variables N and M respectively. Then, define a new function P_S by

$$P_S(t) = P_N(P_M(t)).$$

Consider a random variable S with pgf P_S. The distribution of S is called a **compound distribution** (and S is called **compound random variable**), where:

- distribution of N is called **primary distribution** ($\leftrightarrow$ outer pgf);
- distribution of M is called **secondary distribution** ($\leftrightarrow$ inner pgf).

[Mnemonic: Output from the pgf of *secondary* distribution is fed into the pgf of *primary* distribution:

$$t \xrightarrow[\text{secondary}]{P_M} P_M(t) \xrightarrow[\text{primary}]{P_N} P_S(t)$$

]

5.1.3 A practical interpretation of S is as follows. Let:

- N: number of accidents arising in a portfolio of risks, with pgf P_N
- $M_1, M_2, \ldots$: number of claims from each accident
- S: total number of claims

Then, $S = M_1 + \cdots + M_N$.

Remarks:

- The number of summands is N, a random variable.
- The equation $S = M_1 + \cdots + M_N$ means that given $N = n$, we have $S = M_1 + \cdots + M_n$.
- When $N = 0$, we set $S = 0$.

Theorem 5.1.a. Suppose that $M_1, M_2, \ldots$ are i.i.d. (with the same distribution as random variable M) and are independent of N. The pgf of S is

$$P_S(t) = P_N(P_M(t)).$$

[Note: This suggests that the definition of compound distribution using pgf (in [5.1.2]) and using sum of N i.i.d. random variables (here) are equivalent (in the sense that both result in the same distribution).]

Proof: Note that

$$
\begin{aligned}
P_S(t) &= \sum_{k=0}^{\infty} t^k \mathbb{P}(S=k) \\
&= \sum_{k=0}^{\infty} t^k \sum_{n=0}^{\infty} \mathbb{P}(S=k|N=n)\,\mathbb{P}(N=n) \\
&= \sum_{n=0}^{\infty} \mathbb{P}(N=n) \sum_{k=0}^{\infty} t^k \mathbb{P}(M_1+\cdots+M_n=k|N=n) \\
&= \sum_{n=0}^{\infty} \mathbb{P}(N=n) \sum_{k=0}^{\infty} t^k \mathbb{P}(M_1+\cdots+M_n=k) && \text{(by independence)} \\
&= \sum_{n=0}^{\infty} \mathbb{P}(N=n)\, P_{M_1+\cdots+M_n}(t) \\
&= \sum_{n=0}^{\infty} \mathbb{P}(N=n)\, [P_M(t)]^n && \text{(proposition 3.1.d, i.i.d. assumption)} \\
&= \mathbb{E}\left[[P_M(t)]^N\right] \\
&= P_N(P_M(t)).
\end{aligned}
$$

[Note: $P_{M_1+\cdots+M_n}$ and P_M are the pgf's of $M_1+\cdots+M_n$ and M respectively.] □

5.1.4 Consider the setting in [5.1.2]. The mean and variance of S are as follows.

- $\mathbb{E}[S] = \mathbb{E}[N]\,\mathbb{E}[M]$
- $\mathrm{Var}(S) = \mathbb{E}[N]\,\mathrm{Var}(M) + (\mathbb{E}[M])^2\,\mathrm{Var}(N)$

Proof: We shall use the equivalent definition suggested by theorem 5.1.a to prove these formulas. Firstly, note that

$$\mathbb{E}[S|N=n] = \mathbb{E}[M_1+\cdots+M_N|N=n] = \mathbb{E}[M_1+\cdots+M_n|N=n] = \mathbb{E}[M_1+\cdots+M_n] = n\mathbb{E}[M].$$

Hence, $\mathbb{E}[S|N] = N\mathbb{E}[M]$. Thus, by law of total expectation,

$$\mathbb{E}[S] = \mathbb{E}[\mathbb{E}[S|N]] = \mathbb{E}[N\mathbb{E}[M]] = \mathbb{E}[M]\mathbb{E}[N].$$

Next, for the variance formula, first consider:

$$
\begin{aligned}
\mathrm{Var}(S|N=n) &= \mathrm{Var}(M_1+\cdots+M_N|N=n) \\
&= \mathrm{Var}(M_1+\cdots+M_n|N=n) \\
&= \mathrm{Var}(M_1+\cdots+M_n) && \text{(by independence)} \\
&= n\,\mathrm{Var}(M) && \text{(i.i.d. assumption).}
\end{aligned}
$$

Thus, $\mathrm{Var}(S|N) = N\,\mathrm{Var}(M)$, and by law of total variance,

$$\mathrm{Var}(S) = \mathbb{E}[\mathrm{Var}(S|N)] + \mathrm{Var}(\mathbb{E}[S|N]) = \mathbb{E}[N\mathrm{Var}(M)] + \mathrm{Var}(N\mathbb{E}[M]) = \mathrm{Var}(M)\mathbb{E}[N] + (\mathbb{E}[M])^2\,\mathrm{Var}(N).$$

□

5.2 Panjer's Recursion

5.2.1 Consider again the setting in [5.1.2]. Let:

- $g_k = \mathbb{P}(S=k)$;

- $p_k = \mathbb{P}(N = k)$;
- $f_k = \mathbb{P}(M = k)$.

Suppose that p_k and f_k are known for any $k \geq 0$. Then, we are interested in finding the distribution of S, i.e., the value of g_k for any $k \geq 0$. Unfortunately, in general there is no simple way to do this.

5.2.2 However, in the special case where N is in the $(a, b, 0)$ class, we have the following result that allows us to compute g_k recursively.

Theorem 5.2.a (Panjer's recursion ($(a, b, 0)$ class)). In this special case, for any $k \geq 1$ (in the support of S),

$$g_k = \frac{1}{1 - af_0} \sum_{j=1}^{k} \left(a + \frac{bj}{k}\right) f_j g_{k-j}.$$

Proof: Omitted (See, e.g., proof of theorem 7.1 in Klugman et al. (2019)). □

Remarks:

- The a and b are the constants a and b in the $(a, b, 0)$ class characterization of S.
- To use the Panjer's recursion, we need the initial value g_0. It can be obtained by

$$g_0 = P_S(0) = P_N(P_M(0)) = P_N(f_0).$$

- Some special cases:
  - $(k = 1)$ $g_1 = \dfrac{1}{1 - af_0}\left(a + \dfrac{b \cdot 1}{1}\right) f_1 g_0$
  - $(k = 2)$ $g_2 = \dfrac{1}{1 - af_0}\left[\left(a + \dfrac{b \cdot 1}{2}\right) f_1 g_1 + \left(a + \dfrac{b \cdot 2}{2} f_2 g_0\right)\right]$
  - $(k = 3)$ $g_3 = \dfrac{1}{1 - af_0}\left[\left(a + \dfrac{b \cdot 1}{3}\right) f_1 g_2 + \left(a + \dfrac{b \cdot 2}{3} f_2 g_1\right) + \left(a + \dfrac{b \cdot 3}{3} f_3 g_0\right)\right]$

5.2.3 When N is in the $(a, b, 1)$ class instead, we can slightly modify the Panjer's recursion to obtain the appropriate recursive formula as follows.

Theorem 5.2.b (Panjer's recursion ($(a, b, 1)$ class)). In this special case, for any $k \geq 1$,

$$g_k = \frac{1}{1 - af_0}\left\{[p_1 - (a + b)p_0] f_k + \sum_{j=1}^{k} \left(a + \frac{bj}{k}\right) f_j g_{k-j}\right\}.$$

Proof: Similar to the proof for theorem 5.2.a. □

5.3 Compound Poisson Frequency Distributions

5.3.1 If S is a compound random variable with the *primary* distribution being $\mathrm{Poi}(\lambda)$, then S follows a **compound Poisson distribution** (and S is called a **compound Poisson random variable**).

5.3.2 Recall from [3.2.2] that the pgf of $N \sim \mathrm{Poi}(\lambda)$ is

$$P_N(s) = \exp\{\lambda(s - 1)\}.$$

Thus, the pgf of a compound Poisson random variable S is

$$P_S(t) = P_N(P_M(t)) = \boxed{\exp\{\lambda(P_M(t) - 1)\}}$$

where P_M is the pgf for the secondary distribution.

5.3.3 Conversely, if S is a discrete random variable whose pgf can be expressed as

$$P_S(t) = \exp\{\lambda(Q(t) - 1)\}$$

for some function Q which serves as a pgf of a discrete random variable, then S is a compound Poisson random variable with primary distribution being $\mathrm{Poi}\,(\lambda)$ and secondary distribution identified by the pgf Q.

[Note: This holds since same pgf implies same distribution, by corollary 3.1.b.]

5.3.4 Recall the convolution result for Poisson distribution (theorem 3.2.a). It turns out that analogous result holds also for *compound* Poisson distribution:

Theorem 5.3.a. Let $S_1, \dots, S_k$ be k independent compound Poisson random variables with Poisson parameters $\lambda_1, \dots, \lambda_k$ respectively. Suppose that the pmf of the secondary distribution for S_i is given by $q_i(n)$, for any $i = 1, \dots, k$. Then, the sum $S = S_1 + \dots + S_k$ is also a compound Poisson random variable with Poisson parameter $\lambda = \lambda_1 + \dots + \lambda_k$, and the pmf of the secondary distribution is given by

$$q(n) = \frac{\lambda_1}{\lambda} q_1(n) + \dots + \frac{\lambda_k}{\lambda} q_k(n).$$

[Note: $q(n)$ is a legitimate pmf since $q(n) \geq 0$ (as $q_1(n), \dots, q_k(n) \geq 0$) for any $n \in \mathbb{N}_0$ and

$$\sum_{n=0}^{\infty} q(n) = \frac{\lambda_1}{\lambda} \sum_{n=0}^{\infty} q_1(n) + \dots + \frac{\lambda_k}{\lambda} \sum_{n=0}^{\infty} q_k(n) = \frac{\lambda_1 + \dots + \lambda_k}{\lambda} = 1.$$

]

Proof: Let Q_i be the pgf of the secondary distribution for S_i, i.e.,

$$Q_i(t) = \sum_{n=0}^{\infty} t^n q_i(n).$$

Then, for any $i = 1, \dots, k$, the pgf of S_i can be expressed as

$$P_{S_i}(t) = \exp\{\lambda_i(Q_i(t) - 1)\}.$$

Also, the pgf of the secondary distribution for S is

$$Q(t) = \sum_{n=0}^{\infty} t^n q(n) = \frac{\lambda_1}{\lambda} \sum_{n=0}^{\infty} t^n q_1(n) + \dots + \frac{\lambda_k}{\lambda} \sum_{n=0}^{\infty} t^n q_k(n) = \frac{\lambda_1}{\lambda} Q_1(t) + \dots + \frac{\lambda_k}{\lambda} Q_k(t).$$

Since $S_1, \dots, S_k$ are independent, by proposition 3.1.d, the pgf of S is

$$P_S(t) = \prod_{i=1}^{k} P_{S_i}(t) = \prod_{i=1}^{k} \exp\{\lambda_i(Q_i(t) - 1)\} = \exp\left\{\sum_{i=1}^{k} \lambda_i Q_i(t) - \sum_{i=1}^{k} \lambda_i\right\} = \exp\{\lambda(Q(t) - 1)\}.$$

Thus, the result follows by [5.3.3]. □

6 Coverage Modifications

6.0.1 In this section, we will discuss different kinds of *coverage modifications*: modifications on the "coverage"/term for an insurance on some loss. Mathematically speaking, with coverage modifications, the amount of actual payment made by the insurer is obtained by modifying the loss suffered X.

6.0.2 Some of the coverage modifications here have been briefly discussed in section 1, namely *deductibles* and *policy limit*. But here we will discuss more varieties of coverage modifications in more details.

6.1 Ordinary Deductibles

6.1.1 Consider an insurance policy with a *per-loss* (ordinary) deductible d. Then, for every loss claimed,

- if the loss $X \le d$, then there is no payment;
- if the loss $X > d$, then the payment amount is $X - d$.

The first d dollars of the loss is borne by the policyholder → avoid *moral hazard*.

[Note: If there is no such deductible and the full loss amount is covered by the insurance, may be incentivized to take "too much" risk as is not responsible for bearing the loss. This issue is known as **moral hazard**.]

6.1.2 Consider a loss random variable X. With the per-loss deductible d, we can modify the loss X in the following ways:

- loss X → the **per-payment variable** (excess loss variable):

$$Y^P = (X - d | X > d) = \begin{cases} \text{undefined} & \text{if } X \le d; \\ X - d & \text{if } X > d. \end{cases}$$

  [Note: The *per-payment* variable is the payment amount (or cost) *per payment* (which is triggered when $X > d$).]

- loss X → the **per-loss variable** (stop loss variable):

$$Y^L = (X - d)_+.$$

  [Note: The *per-loss* variable is the payment amount *per loss* (which is triggered when $X > d$).]

6.1.3 We can also write the per-payment variable as follows:

$$Y^P = \begin{cases} \text{undefined} & \text{if } Y^L \le 0; \\ Y^L & \text{if } Y^L > 0 \end{cases} = \boxed{(Y^L | Y^L > 0)}.$$

This means that the following three distributions are equal:

(a) distribution of Y^P

(b) conditional distribution of $X - d$ given $X > d$

(c) conditional distribution of Y^L given $Y^L > 0$

6.1.4 Some distributional quantities of Y^P are as follows.

- cdf:

$$F_{Y^P}(y) = \mathbb{P}\left(Y^P \le y\right) = \mathbb{P}(X - d \le y | X > d) = \frac{\mathbb{P}(d < X \le y + d)}{\mathbb{P}(X > d)} = \boxed{\frac{F_X(y+d) - F_X(d)}{1 - F_X(d)}}, \quad y > 0.$$

- survival function:

$$S_{Y^P}(y) = 1 - F_{Y^P}(y) = 1 - \frac{F_X(y+d) - F_X(d)}{\underbrace{1 - F_X(d)}_{S_X(d)}} = \frac{\overbrace{S_X(d) + F_X(d)}^{1} - F_X(y+d)}{S_X(d)} = \boxed{\frac{S_X(y+d)}{S_X(d)}}, \quad y > 0.$$

- pdf:

$$f_{Y^P}(y) = \frac{\mathrm{d}}{\mathrm{d}y} F_{Y^P}(y) = \frac{F_X'(y+d) \cdot \overbrace{\frac{\mathrm{d}}{\mathrm{d}y}(y+d)}^{1}}{S_X(d)} = \boxed{\frac{f_X(y+d)}{S_X(d)}}, \quad y > 0.$$

[Note: When X and Y^P are discrete, the pmf of Y^P (f_{Y^P}) takes the same form, but derived differently:

$$f_{Y^P}(y) = F_{Y^P}(y+1) - F_{Y^P}(y) = \frac{F_X(y+1+d) - F_X(y+d)}{S_X(d)} = \frac{f_X(y+d)}{S_X(d)}, \quad y = 1, 2, \ldots,$$

where f_X is the pmf of X, and $f_{Y^P}(y) = 0$ elsewhere.]

- hazard rate function:

$$h_{Y^P}(y) = \frac{f_{Y^P}(y)}{S_{Y^P}(y)} = \frac{f_X(y+d)}{S_X(y+d)} = \boxed{h_X(y+d)}, \quad y > 0.$$

6.1.5 Some distributional quantities of Y^L are as follows.

- survival function:

$$S_{Y^L}(y) = \boxed{S_X(y+d)}, \quad y \geq 0.$$

[Note: Recall that $(X-d)_+ > y \iff X > y+d$ for any $y \geq 0$ (see the proof of proposition 1.2.a).]

- cdf:

$$F_{Y^L}(y) = 1 - S_{Y^L}(y) = \boxed{F_X(y+d)}, \quad y \geq 0.$$

- probability function (mixed): For any $y > 0$, (continuous; pdf)

$$f_{Y^L}(y) = \frac{\mathrm{d}}{\mathrm{d}y} F_{Y^P}(y) = \boxed{f_X(y+d)}.$$

When $y = 0$, (discrete; pmf)

$$f_{Y^L}(y) = \mathbb{P}\left(Y^L = 0\right) = \mathbb{P}(X \leq d) = \boxed{F_X(d)}.$$

[Note: When X and Y^P are discrete, the probability function f_{Y^L} is pmf, and takes the same form:

$$f_{Y^L}(y) = F_{Y^L}(y+1) - F_{Y^L}(y) = F_X(y+1+d) - F_X(y+d) = f_X(y+d), \quad y = 1, 2, \ldots,$$

where f_X is the pmf of X (and $f_{Y^L}(0) = F_X(d)$ still).]

- hazard rate function:

$$h_{Y^L}(y) = \frac{f_{Y^L}(y)}{S_{Y^L}(y)} = \frac{f_X(y+d)}{S_X(y+d)} = \boxed{h_X(y+d)}, \quad y > 0.$$

[Note: This is the same as $h_{Y^P}(y)$.]

6.1.6 Here we recall formulas for computing the means of Y^L and Y^P, which have been discussed in section 1 (in the language of stop loss and excess loss variables):

- $$\mathbb{E}\left[Y^L\right] \overset{[1.2.3]}{=} \boxed{\int_d^\infty (x-d) f_X(x)\,\mathrm{d}x} \overset{\text{prop. 1.2.a}}{=} \boxed{\int_d^\infty S_X(x)\,\mathrm{d}x} \overset{[1.4.4]}{=} \boxed{\mathbb{E}[X] - \mathbb{E}[X \wedge d]}.$$

- $$\mathbb{E}\left[Y^P\right] \overset{[1.3.5]}{=} \boxed{\frac{\mathbb{E}\left[Y^L\right]}{\mathbb{P}(X > d)}} = \boxed{\frac{\mathbb{E}\left[Y^L\right]}{\mathbb{P}(Y^L > 0)}}.$$

6.2 Franchise Deductibles

6.2.1 A **franchise deductible** is a modified version of ordinary deductible where the deductible amount is added on top of the payment amount \$ *when the payment amount \$ is positive.*

6.2.2 Hence, with the per-loss *franchise* deductible d, we can modify the loss X in the following ways:

- loss X → the **per-payment variable** (excess loss variable) *in this context*:

$$Y^P = (X|X > d) = \begin{cases} \text{undefined} & \text{if } X \leq d; \\ X & \text{if } X > d. \end{cases}$$

[Note: The *per-payment* variable is the payment amount *per payment* (which is triggered when $X > d$).]

- loss X → the **per-loss variable** (stop loss variable) *in this context*:

$$Y^L = \begin{cases} 0 & \text{if } X \leq d; \\ X & \text{if } X > d. \end{cases}$$

[Note: The *per-loss* variable is the payment amount *per loss* (which is triggered when $X > d$).]

6.2.3 Likewise, we can also write the per-payment variable as follows:

$$Y^P = \begin{cases} \text{undefined} & \text{if } Y^L \leq 0; \\ Y^L & \text{if } Y^L > 0 \end{cases} = \boxed{(Y^L|Y^L > 0)}.$$

This means that the following three distributions are equal:

(a) distribution of Y^P

(b) conditional distribution of X given $X > d$

(c) conditional distribution of Y^L given $Y^L > 0$

6.2.4 Some distributional quantities of Y^P are as follows.

- cdf:

$$F_{Y^P}(y) = \mathbb{P}\left(Y^P \leq y\right) = \mathbb{P}\left(X \leq y|X > d\right) = \frac{\mathbb{P}\left(d < X \leq y\right)}{\mathbb{P}\left(X > d\right)} = \boxed{\frac{F_X(y) - F_X(d)}{1 - F_X(d)}}, \quad y > d.$$

[Note: We have $F_{Y^P}(y) = 0$ for any $y \leq d$.]

- survival function:

$$S_{Y^P}(y) = \boxed{\frac{S_X(y)}{S_X(d)}}, \quad y > d.$$

[Note: We have $S_{Y^P}(y) = 1$ for any $y \leq d$.]

- pdf:

$$f_{Y^P}(y) = \frac{\mathrm{d}}{\mathrm{d}y} F_{Y^P}(y) = \frac{F_X'(y)}{S_X(d)} = \boxed{\frac{f_X(y)}{S_X(d)}}, \quad y > d.$$

Remarks:

– We have $f_{Y^P}(y) = 0$ for any $y \leq d$.

– When X and Y^P are discrete, the pmf of Y^P (f_{Y^P}) takes the same form, but derived differently:

$$f_{Y^P}(y) = F_{Y^P}(y+1) - F_{Y^P}(y) = \frac{F_X(y+1) - F_X(y)}{S_X(d)} = \frac{f_X(y)}{S_X(d)}, \quad y = d+1, d+2, \ldots,$$

where f_X is the pmf of X, and $f_{Y^P}(y) = 0$ elsewhere.

- hazard rate function:

$$h_{Y^P}(y) = \frac{f_{Y^P}(y)}{S_{Y^P}(y)} = \frac{f_X(y)}{S_X(y)} = \boxed{h_X(y)}, \quad y > d.$$

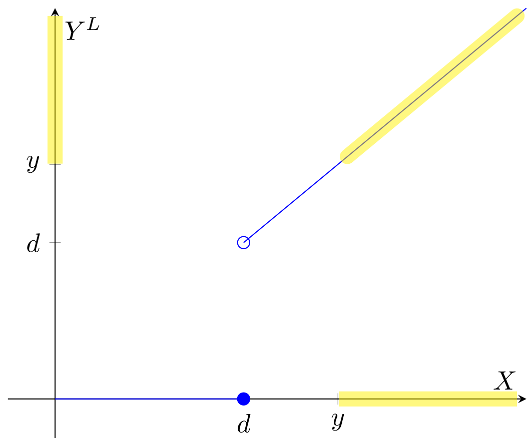
[Note: We have $h_{Y^P}(y) = 0$ for any $y \leq d$.]

6.2.5 Some distributional quantities of Y^L are as follows.

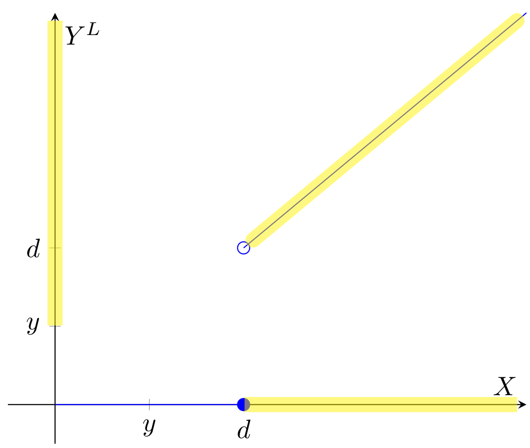
- survival function:

$$S_{Y^L}(y) = \boxed{\begin{cases} S_X(y) & \text{if } y > d; \\ S_X(d) & \text{if } 0 \leq y \leq d \end{cases}}.$$

[Note: We have $Y^L > y \iff X > y$ for any $y > d$.



Also, we have $Y^L > y \iff X > d$ for any $0 \leq y \leq d$.



]

- cdf:

$$F_{Y^L}(y) = 1 - S_{Y^L}(y) = \boxed{F_X(y)}, \quad y \geq 0.$$

- probability function (mixed): For any $y > d$, (continuous; pdf)

$$f_{Y^L}(y) = \frac{\mathrm{d}}{\mathrm{d}y} F_{Y^P}(y) = \boxed{f_X(y)},$$

and $f_{Y^L}(y) = 0$ when $0 < y \leq d$ or $y < 0$.

When $y = 0$, (discrete; pmf)

$$f_{Y^L}(y) = \mathbb{P}\left(Y^L = 0\right) = \mathbb{P}(X \le d) = \boxed{F_X(d)}.$$

[Note: When X and Y^P are discrete, the probability function f_{Y^L} is pmf, and takes the same form:

$$f_{Y^L}(y) = F_{Y^L}(y+1) - F_{Y^L}(y) = F_X(y+1) - F_X(y) = f_X(y), \quad y = d, d+1, \ldots,$$

where f_X is the pmf of X (and $f_{Y^L}(0) = F_X(d)$ still).]

- hazard rate function:

$$h_{Y^L}(y) = \frac{f_{Y^L}(y)}{S_{Y^L}(y)} = \frac{f_X(y)}{S_X(y)} = \boxed{h_X(y)}, \quad y > d.$$

[Note: This is the same as $h_{Y^P}(y)$.]

6.2.6 To compute the means of Y^L and Y^P in the case of franchise deductible, consider first the following. Let

$$W = \begin{cases} 0 & \text{if } X \le d; \\ -d & \text{if } X > d. \end{cases}$$

Then,

$$Y^L + W = \begin{cases} 0 & \text{if } X \le d; \\ X - d & \text{if } X > d \end{cases} = (X-d)_+.$$

From this relationship, we can readily derive the following formula for $\mathbb{E}\left[Y^L\right]$:

$$\mathbb{E}\left[Y^L\right] = \mathbb{E}\left[(X-d)_+\right] - \underbrace{\mathbb{E}[W]}_{-d\cdot\mathbb{P}(X>d)} = \boxed{\mathbb{E}\left[(X-d)_+\right] + d \cdot \mathbb{P}(X > d)}.$$

For the per-payment variable Y^P, the formula takes the same form as the case for ordinary deductible:

$$\mathbb{E}\left[Y^P\right] = \boxed{\frac{\mathbb{E}\left[Y^L\right]}{\mathbb{P}\left(Y^L > 0\right)}} = \boxed{\frac{\mathbb{E}\left[Y^L\right]}{\mathbb{P}(X > d)}}$$

(but of course the meaning of Y^L here is different from that for ordinary deductible).

6.3 Loss Elimination Ratio

6.3.1 The loss elimination ratio quantifies the effect of an *ordinary* deductible in lowering the expected payment made by the insurer per loss (how much (expected) loss *for the insurer* (taken out of insurer's pocket) is "eliminated").

6.3.2 More precisely, the **loss elimination ratio** is the ratio of the decrease in the expected payment made by the insurer per loss with an ordinary deductible to the expected payment without the deductible, i.e.,

$$\text{loss elimination ratio} = \frac{\mathbb{E}[X] - \mathbb{E}\left[(X-d)_+\right]}{\mathbb{E}[X]} = \frac{\mathbb{E}[X \wedge d]}{\mathbb{E}[X]}.$$

6.4 Inflation

6.4.1 In practice, there is often a *delay* between the time at which the loss X is triggered (occurrence of accident) and the time at which the payment is made by the insurer, since it takes time for to "process" a claim .

6.4.2 In case the *inflation rate* is very high, such delay can cause a substantial drop in the *real worth* of the payment received by the policyholder. To protect against this inflation risk, we can modify the terms of the insurance to incorporate also the inflation element.

6.4.3 More precisely, let X be the loss before accounting for inflation, and suppose that the inflation rate (for the whole delay period) is r. If there were no inflation, the insurer needs to make a payment of X (loss *before* inflation), assuming no deductible. However, after accounting for inflation, to preserve the real worth, the payment made would be adjusted to $X' = X(1+r)$ (loss *after* inflation).

6.4.4 For an insurance with (ordinary) deductible d which incorporates inflation, the expected payment made by per loss is

$$\mathbb{E}\left[Y^L\right] = \mathbb{E}\left[(X'-d)_+\right] = \mathbb{E}\left[((1+r)X-d)_+\right] = \boxed{(1+r)\mathbb{E}\left[\left(X-\frac{d}{1+r}\right)_+\right]}.$$

6.4.5 For an insurance with (ordinary) deductible d which incorporates inflation, the expected payment made by per payment is

$$\mathbb{E}\left[Y^P\right] = \frac{\mathbb{E}\left[Y^L\right]}{\mathbb{P}\left(X'>d\right)} = \boxed{\frac{\mathbb{E}\left[Y^L\right]}{\mathbb{P}\left(X>\frac{d}{1+r}\right)}}.$$

6.5 Policy Limits

6.5.1 Consider an insurance policy with a policy limit u. Then, recall that:

- If the loss $X \le u$, then the insurer pays the full amount u to the policyholder.
- if the loss $X > u$, then only pays u dollars to.

The policy limit u is the maximum amount of payment made by.

6.5.2 With the policy limit u, the loss X is modified to

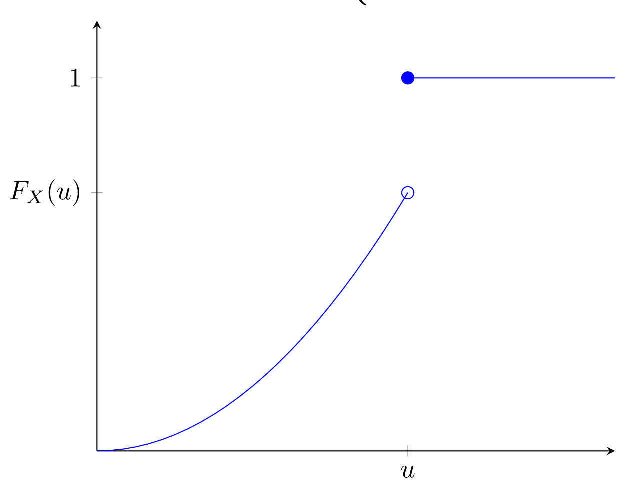
$$Y = X \wedge u,$$

which is the payment made by.

6.5.3 Some distributional quantities of $Y = X \wedge u$ are as follows.

- cdf:

$$F_Y(y) = \mathbb{P}\left(X \wedge u \le y\right) = \begin{cases} F_X(y) & \text{if } y < u; \\ 1 & \text{if } y \ge u. \end{cases}$$



- probability function (mixed):

$$f_Y(y) = \begin{cases} f_X(y) & \text{if } y < u \text{ (continuous; pdf);} \\ 1 - F_X(u) & \text{if } y = u \text{ (discrete; pmf).} \end{cases}$$

6.5.4 Now we consider an insurance with policy limit u which incorporates inflation (as in the setting in [6.4.3]). Then, the expected payment made by per loss is

$$\mathbb{E}\left[X' \wedge u\right] = \mathbb{E}\left[\{(1+r)X\} \wedge \left\{(1+r) \cdot \frac{u}{1+r}\right\}\right] = \boxed{(1+r)\mathbb{E}\left[X \wedge \frac{u}{1+r}\right]}.$$

6.6 Coinsurance

6.6.1 The final coverage modification introduced here is the *coinsurance*. The idea is that both parties (policyholder and insurer) contribute to the insurance coverage together (hence "co").

6.6.2 For an insurance policy with **coinsurance** element added, the insurer only pays a certain fixed proportion α of the original payment amount, where $\alpha \in [0, 1]$ (→ the minimum contribution is to have no contribution for both parties → no "negative" contribution is allowed![5]).

Hence, assuming no other coverage modifications, the payment made by is $X' = \alpha X$.

6.7 General Insurance

6.7.1 Here, a **general insurance** refers generally to any insurance with possibly multiple kinds of coverage modifications.

[Note: Sometimes the term "general insurance" is used to mean "non-life insurance", but this is not the case here.]

6.7.2 Here we consider a general insurance with the following coverage modifications:

(a) inflation (rate: r)
(b) policy limit u
(c) (ordinary) deductible d
(d) coinsurance (proportion: α)

Remarks:

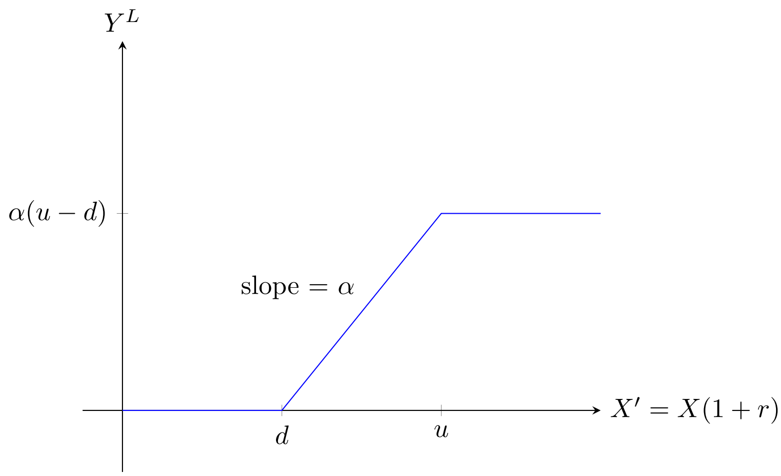
- We assume the modifications are applied *in the order above*:

$$X \overset{\text{inflation}}{\to} X(1+r) \overset{\text{policy limit}}{\to} X(1+r) \wedge u \overset{\text{deductible}}{\to} [X(1+r) \wedge u - d]_+ \overset{\text{coinsurance}}{\to} \alpha[X(1+r) \wedge u - d]_+.$$

- Also, we assume that $u > d$.

6.7.3 We can graphically show the final payment amount per loss $Y^L = \alpha[X(1+r) \wedge u - d]_+$ as follows:

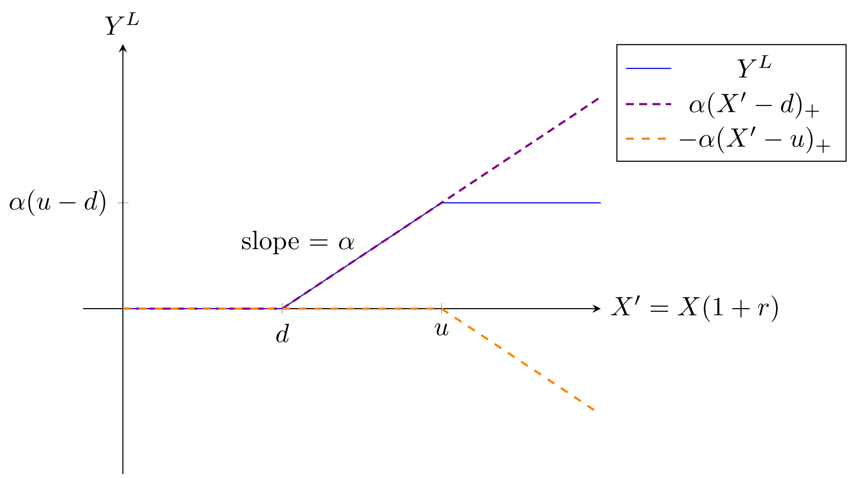
[5] A negative contribution effectively allows the party to *benefit from loss* , which can lead to some moral issues.



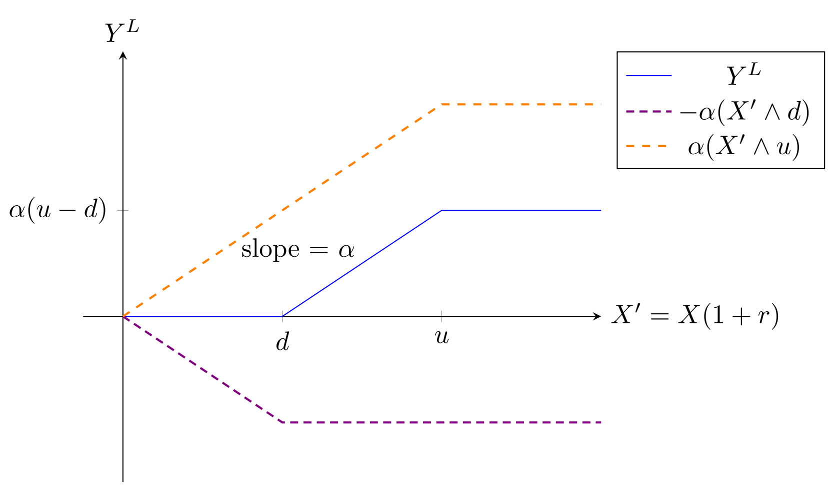
[Note: It has a similar shape as the payoff graph of *bull spread* in STAT3905.]

6.7.4 Based on the graph in [6.7.3], we can deduce the following formulas for Y^L:

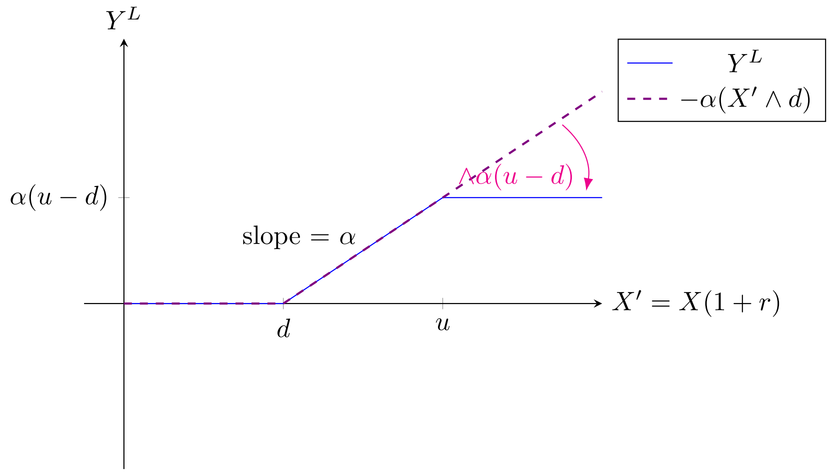
(a) $Y^L = \boxed{\alpha[(X'-d)_+ - (X'-u)_+]}$



(b) $Y^L = \boxed{\alpha[(X' \wedge u) - (X' \wedge d)]}$



(c) $Y^L = \boxed{\alpha[(X' - d)_+ \wedge (u - d)]}$



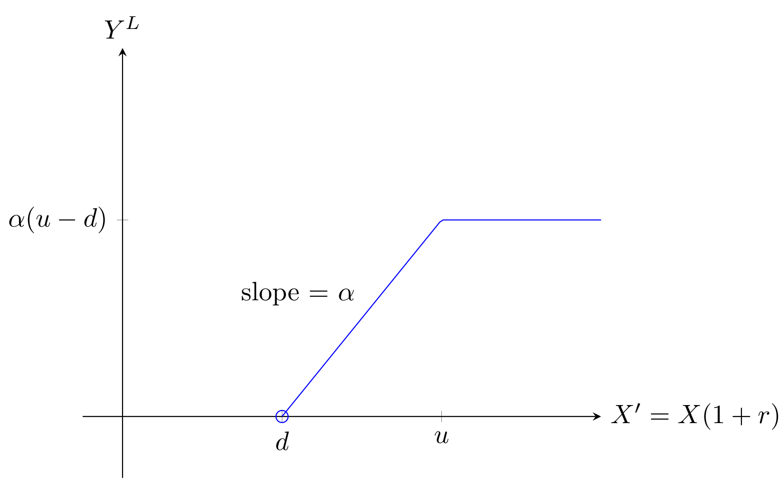
6.7.5 Also, from the graph in [6.7.3], we see that despite the "nominal" policy limit applied is u, the *true* policy limit for the resulting insurance (i.e., the maximum amount of payment made by the insurer) is $\alpha(u - d)$.

On the other hand, no additional payment is made by the insurer when the loss after inflation X' exceeds u, so we call u the **maximum covered loss** in this case.

6.7.6 In a similar manner, we can graphically show the final payment amount *per payment*

$$Y^P = (Y^L | Y^L > 0)$$

as follows:



6.7.7 By the expression in [6.7.4]b, we can obtain the means of Y^L and Y^P:

- $\mathbb{E}\left[Y^L\right] = \boxed{\alpha\{\mathbb{E}\left[X' \wedge u\right] - \mathbb{E}\left[X' \wedge d\right]\}} = \boxed{\alpha(1+r)\left\{\mathbb{E}\left[X \wedge \frac{u}{1+r}\right] - \mathbb{E}\left[X \wedge \frac{d}{1+r}\right]\right\}}.$
- $\mathbb{E}\left[Y^P\right] = \dfrac{\mathbb{E}\left[Y^L\right]}{\mathbb{P}\left(X' > d\right)} = \boxed{\dfrac{\mathbb{E}\left[Y^L\right]}{\mathbb{P}\left(X > \dfrac{d}{1+r}\right)}}.$

6.7.8 The *second moments* of Y^L and Y^P can be found as follows.

Proposition 6.7.a. Let $u^* = \dfrac{u}{1+r}$ and $d^* = \dfrac{d}{1+r}$ (with $u > d \implies u^* > d^*$). Then,

$$\mathbb{E}\left[(Y^L)^2\right] = \alpha^2(1+r)^2\left\{\mathbb{E}\left[(X \wedge u^*)^2\right] - \mathbb{E}\left[(X \wedge d^*)^2\right] - 2d^*\mathbb{E}\left[X \wedge u^*\right] + 2d^*\mathbb{E}\left[X \wedge d^*\right]\right\}$$

and

$$\mathbb{E}\left[(Y^P)^2\right] = \frac{\mathbb{E}\left[(Y^L)^2\right]}{\mathbb{P}\left(X > d^*\right)}.$$

<u>Proof</u>: By the expression in [6.7.4]b, we have

$$Y^L = \alpha[(X' \wedge u) - (X' \wedge d)] = \alpha(1+r)[(X \wedge u^*) - (X \wedge d^*)].$$

Thus,

$$\begin{aligned}\left(\frac{Y^L}{\alpha(1+r)}\right)^2 &= [(X \wedge u^*) - (X \wedge d^*)]^2 \\ &= (X \wedge u^*)^2 + (X \wedge d^*) - 2(X \wedge u^*)(X \wedge d^*) \\ &= (X \wedge u^*)^2 - (X \wedge d^*) - 2(X \wedge d^*)[(X \wedge u^*) - (X \wedge d^*)].\end{aligned}$$

The formula for $\mathbb{E}\left[(Y^L)^2\right]$ then follows since

$$(X \wedge d^*)[(X \wedge u^*) - (X \wedge d^*)] = \begin{cases} d^*(u^* - d^*) & \text{if } X > u^*; \\ d^*(X - d^*) & \text{if } d^* < X \leq u^*; \\ 0 & \text{if } X \leq d^* \end{cases} = d^*[(X \wedge u^*) - (X \wedge d^*)].$$

For $\mathbb{E}\left[(Y^P)^2\right]$, note that $Y^L > 0 \iff X(1+r) > d \iff X > d^*$ (see the graph in [6.7.3]), so

$$\mathbb{E}\left[(Y^P)^2\right] = \mathbb{E}\left[(Y^L)^2 | Y^L > 0\right] = \frac{\mathbb{E}\left[(Y^L)^2\mathbf{1}_{\{Y^L>0\}}\right]}{\mathbb{P}\left(Y^L > 0\right)} = \frac{\mathbb{E}\left[(Y^L)^2\mathbf{1}_{\{Y^L>0\}} + 0^2 \cdot \mathbf{1}_{\{Y^L=0\}}\right]}{\mathbb{P}\left(X > d^*\right)} = \frac{\mathbb{E}\left[(Y^L)^2\right]}{\mathbb{P}\left(X > d^*\right)}.$$

□

6.8 Impact of Deductibles on Claim Frequency

6.8.1 The presence of deductibles decreases (or at least does not increase) claim frequency (number of payments made by insurer) since some claims that present without deductibles may vanish (due to the loss amounts not exceeding the deductible) ➔ some losses do not trigger payment from 🏢 ➔ number of losses and number of payments can differ.

6.8.2 Suppose that there are N^L (random variable) independent losses, and let X_j be the amount/severity of jth loss. For a given insurance policy with possibly coverage modification, let v be the probability that a loss results in a payment. (For example, $v = \mathbb{P}\left(X > d\right)$ for an insurance with just an ordinary deductible of d.)

Define the indicator random variable I_j by

$$I_j = \begin{cases} 1 & \text{if } j\text{th loss results in payment;} \\ 0 & \text{otherwise.} \end{cases}$$

Then, we can express the total number of payments N^P as a *compound random variable*:

$$N^P = I_1 + I_2 + \cdots + I_{N^L}.$$

6.8.3 To obtain the pgf of N^P, first note that $I_1, I_2, \ldots$ are i.i.d. $\text{Ber}(v)$ random variables (by the independence assumption on the losses). Thus, they have the common pgf $P_I(t) = t^0(1-v) + t^1 v = 1 - v + vt$. Hence, the pgf of the compound random variable N^P is

$$P_{N^P}(t) = P_{N^L}(P_I(t)) = \boxed{P_{N^L}\big((1-v) + vt\big)}.$$

6.8.4 By [5.1.4], the mean and variance of the compound random variable N^P are: (let $I \sim \text{Ber}(v)$)

- $\mathbb{E}\left[N^P\right] = \mathbb{E}\left[N^L\right]\mathbb{E}\left[I\right] = \boxed{v\mathbb{E}\left[N^L\right]}$.
- $\text{Var}\left(N^P\right) = \mathbb{E}\left[N^L\right]\text{Var}\left(I\right) + \left(\mathbb{E}\left[I\right]\right)^2\text{Var}\left(N^L\right) = \boxed{v(1-v)\mathbb{E}\left[N^L\right] + v^2\,\text{Var}\left(N^L\right)}$.

6.8.5 The following result suggests a relationship between N^L and N^P under the condition that N^L is in the $(a, b, 0)$ or $(a, b, 1)$ class.

Theorem 6.8.a. Suppose that N^L is in the $(a, b, 0)$ or $(a, b, 1)$ class. Then, N^P and N^L belong to the same parametric family of distributions (with possibly different parameters)[6].

[6] In other words, their pgf take the same form, with possibly different parameters.

7 Aggregate Loss Model

7.0.1 Previously, when multiple losses are involved, we consider them in isolation. Here, we study the behaviour of their *sum* (called **aggregate loss**).

7.0.2 A practical setting where such sum arises is when an insurer sells many insurance policies ➔ pooling many risks and having "diversification".

7.1 Collective Risk Model

7.1.1 In the **collective risk model**, the aggregate loss is a sum S of N (random variable) individual losses with amounts $X_1, \dots, X_N$:

$$S = X_1 + \dots + X_N.$$

Furthermore we impose the following:

- When $N = 0$, we have $S = 0$.
- $X_1, X_2, \dots$ are i.i.d. random variables (with the same distribution as a random variable X) and are independent of N.

Remarks:

- This setting is similar to [5.1.3].
- The random variable N is called **claim count** (or **claim frequency**).
- Each random variable X_j is called **severity**/**single loss**/**individual loss**.
- The aggregate loss S is a *compound random variable* with primary (secondary) distribution being claim frequency distribution (severity distribution).

We work in the collective risk model henceforth in this section.

7.1.2 The pgf of S is

$$P_S(t) = \boxed{P_N(P_X(t))}$$

where P_N and P_X are the pgfs of N and X respectively.

7.1.3 The mgf of S is

$$M_S(t) = \mathbb{E}\left[\left(e^t\right)^S\right] = P_S(e^t) = P_N(P_X(e^t)) = \boxed{P_N(M_X(t))}$$

where M_X is the mgf of X.

7.1.4 By [5.1.4], the mean and variance of S are:

- $\mathbb{E}[S] = \boxed{\mathbb{E}[N]\,\mathbb{E}[X]}$.
- $\mathrm{Var}(S) = \boxed{\mathbb{E}[N]\,\mathrm{Var}(X) + \mathrm{Var}(N)\,(\mathbb{E}[X])^2}$.

7.1.5 Consider the collective risk model equipped with the setting in [6.8.2]. Suppose that the given insurance policy has an ordinary/franchise deducible of d applied to individual losses. Then, the probability that a loss results in a payment is

$$v = \mathbb{P}(X > d).$$

Recall that the total number of payments is

$$N^P = I_1 + I_2 + \dots + I_{N^L}$$

where N^L is the total number of losses.

Now, for any $i = 1, 2, \dots$, let the per-loss and per-payment variables be respectively:

- $Y_i^L = (X_i - d)_+$.
- $Y_i^P = (X_i - d | X_i > d)$.

7.1.6 In this setting, the **aggregate payment** can be calculated in the following ways:

- per-loss perspective: aggregate payment = $\boxed{Y_1^L + \cdots + Y_{N^L}^L}$ [Note: This sums "payments" associated to all losses (N^L of them), including those with zero value.]
- per-payment perspective: aggregate payment = $\boxed{Y_1^P + \cdots + Y_{N^P}^P}$ [Note: This sums only positive payments (N^P of them).]

Both methods result in the same aggregate payment since the difference between two expressions is just a sum of "payments" of zero value.

7.2 Convolution

7.2.1 To obtain more distributional quantities for the aggregate loss S in the collective risk model, we need the concept of *convolution*.

7.2.2 Let $X_1, X_2, \ldots$ be i.i.d. continuous random variables with the same cdf F_X and pdf f_X. Define

$$S_n = X_1 + \cdots + X_n$$

which is a sum of n i.i.d. random variables. Denote its cdf by F_X^{*n}:

$$F_X^{*n}(x) = \mathbb{P}(X_1 + \cdots + X_n \leq x) \quad \text{for any } x \in \mathbb{R}.$$

7.2.3 To find the cdf F_X^{*n}, we can perform the following recursive process:

(a) Define $F_X^{*1} = F_X$ (the common cdf of $X_1, X_2, \ldots$).

(b) Use the following recursive formula to find $F_X^{*2}, F_X^{*3}, \ldots$, up to F_X^{*n}:

$$\begin{aligned}
F_X^{*k}(x) &= \mathbb{P}(X_1 + \cdots + X_k \leq x) \\
&= \int_{-\infty}^{\infty} \mathbb{P}(X_1 + \cdots + X_k \leq x | X_k = y)\, f_X(y)\, \mathrm{d}y \quad \text{(law of total probability, continuous case)} \\
&= \int_{-\infty}^{\infty} \mathbb{P}(X_1 + \cdots + X_{k-1} \leq x - y)\, f_X(y)\, \mathrm{d}y \\
&= \boxed{\int_{-\infty}^{\infty} F_X^{*(k-1)}(x-y) f_X(y)\, \mathrm{d}y}
\end{aligned}$$

for any $k = 2, 3, \ldots$.

[Note: The operation here is a special case of *convolution* (mathematical concept) in the context of probability distributions.]

7.2.4 In case $X_1, X_2, \ldots$ are *nonnegative* (e.g., when they represent individual losses in the collective risk model), the recursive formula can be expressed as:

$$F_X^{*k}(x) = \boxed{\int_0^x F_X^{*(k-1)}(x-y) f_X(y)\, \mathrm{d}y}$$

for any $k = 2, 3, \ldots$ (since the integrand is zero when $y \notin (0, x)$).

Furthermore, the recursive formula for *pdf* can be obtained by differentiating both sides of this recursive formula:

$$f_X^{*k}(x) = \boxed{\int_0^x f_X^{*(k-1)}(x-y) f_X(y)\, \mathrm{d}y}$$

for any $k = 2, 3, \ldots$, where f_X^{*n} denotes the pdf of $X_1 + \cdots + X_n$.

7.2.5 If $X_1, X_2, \ldots$ become *discrete* (and are still nonnegative & i.i.d., with the same pmf f_X), then the recursive formulas in [7.2.4] become:

$$F_X^{*k}(x) = \boxed{\sum_{y=0}^{x} F_X^{*(k-1)}(x-y) f_X(y)}, \quad x = 0, 1, \ldots,$$

for any $k = 2, 3, \ldots$, and

$$f_X^{*k}(x) = \boxed{\sum_{y=0}^{x} f_X^{*(k-1)}(x-y) f_X(y)}, \quad x = 0, 1, \ldots,$$

for any $k = 2, 3, \ldots$, where f_X^{*n} denotes the pmf of $X_1 + \cdots + X_n$.

7.3 Cdf and Probability Function of the Aggregate Loss

7.3.1 Having the recursive formulas from the concept of convolution, we can derive the cdf and probability function of the aggregate loss S in the collective risk model. Henceforth we let $p_n = \mathbb{P}(N = n)$ be the pmf of the claim frequency N.

7.3.2 The cdf of the aggregate loss S is

$$\begin{aligned}
F_S(x) &= \sum_{n=0}^{\infty} p_n \cdot \mathbb{P}(S \leq x | N = n) && \text{(law of total probability)} \\
&= \sum_{n=0}^{\infty} p_n \cdot \mathbb{P}(X_1 + \cdots + X_n \leq x | N = n) \\
&= \sum_{n=0}^{\infty} p_n \cdot \mathbb{P}(X_1 + \cdots + X_n \leq x) && \text{(by independence assumption)} \\
&= \boxed{\sum_{n=0}^{\infty} p_n \cdot F_X^{*n}(x)}.
\end{aligned}$$

7.3.3 Based on the formula in [7.3.2], the probability function of the aggregate loss S is thus given by:

- (S is continuous): differentiating ➔ pdf is

$$f_S(x) = \boxed{\sum_{n=0}^{\infty} p_n \cdot f_X^{*n}(x)},$$

  where f_X^{*n} is the pdf of $X_1 + \cdots + X_n$.
- (S is discrete): pmf is

$$f_S(x) = F_S(x+1) - F_S(x) = \sum_{n=0}^{\infty} p_n \cdot [F_X^{*n}(x+1) - F_X^{*n}(x)] = \boxed{\sum_{n=0}^{\infty} p_n \cdot f_X^{*n}(x)},$$

  where $x = 0, 1, \ldots$ and f_X^{*n} is the pmf of $X_1 + \cdots + X_n$.

7.4 Panjer's Recursion for Discrete Severity

7.4.1 When the severity is discrete, the aggregate loss in the collective risk model S "matches" with the setting in [5.1.3] (just with different notations and "labels"), so *Panjer's recursion* is applicable for finding the distribution of S recursively.

7.4.2 By theorem 5.2.a, when the claim frequency N is in the $(a,b,0)$ class, the pmf of S can be obtained recursively by

$$f_S(x) = \boxed{\frac{1}{1-af_X(0)}\sum_{y=1}^{x}\left(a+\frac{by}{x}\right)f_X(y)f_S(x-y)}$$

for any $x \geq 1$ (in the support of S).

[Note: The starting value $f_S(0)$ can be obtained by

$$f_S(0) = P_S(0) = P_N(P_X(0)) = P_N(f_X(0)).$$

]

7.4.3 By theorem 5.2.b, when the claim frequency N is in the $(a,b,0)$ class, we have:

$$f_S(x) = \boxed{\frac{1}{1-af_X(0)}\left\{[p_1-(a+b)p_0]f_X(x)+\sum_{y=1}^{x}\left(a+\frac{by}{x}\right)f_X(y)f_S(x-y)\right\}}$$

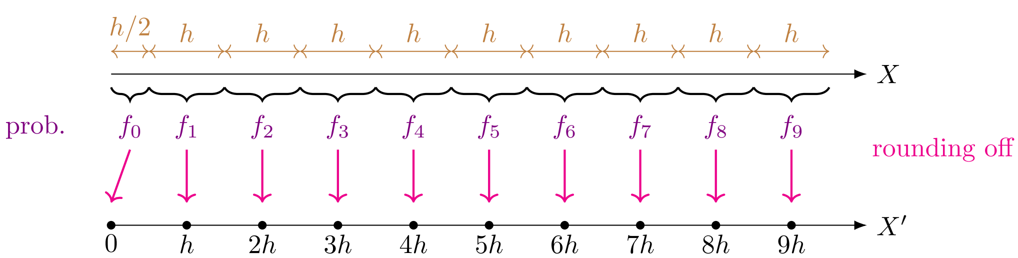
for any $x \geq 1$ (in the support of S).

7.5 Method of Rounding

7.5.1 To apply the recursive formulas in [7.4.2] and [7.4.3], the severity X needs to be *discrete*. Thus, they are not applicable for *continuous* severity. Nevertheless, we may still "apply" the formulas by *approximating* the continuous severity distribution by a discrete distribution (known as **discretization**). Here, we introduce a way of discretization: *method of rounding*.

7.5.2 The **method of rounding** is as follows. Let X be the original continuous severity with cdf F_X. Then, we approximate X by a discrete random variable X' defined by:

$$X' = \begin{cases} 0 & \text{w.p. } f_0 = F_X(h/2); \\ h & \text{w.p. } f_1 = F_X(h+h/2) - F_X(h-h/2); \\ 2h & \text{w.p. } f_2 = F_X(2h+h/2) - F_X(2h-h/2); \\ \vdots & \vdots \\ jh & \text{w.p. } f_j = F_X(jh+h/2) - F_X(jh-h/2); \\ \vdots & \vdots \end{cases}$$



Remarks:

- "w.p." means "with probability".
- We call h the **span**.

7.6 Stop-Loss Premium

7.6.1 Practically, it is common for an insurance to not apply deductible *individually* on each loss claimed, but apply it on the losses claimed for a certain period (say a year) *in aggregate*.

7.6.2 Consider an aggregate loss S (sum of individual losses in a certain period). An insurance on the aggregate loss S subject to a deductible d is called **stop-loss insurance**. The expected payment per loss of this insurance is called the **(net) stop-loss premium**:

$$\mathbb{E}\left[(S-d)_+\right].$$

Remarks:

- Here we are not required to work in the collective risk model.
- The insurance "stops" the aggregate loss suffered by the policyholder (once d dollar is reached) → hence "stop-loss insurance".
- If one uses the *equivalence principle* (without considering expenses) to price the insurance policy, the net premium obtained would be $\mathbb{E}\left[(S-d)_+\right]$ (the expected amount of benefit) → hence "net stop-loss premium".

7.6.3 Here we recall some formulas for computing the stop-loss premium discussed in section 1:

- (proposition 1.2.a)

$$\mathbb{E}\left[(S-d)_+\right] = \boxed{\int_d^\infty [1-F_S(x)]\,\mathrm{d}x}$$

  where F_S is the cdf of S.

- ([1.2.3]) If S is continuous with pdf f_S, then

$$\mathbb{E}\left[(S-d)_+\right] = \boxed{\int_d^\infty (x-d)f_S(x)\,\mathrm{d}x}.$$

  If S is discrete with pmf f_S, then

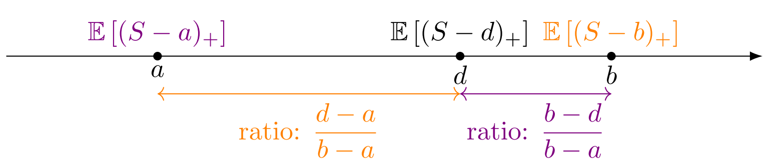
$$\mathbb{E}\left[(S-d)_+\right] = \boxed{\sum_{s_j>d}(s_j-d)\cdot\mathbb{P}\left(S=s_j\right)}.$$

7.6.4 If there is an interval in which the aggregate loss S has no probability to lie, the following result may simplify calculations.

Proposition 7.6.a. Suppose that $\mathbb{P}\left(a<S<b\right)=0$ for some $a<b$. Then, for any $d\in[a,b]$,

$$\mathbb{E}\left[(S-d)_+\right] = \frac{b-d}{b-a}\cdot\mathbb{E}\left[(S-a)_+\right] + \frac{d-a}{b-a}\cdot\mathbb{E}\left[(S-b)_+\right].$$

[Note: This means that the stop-loss premium can be calculated using *linear interpolation*.]



Proof: By assumption, the cdf

$$F_S(x) = F_S(a)$$

for any $x \in [a, b)$. Hence,

$$\begin{aligned}\mathbb{E}[(S-d)_+] &= \int_d^\infty [1 - F_S(x)]\,\mathrm{d}x \\ &= \int_a^\infty [1 - F_S(x)]\,\mathrm{d}x - \int_a^d [1 - F_S(a)]\,\mathrm{d}x \\ &= \mathbb{E}[(S-a)_+] - [1 - F_S(a)](d-a).\end{aligned}$$

Setting $d = b$ here gives

$$\mathbb{E}[(S-b)_+] = \mathbb{E}[(S-a)_+] - [1 - F_S(a)](b-a) \implies 1 - F_S(a) = \frac{\mathbb{E}[(S-b)_+] - \mathbb{E}[(S-a)_+]}{b-a},$$

so it follows that

$$\mathbb{E}[(S-d)_+] = \mathbb{E}[(S-a)_+] - \frac{\mathbb{E}[(S-b)_+] - \mathbb{E}[(S-a)_+]}{b-a}\cdot(d-a) = \frac{b-d}{b-a}\cdot\mathbb{E}[(S-a)_+] + \frac{d-a}{b-a}\cdot\mathbb{E}[(S-b)_+].$$

□

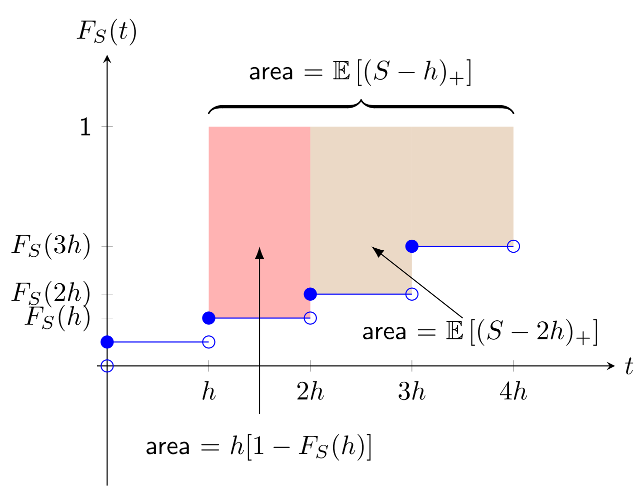
7.6.5 When the aggregate loss S is *discrete*, the stop-loss premiums are related as follows.

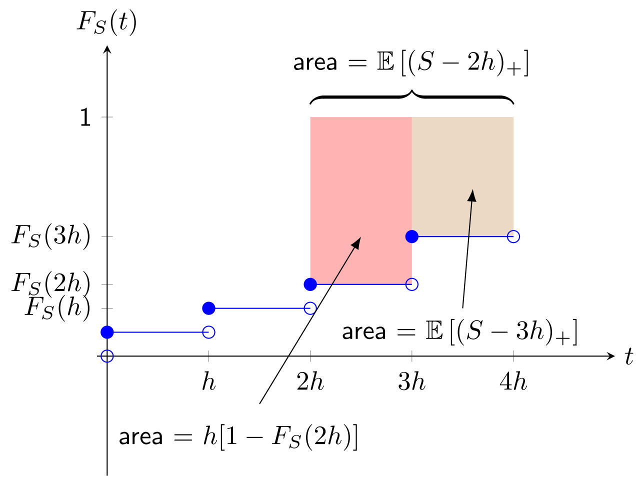
Proposition 7.6.b. Suppose that S is a discrete random variable with support $\{0, h, 2h, \dots\}$ for some fixed $h > 0$. Then, for any $j \in \mathbb{N}_0$,

$$\mathbb{E}[(S-(j+1)h)_+] = \mathbb{E}[(S-jh)_+] - h[1 - F_S(jh)]$$

where F_S is the cdf of S.

Proof:





Based on the geometrical interpretation of expectation (area between the graph of F_S and the line $y = 1$, starting at $t = d$ for $\mathbb{E}\left[(S-d)_+\right]$), the result follows readily. $\square$

8 Risk Measures

8.1 Introduction

8.1.1 For a loss, we use a nonnegative loss random variable X to measure it. Now, consider a *risk*, which is an exposure to possibility of loss. The main goal in this section is to study how to *measure* a *risk* (quantifying exposure/risk "level"). It turns out that, unlike a loss, there are multiple ways to measure a risk.

8.1.2 A **risk measure** is a function $\rho : \mathcal{X} \to \mathbb{R}$ where $\mathcal{X}$ is the set of all possible nonnegative loss random variables. For any loss random variable X, we assign a number $\rho(X)$ to it, which serves as a measure for the risk corresponding to that loss.

The value $\rho(X)$ is to be interpreted as the amount of *capital* needed for the insurer to *protect* against the loss X. Then, larger value of $\rho(X)$ indicates higher risk.

[Note: This interpretation is more natural if the function ρ satisfies certain properties, which will be discussed in section 8.8. However, these properties are not *required* in the definition.]

8.2 Premium Principles

8.2.1 The first use of risk measures in actuarial science was the development of *premium principles* (e.g., equivalence principle in STAT3901). The premium principles are applied to a loss distribution (distribution of loss X) for determining a suitable premium to charge for insuring the loss X.

8.2.2 Some examples of premium principles expressed in the language of risk measures are:

- **expected value premium principle**: $\rho(X) = (1 + k)\mathbb{E}[X]$ for some $k \geq 0$
- **standard deviation premium principle**: $\rho(X) = \mathbb{E}[X] + k\sqrt{\mathrm{Var}(X)}$ for some $k \geq 0$
- **variance premium principle**: $\rho(X) = \mathbb{E}[X] + k\,\mathrm{Var}(X)$ for some $k \geq 0$

[Note: To follow one of these premium principle, the premium to be charged for insuring the loss X is $\rho(X)$.
]

8.2.3 Each of these premium principles gives a premium that is *at least* the expected loss $\mathbb{E}[X]$. The excess amount serves as a "cushion" against adverse experience. Such excess amount is known as **premium loading**:

$$\text{premium loading} = \text{premium} - \mathbb{E}[X].$$

In the standard deviation and variance principles, the premium loading is related to the variance of the loss X:

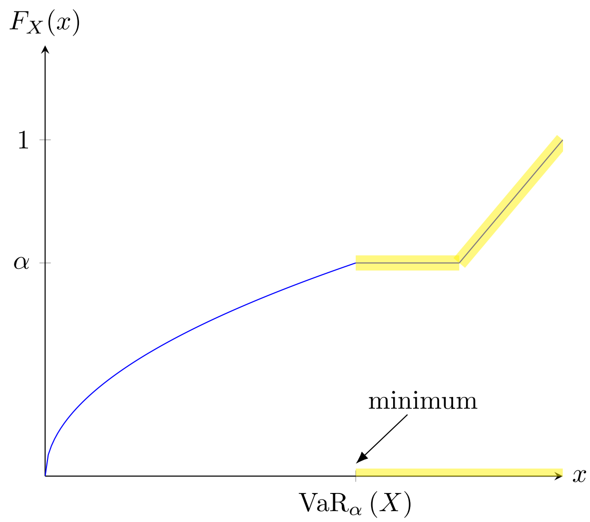
- standard deviation premium principle: premium loading $= \alpha\sqrt{\mathrm{Var}(X)}$
- variance premium principle: premium loading $= \alpha\,\mathrm{Var}(X)$

8.3 Value-at-Risk

8.3.1 A popular risk measure is the *value-at-risk* (VaR). Let $\alpha \in (0, 1)$. [Warning: Here 0 and 1 are *excluded*!] The **value-at-risk at confidence level α** (or **α-VaR**) of a loss X is the αth quantile (or $100\alpha\%$ percentile) of X:

$$\mathrm{VaR}_\alpha(X) = \min\{x \geq 0 : F_X(x) \geq \alpha\}$$

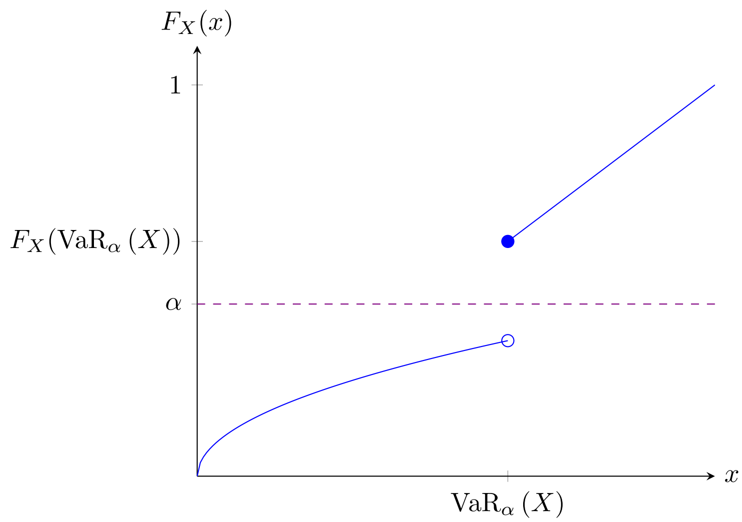
where F_X is the cdf of X.



Remarks:

- In most cases, α-VaR is the value that is not exceeded by the loss X with probability α, i.e., $\mathbb{P}(X \leq \mathrm{VaR}_\alpha(X)) = \alpha$. (See [8.3.2] for an example where this is not the case.)
- It can be interpreted as the amount of capital required to ensure that the loss can be "absorbed" by the insurer 🏢 (so that 🏢 would not bankrupt) with a high degree of certainty (when α is large).
- If F_X is continuous and strictly increasing, we have $\mathrm{VaR}_\alpha(X) = F_X^{-1}(\alpha)$.

8.3.2 If F_X is continuous, then we always have $F_X(\mathrm{VaR}_\alpha(X)) = \alpha$. Thus, α-VaR is the value that is not exceeded by the loss X with probability α. However, this is not necessarily the case if F_X is not continuous:

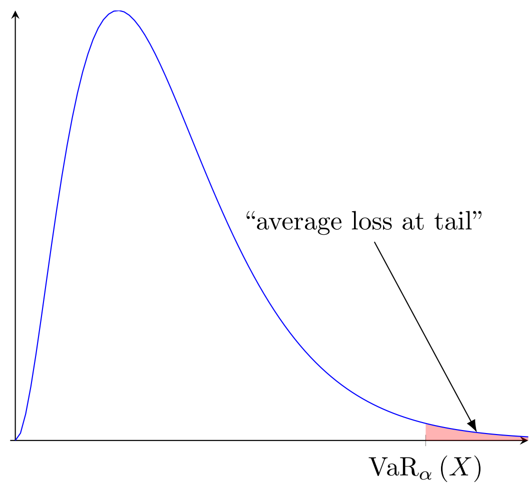


In this case, we have $F_X(\mathrm{VaR}_\alpha(X)) > \alpha$.

8.4 Conditional Tail Expectation

8.4.1 Let $\alpha \in (0, 1)$. The **conditional tail expectation at confidence level α** (or **α-CTE**) is the expected loss given that the loss exceeds its α-VaR:

$$\mathrm{CTE}_\alpha(X) = \mathbb{E}[X | X > \mathrm{VaR}_\alpha(X)].$$

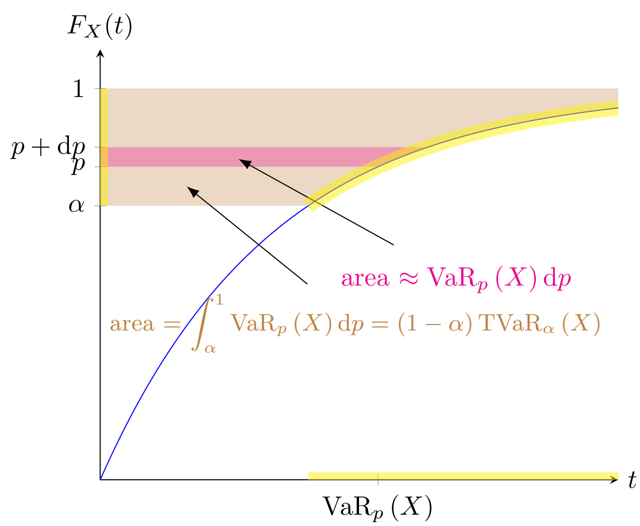


[Note: From [8.3.2], the probability $\mathbb{P}(X > \mathrm{VaR}_{\alpha}(X))$ is $1-\alpha$ if F_X is continuous, and it is not necessarily $1-\alpha$.]

8.5 Tail Value-at-Risk

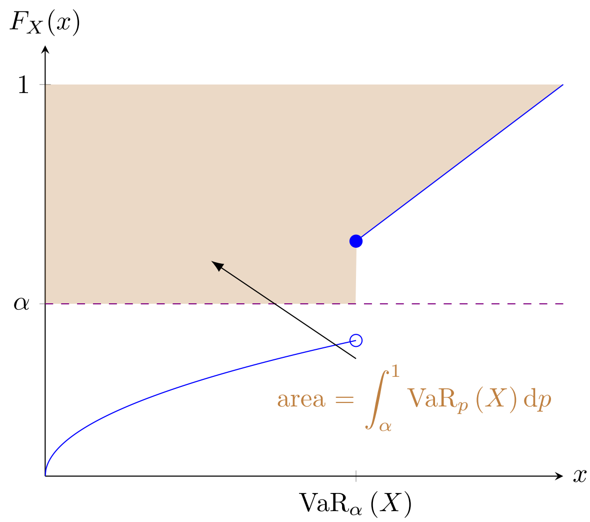
8.5.1 Let $\alpha \in (0,1)$. The **tail value-at-risk at confidence level α** (or **α-TVaR**) of the loss X is the "average" of the VaR at confidence level not less than α:

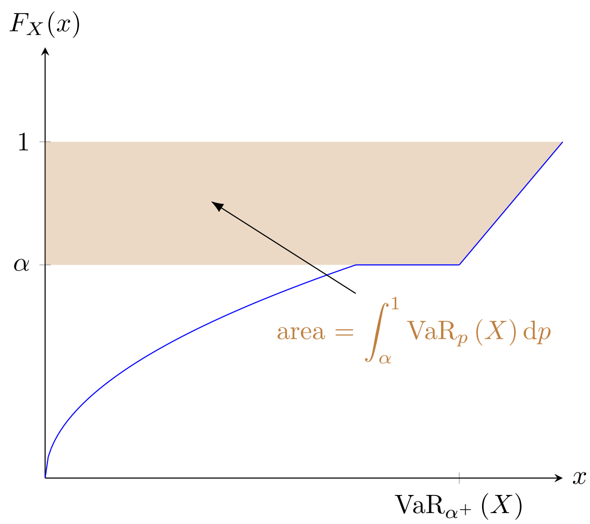
$$\mathrm{TVaR}_{\alpha}(X) = \frac{1}{1-\alpha}\int_{\alpha}^{1} \mathrm{VaR}_p(X)\,\mathrm{d}p\,.$$



[Note: We consider VaR at confidence level in the "right tail".]

8.5.2 The following are several more cases represented geometrically:





[Note: α^+ represents a value that is "just" larger than α.]

8.6 Expected Shortfall

8.6.1 Let $\alpha \in (0,1)$. The **expected shortfall at confidence level α** (or **α-ESF**) of the loss X is the expected value of shortfall (with respect to α-VaR):

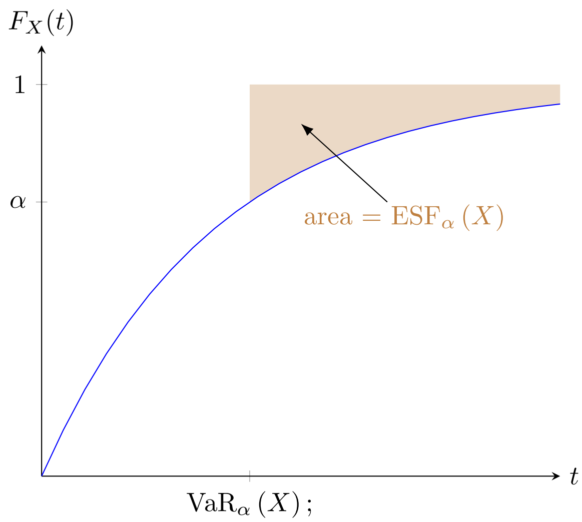
$$\mathrm{ESF}_\alpha(X) = \mathbb{E}\left[(X - \mathrm{VaR}_\alpha(X))_+\right].$$

[Note: The shortfall with respect to α-VaR is the excess amount of loss over the α-VaR (capital reserved based on VaR) when the loss exceeds the α-VaR, and is zero otherwise.]

8.6.2 By proposition 1.2.a, we can express $\mathrm{ESF}_\alpha(X)$ as:

$$\mathrm{ESF}_\alpha(X) = \int_{\mathrm{VaR}_\alpha(X)}^{\infty} [1 - F_X(x)]\,\mathrm{d}x.$$

We can represent the expected shortfall $\mathrm{ESF}_\alpha(X)$ geometrically as follows:

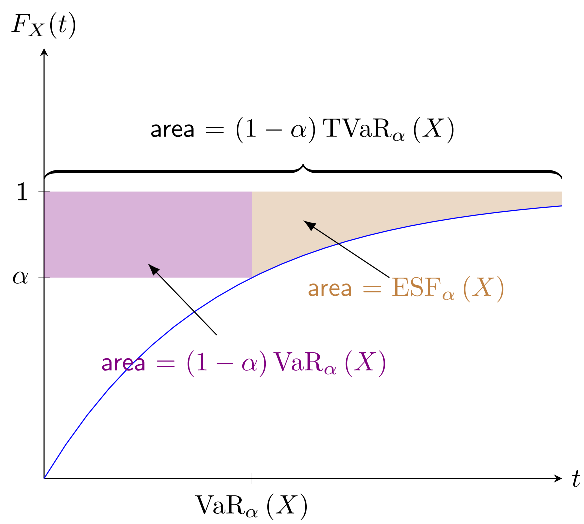


8.7 Relationships Between VaR, CTE, TVaR and ESF

8.7.1 For any $\alpha \in (0,1)$, we have:

(a) $\mathrm{TVaR}_\alpha(X) = \boxed{\mathrm{VaR}_\alpha(X) + \frac{1}{1-\alpha}\cdot \mathrm{ESF}_\alpha(X)}$

Proof:



From this, we see that

$$(1-\alpha)\,\mathrm{TVaR}_\alpha(X) = (1-\alpha)\,\mathrm{VaR}_\alpha(X) + \mathrm{ESF}_\alpha(X),$$

which implies that

$$\mathrm{TVaR}_\alpha(X) = \mathrm{VaR}_\alpha(X) + \frac{1}{1-\alpha}\cdot \mathrm{ESF}_\alpha(X).$$

(The argument is similar when the cdf has other shape.) $\square$

(b) $\mathrm{CTE}_\alpha(X) = \boxed{\mathrm{TVaR}_{F_X(\mathrm{VaR}_\alpha(X))}(X)}$ (when $0 < F_X(\mathrm{VaR}_\alpha(X)) < 1$)

Proof: We have

$$\begin{aligned}\mathrm{CTE}_\alpha(X) = \mathbb{E}[X|X > \mathrm{VaR}_\alpha(X)] = \mathrm{VaR}_\alpha(X) + \underbrace{\mathbb{E}[X - \mathrm{VaR}_\alpha(X)|X > \mathrm{VaR}_\alpha(X)]}_{e_X(\mathrm{VaR}_\alpha(X))} \\ \overset{[1.3.5]}{=} \mathrm{VaR}_\alpha(X) + \frac{\mathrm{ESF}_\alpha(X)}{1 - F_X(\mathrm{VaR}_\alpha(X))} \\ = \mathrm{VaR}_{F_X(\mathrm{VaR}_\alpha(X))}(X) + \frac{1}{1 - F_X(\mathrm{VaR}_\alpha(X))} \cdot \mathrm{ESF}_{F_X(\mathrm{VaR}_\alpha(X))}(X) \\ \overset{[8.7.1]a}{=} \mathrm{TVaR}_{F_X(\mathrm{VaR}_\alpha(X))}(X).\end{aligned}$$

For the second-to-last equality, see the proof of [8.7.1]c. □

(c) $\mathrm{CTE}_\alpha(X) = \boxed{\mathrm{VaR}_\alpha(X) + \frac{1}{1 - F_X(\mathrm{VaR}_\alpha(X))} \cdot \mathrm{ESF}_\alpha(X)}$ (when $F_X(\mathrm{VaR}_\alpha(X)) < 1$)

Proof: Apply [8.7.1]b followed by [8.7.1]a, and note that

$$\mathrm{VaR}_{F_X(\mathrm{VaR}_\alpha(X))}(X) = \min\{x \geq 0 : F_X(x) \geq F_X(\mathrm{VaR}_\alpha(X))\} = \mathrm{VaR}_\alpha(X),$$

and

$$\mathrm{ESF}_{F_X(\mathrm{VaR}_\alpha(X))}(X) = \mathbb{E}\left[\left[X - \mathrm{VaR}_{F_X(\mathrm{VaR}_\alpha(X))}(X)\right]_+\right] = \mathbb{E}\left[[X - \mathrm{VaR}_\alpha(X)]_+\right] = \mathrm{ESF}_\alpha(X).$$

□

[Note: If $F_X(\mathrm{VaR}_\alpha(X)) = \alpha$ (which is the case when F_X is continuous), then we can simplify the formula in [8.7.1]b to:

$$\mathrm{CTE}_\alpha(X) = \boxed{\mathrm{TVaR}_\alpha(X)}.$$

This shows CTE and TVaR are equivalent in this case!]

8.8 Coherent Risk Measures

8.8.1 When the risk measure ρ satisfies certain properties, the interpretation of $\rho(X)$ as amount of capital needed for the insurer to protect against the loss X is more natural.

8.8.2 The properties are as follows. [Note: The letters X and Y denote arbitrary nonnegative loss random variables (in the set $\mathcal{X}$) in the following.]

- **translation invariance** (TI):

$$\rho(X + c) = \rho(X) + c \quad \text{for any constant } c.$$

  [Note: This means adding a constant amount of loss (positive or negative) implies addition of the same amount to the required capital for protecting that loss.]

- **positive homogeneity** (PH):

$$\rho(\lambda X) = \lambda\rho(X) \quad \text{for any constant } \lambda > 0.$$

  [Note: This means changing the unit of loss only leads to the corresponding unit change for the amount of capital required (but not change in the "actual" amount).[7]]

[7] For example, suppose that the exchange rate is 1 USD = 160 JPY. Let X be the loss in USD. Suppose the amount of capital required in USD is $\rho(X) = 100$ (USD). Changing the currency unit to JPY, the loss becomes $160X$ (in JPY), and the capital required in JPY should be

$$160\rho(X) = 16000 \text{ (JPY)}$$

in order to have no change in the "actual" amount. Hence, we should have

$$\rho(160X) = 160\rho(X).$$

- **monotonicity** (M):
$$\mathbb{P}(X \leq Y) = 1 \implies \rho(X) \leq \rho(Y).$$
[Note: This means when a loss Y is not less than another loss X (with probability 1), the required capital for protecting the loss Y should also be not less than that for the loss X.]
- **subadditivity** (S):
$$\rho(X+Y) \leq \rho(X) + \rho(Y).$$
[Note: This means diversification (adding/combining different losses together) cannot possibly make the resulting *total* amount of capital required greater than what we would have if there were no diversification.]

A risk measure satisfying TI, PH, M, and S is called a **coherent risk measure**.

8.8.3 The following summarizes the coherence of some common risk measures.

Risk measure	Coherent?
$\mathbb{E}[\cdot]$	yes[8]
expected value premium principle	no (fails TI) (when $k > 0$)
standard deviation premium principle	no (fails M) (when $k > 0$)
variance premium principle	no (fails M) (when $k > 0$)
VaR	no (fails S)
CTE	no (fails S)
ESF	no (fails S)
TVaR	yes[9]

[8] TI, PH, and S follow from linearity of expectation. M follows from monotonicity of expectation.
[9] It turns out to be highly non-trivial to prove this.

References

Klugman, S. A., Panjer, H. H., & Willmot, G. E. (2019). *Loss models: From data to decisions* (5th ed.). John Wiley & Sons.

Sundt, B., & Jewell, W. S. (1981). Further results on recursive evaluation of compound distributions. *ASTIN Bulletin*, *12*(1), 27–39.

Concepts and Terminologies

Results

Section 1

- proposition 1.1.a: formula for $\mathbb{E}[X]$ based on survival function
- corollary 1.1.b: formula for $\mathbb{E}[X]$ based on survival function where X is discrete
- [1.2.3]: direct formulas for $\mathbb{E}[(X-d)_+]$ (continuous and discrete cases)
- proposition 1.2.a: formula for $\mathbb{E}[(X-d)_+]$ based on survival function
- [1.3.3]: formulas for MRL $e_X(d)$
- [1.3.4]: formulas for kth moment of excess loss $e_X^k(d)$
- [1.3.5]: relationship between MRL $e_X(d)$ and $\mathbb{E}[(X-d)_+]$
- [1.4.3]: formulas for $\mathbb{E}\left[(X \wedge u)^k\right]$
- [1.4.4]: relationship between stop loss and limited loss variables
- proposition 1.5.a: existence of kth (positive) moment implies existence of all smaller positive moments
- [1.5.7]: equality of limit of ratio of survival functions and limit of ratio of pdfs
- proposition 1.5.b: X has DFR (IFR) implies X has IMRL (DMRL)

Section 2

- [2.1.2]: expression of probability function from mixing finitely many random variables
- [2.1.3]: expression of probability function from mixing countably infinitely many random variables
- [2.1.4]: expression of probability function from mixing uncountably infinitely many random variables
- proposition 2.1.a: expression of cdf from mixing random variables
- theorem 2.2.a: law of total expectation
- proposition 2.2.b: law of total variance

Section 3

- theorem 3.1.a: probability generating property of pgf
- corollary 3.1.b: same pgf implies same distribution
- proposition 3.1.c: (partial) moment generating property of pgf
- proposition 3.1.d: pgf of a sum of independent nonnegative discrete random variables
- [3.2.2]: pgf of $N \sim \text{Poi}(\lambda)$
- [3.2.3]: mean and variance of $N \sim \text{Poi}(\lambda)$
- theorem 3.2.a: convolution for Poisson distribution
- theorem 3.2.b: thinning for Poisson distribution
- [3.3.3]: mean and variance of a mixed Poisson random variable

- proposition 3.3.a: pgf of N when $N|\Theta \sim \text{Poi}(\lambda\Theta)$ (special case for mixed Poisson distribution)
- [3.4.4]: pgf of $N \sim \text{NB}(r, \beta)$
- [3.4.5]: mean and variance of $N \sim \text{NB}(r, \beta)$
- proposition 3.4.a: $N|\Lambda \sim \text{Poi}(\Lambda)$ and $\Lambda \sim \text{Gamma}(\alpha, \theta) \implies N \sim \text{NB}(\alpha, \theta)$
- [3.5.2]: pgf of $N \sim \text{Geom}(\beta)$
- [3.5.4]: $N|\Lambda \sim \text{Poi}(\Lambda)$ and $\Lambda \sim \text{Exp}(\theta) \implies N \sim \text{Geom}(\theta)$
- proposition 3.5.a: memoryless property of geometric distribution
- [3.6.2]: pgf of $N \sim \text{Bin}(n, p)$
- [3.6.3]: mean and variance of $N \sim \text{Bin}(n, p)$

Section 4

- theorem 4.1.a: the only types of distributions in the $(a, b, 0)$ class
- [4.1.5]: characterization of $\text{Poi}(\lambda)$ using the $(a, b, 0)$ class
- [4.1.6]: characterization of $\text{NB}(r, \beta)$ using the $(a, b, 0)$ class
- [4.1.7]: characterization of $\text{Bin}(m, q)$ using the $(a, b, 0)$ class
- [4.3.3]: formula for constant c in the zero-modified $(a, b, 1)$ class
- [4.3.4]: formula for kth moment of a random variable in the zero-modified $(a, b, 1)$ class
- [4.3.5]: formula for pgf of a random variable in the zero-modified $(a, b, 1)$ class
- [4.3.7]: formulas for constant c and pgf of a random variable in the zero-truncated $(a, b, 1)$ class
- [4.4.6]: steps for computing $\text{ETNB}(r, \beta)$ probabilities

Section 5

- theorem 5.1.a: equivalent definition of compound distribution using sum of N i.i.d. random variables
- [5.1.4]: mean and variance of compound distribution
- theorem 5.2.a: Panjer's recursion ($(a, b, 0)$ class)
- theorem 5.2.b: Panjer's recursion ($(a, b, 1)$ class)
- [5.3.2]: pgf of a compounded Poisson random variable
- [5.3.3]: expression of a pgf for the underlying random variable to follow compound Poisson distribution
- theorem 5.3.a: convolution for compound Poisson distribution

Section 6

- [6.1.3]: expression of Y^P in terms of Y^L (ordinary deductible)
- [6.1.4]: distributional quantities of Y^P (ordinary deductible)
- [6.1.5]: distributional quantities of Y^L (ordinary deductible)
- [6.1.6]: means of Y^L and Y^P (ordinary deductible)
- [6.2.3]: expression of Y^P in terms of Y^L (franchise deductible)
- [6.2.4]: distributional quantities of Y^P (franchise deductible)
- [6.2.5]: distributional quantities of Y^L (franchise deductible)
- [6.2.6]: means of Y^L and Y^P (franchise deductible)
- [6.4.4]: expression of Y^L with inflation incorporated in the insurance
- [6.4.5]: expression of Y^P with inflation incorporated in the insurance
- [6.5.3]: distributional quantities of $Y = X \wedge u$
- [6.5.4]: expected payment made by insurer per loss for an insurance with policy limit u which incorporates inflation
- [6.7.4]: expressions of Y^L for general insurance
- [6.7.7]: means of Y^L and Y^P for general insurance
- proposition 6.7.a: second moments of Y^L and Y^P for general insurance
- [6.8.3]: pgf of N^P with deductible considered
- [6.8.4]: mean and variance of N^P with deductible considered
- theorem 6.8.a: same family for N^P and N^L when N^L is in the $(a, b, 0)$ or $(a, b, 1)$ class

Section 7

- [7.1.2]: pgf of S in the collective risk model
- [7.1.3]: mgf of S in the collective risk model
- [7.1.4]: mean and variance of S in the collective risk model
- [7.1.6]: formulas for aggregate payment in the collective risk
- [7.2.3]: steps for performing convolution for a continuous random variable
- [7.2.4]: recursive formulas for cdf and pdf for convolution (continuous random variable)
- [7.2.5]: recursive formulas for cdf and pdf for convolution (discrete random variable)
- [7.3.2]: cdf of S in the collective risk model
- [7.3.3]: probability function of S in the collective risk model
- [7.4.2]: Panjer's recursion ($(a, b, 0)$ class) for S in the collective risk model
- [7.4.3]: Panjer's recursion ($(a, b, 1)$ class) for S in the collective risk model
- proposition 7.6.a: linear interpolation formula for stop-loss premium
- proposition 7.6.b: relationship between stop-loss premiums when the aggregate loss S is discrete

Section 8

- [8.7.1]a: formula of TVaR in terms of VaR and ESF
- [8.7.1]b: formula of CTE in terms of TVaR
- [8.7.1]c: formula of CTE in terms of VaR and ESF
- [8.8.3]: coherence of some common risk measures