CA MAS-I Chapter 1

Overview - Learn MAS-I - Coaching Actuaries

5m

Part of this exam focuses on the application of probability. Specifically, it teaches us how to apply our probability knowledge in building a risk model and solving insurance-related problems. Due to the importance of having a strong foundation in probability, this section covers the fundamental concepts of probability.

Many concepts in this section have been introduced in previous exams. However, we do not recommend skipping this section because it also contains new ideas that were not previously covered. In any case, it will be beneficial to review the basics.

1.0.1 Probability Functions

5m

There are two types of random variables:

- *Discrete*
  Takes values from a specified set of exact countable values
- *Continuous*
  Takes values from a specified interval or a collection of intervals

Probability Mass Function (PMF)

If a random variable X is discrete, then for any value x, the probability of $X = x$ can be calculated by evaluating its *probability mass function (PMF)* at x, which is commonly denoted by one of the following:

- $\Pr(X = x)$
- $p_X(x)$
- $p(x)$

Coach's Remarks

The subscript X in $p_X(x)$ refers to the random variable X and is usually included to distinguish between multiple random variables. However, if the random variable is implied, it is common to drop the subscript.

The PMF produces probabilities, and all probabilities must be between 0 and 1. The sum of the probabilities of all possible values of a random variable must equal 1. Therefore, a valid PMF must satisfy the following properties:

- $0 \leq p(x) \leq 1$
- $\sum_{\text{all } x} p(x) = 1$

Probability Density Function (PDF)

A *probability density function (PDF)*, or *density function* for short, is the equivalent of a PMF for continuous random variables. For a continuous random variable X, its PDF is denoted by:

- $f_X(x)$
- $f(x)$

Unlike a PMF, evaluating a PDF at a certain value does not produce the probability at that value. In other words,

$$f(x) \neq \Pr(X = x)$$

Instead, calculate probabilities for a continuous random variable by integrating the PDF. More generally,

$$\Pr(a \leq X \leq b) = \int_a^b f(x)\,\mathrm{d}x$$

Using the same logic, we can conclude the probability of X being any particular value is zero:

$$\begin{aligned}\Pr(X = a) &= \Pr(a \leq X \leq a) \\ &= \int_a^a f(x)\,\mathrm{d}x \\ &= 0\end{aligned}$$

This is because, for continuous random variables, the possible values are uncountably infinite, making any exact value immensely improbable. Thus, non-zero probabilities exist over **intervals** of values.

Coach's Remarks

Although evaluating a PDF at a particular value does not yield a probability, it describes the relative likelihood for the continuous random variable to take on a given value. For example, if the PDF evaluated at a is greater than at b, then X is more likely to be in the neighborhood of a than in the neighborhood of b.

Recall that PMFs must satisfy two properties to be valid. There are two analogous properties for PDFs.

- $f(x) \geq 0$
- $\displaystyle\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1$

All PDFs must be non-negative. However, they are allowed to exceed 1 because PDFs do not represent actual probabilities. A PDF above 1 **does not** translate to a probability exceeding 1. The integral of the PDF over all values of the random variable must equal 1.

1.0.2 CDFs, Survival Functions, and Hazard Functions

10m

Cumulative Distribution Function (CDF)

The *cumulative distribution function (CDF)*, often called the *distribution function*, of a random variable X is the probability that X does not exceed a given value. This definition holds true for both discrete and continuous random variables. The CDF is commonly denoted by one of the following:

- $\Pr(X \le x)$
- $F_X(x)$
- $F(x)$

Coach's Remarks

Notice the inequality used in the CDF definition is "less than or equal to" or "not exceeding", i.e. $\le$, rather than "strictly less than", $<$.

This distinction is important for **discrete** distributions because the probability of a particular value might not be zero. In other words, for a **discrete** random variable X, it is possible that

$$\begin{aligned} F(x) &= \Pr(X \le x) \\ &= \Pr(X < x) + p(x) \\ &\ne \Pr(X < x) \end{aligned}$$

However, it is not important to make this distinction for **continuous** random variables. This is because the probability of a particular value is always zero. Therefore, for a **continuous** random variable X,

$$\begin{aligned} F(x) &= \Pr(X \le x) \\ &= \Pr(X < x) + 0 \\ &= \Pr(X < x) \end{aligned}$$

The CDF is computed as a sum for discrete random variables and as an integral for continuous random variables.

$$\begin{aligned} F(x) &= \Pr(X \le x) \\ &= \sum_{t \le x} p(t) && \text{(discrete)} \\ &= \int_{-\infty}^{x} f(t)\,\mathrm{d}t && \text{(continuous)} \end{aligned} \tag{1.0.2.1}$$

The definition for the **continuous** CDF implies the first derivative of the CDF returns the PDF.

$$f(x) = \frac{\mathrm{d}}{\mathrm{d}x} F(x) \tag{1.0.2.2}$$

All CDFs, discrete or continuous, satisfy the following properties:

- $F(-\infty) = 0$
- $F(\infty) = 1$

Survival Function

The *survival function* is the complement of the CDF. Thus, evaluating a survival function at x yields the probability that the random variable exceeds x. The survival function is commonly denoted by one of the following:

- $\Pr(X > x)$

- $\Pr(X > x)$
- $S_X(x)$
- $S(x)$

Coach's Remarks

Since the survival function has a complementary relationship with the CDF, we need to pay attention to its inequality sign.

$$\begin{aligned} S(x) &= \Pr(X > x) \\ &\neq \Pr(X \geq x) \end{aligned}$$

While both inequality signs produce the same continuous event, that is not the case for discrete distributions. Using the correct inequality sign is important for discrete random variables.

The survival function formulas for discrete and continuous random variables are:

$$\begin{aligned} S(x) &= 1 - F(x) \\ &= \Pr(X > x) \\ &= \sum_{t > x} p(t) && \text{(discrete)} \\ &= \int_x^\infty f(t)\,\mathrm{d}t && \text{(continuous)} \end{aligned} \tag{1.0.2.3}$$

If X is **continuous**, we can use Equation 1.0.2.2 and the relationship between $F(x)$ and $S(x)$ to derive a relationship between the PDF and survival function.

$$
\begin{aligned}
f(x) &= \frac{\mathrm{d}}{\mathrm{d}x}[1 - S(x)] \\
&= -\frac{\mathrm{d}}{\mathrm{d}x}S(x)
\end{aligned} \tag{1.0.2.4}
$$

In addition, all survival functions satisfy the following properties that mirror the CDF properties:

- $S(-\infty) = 1$
- $S(\infty) = 0$

Hazard Function

The *hazard function*, commonly known as the *hazard rate* or the *force of mortality* or the *failure rate function*, is the ratio of the PDF of a random variable to its survival function. The hazard function of a continuous random variable X is usually denoted by:

- $h_X(x)$
- $h(x)$

By definition,

$$
\begin{aligned}
h(x) &= \frac{f(x)}{S(x)} \\
&= \frac{-\frac{\mathrm{d}}{\mathrm{d}x}S(x)}{S(x)} \\
&= -\frac{\mathrm{d}}{\mathrm{d}x}[\ln S(x)]
\end{aligned} \tag{1.0.2.5}
$$

It can be interpreted as the PDF evaluated at x, adjusted by the likelihood that the random variable is greater than x. Thus, it measures the likelihood of the random variable at x by inflating the PDF as the random variable becomes less likely to exceed x.

Just like the PDF, $h(x)$ is not a probability. Therefore, a hazard function can exceed 1.

$$h(x) \geq 0$$

Integrating the hazard function produces the *cumulative hazard function*, which is often denoted as $H_X(x)$ or $H(x)$. Substituting Equation 1.0.2.5 into the cumulative hazard function definition will produce a useful relationship between $H(x)$ and $S(x)$:

$$\begin{aligned} H(x) &= \int_{-\infty}^{x} h(t)\,\mathrm{d}t \\ &= \int_{-\infty}^{x} -\frac{\mathrm{d}}{\mathrm{d}t}[\ln S(t)]\,\mathrm{d}t \\ &= -\ln S(x) \end{aligned}$$

Therefore,

$$S(x) = e^{-H(x)} \tag{1.0.2.6}$$

1.0.3 Percentiles and Mode

20m

Percentiles

The $100q^{\text{th}}$ *percentile* of a random variable X is the value π_q that satisfies **both** of these inequalities:

- $\Pr(X < \pi_q) \le q$
- $\Pr(X \le \pi_q) \ge q$

The definition above applies to both discrete and continuous distributions. This definition can be simplified if X is **continuous**. Recall that for a continuous random variable X, it is always true that $\Pr(X < \pi_q) = \Pr(X \le \pi_q)$. Therefore, the two inequalities can be restated as:

$$q \le \Pr(X \le \pi_q) \le q$$

This can be further restated as:

$$F(\pi_q) = \Pr(X \le \pi_q) = q$$

In other words, the $100q^{\text{th}}$ percentile of a continuous random variable X is the value π_q that satisfies $F(\pi_q) = q$.

There are special names for certain percentiles:

- The 25^{th} percentile is the 1^{st} quartile
- The 50^{th} percentile is the *median*, or the 2^{nd} quartile
- The 75^{th} percentile is the 3^{rd} quartile

CONTINUOUS EXAMPLE

Consider the following example:

A random variable X has the following probability distribution:

$$f(x) = 0.5x, \qquad 0 < x < 2$$

Determine the median and the 3rd quartile of X.

The median is the 50th percentile of X, which is $\pi_{0.5}$. The 3rd quartile is its 75th percentile, which is $\pi_{0.75}$.

Determine the CDF of X to compute its percentiles.

$$F(x) = \begin{cases} 0, & x \le 0 \\ 0.25x^2, & 0 < x < 2 \\ 1, & x \ge 2 \end{cases}$$

Since X is continuous, $\pi_{0.5}$ is the value of X such that $F(\pi_{0.5}) = 0.5$. Therefore,

$$\begin{aligned} F(\pi_{0.5}) &= 0.5 \\ 0.25(\pi_{0.5})^2 &= 0.5 \\ \pi_{0.5} &= \boldsymbol{\sqrt{2}} \end{aligned}$$

Similarly, the CDF evaluated at $x = \pi_{0.75}$ must produce a probability of 0.75. Therefore,

$$\begin{aligned} F(\pi_{0.75}) &= 0.75 \\ 0.25(\pi_{0.75})^2 &= 0.75 \end{aligned}$$

$$\pi_{0.75} = \sqrt{\mathbf{3}}$$

DISCRETE EXAMPLE

Consider another example:

A random variable X has the following probability distribution:

$$p(x) = \begin{cases} 0.40, & x = 1 \\ 0.20, & x = 2 \\ 0.15, & x = 5 \\ 0.25, & x = 8 \\ 0, & \text{otherwise} \end{cases}$$

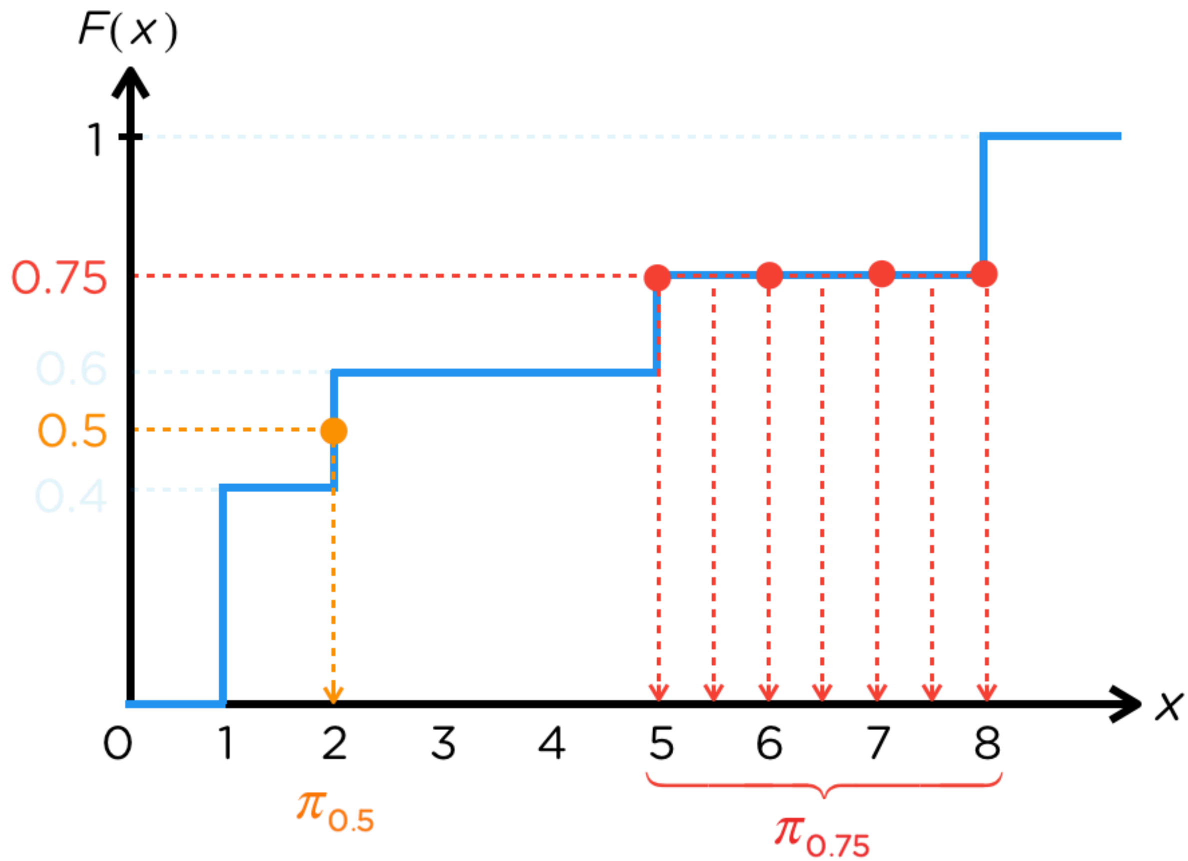
Determine the median and the 3rd quartile of X.

Since X is discrete in this example, the shortcut formula $F(\pi_q) = q$ cannot be used. The full definition of a percentile must be used instead.

Start by sketching a graph of the CDF. This is represented by the blue line in the figure below. To determine the $100q^{\text{th}}$ percentile, draw a horizontal line $F(x) = q$.

- In most cases, this line will intersect the CDF at one point. The x-coordinate of the point of intersection is the $100q^{\text{th}}$ percentile.
- In the case of multiple points of intersection, the x-coordinates of all points of intersection will count as the percentile, i.e. there will be multiple points corresponding to the same percentile.

The following graph illustrates $\pi_{0.5}$ and $\pi_{0.75}$.



The line $F(x) = 0.5$ intersects the CDF at $x = 2$. Therefore, $\pi_{0.5} = \mathbf{2}$.

We can verify that our answer is correct by checking if the percentile satisfies the two inequalities.

- $\Pr(X \le 2) = 0.6 \ge 0.5$
- $\Pr(X < 2) = \Pr(X \le 1) = 0.4 \le 0.5$

This confirms the median of X is 2.

Notice that for any q where $0.4 \le q \le 0.6$, the line $F(x) = q$ will intersect the CDF at $x = 2$. Therefore, 2 is not only the 50th percentile; it is any percentile of X between the 40th and the 60th.

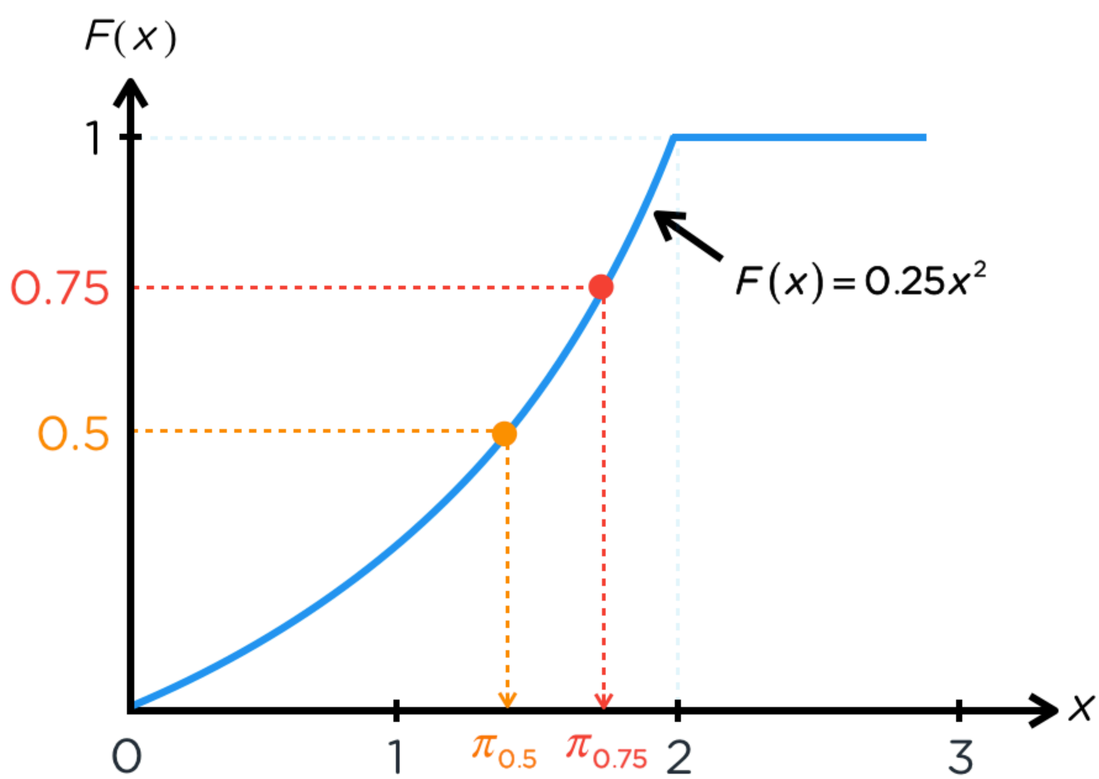
The line $F(x) = 0.75$ does not intersect the CDF at just one point. It intersects the CDF at every point from $x = 5$ to $x = 8$. That means every value in the interval from 5 to 8 (including 5 and 8) is the 75$^{\text{th}}$ percentile. This can be confirmed using the percentile definition, as both inequalities will be satisfied for **any** $\pi_{0.75}$ where $5 \leq \pi_{0.75} \leq 8$.

In short, determining the percentiles of a discrete distribution can be tricky.

- Multiple percentiles can have the same value.
- Multiple values can correspond to the same percentile.

Based on the examples explained above, several conclusions can be made:

1. If X is **continuous**, it is easy to compute its percentiles. This is because the types of continuous random variables that we care about in this exam will have a CDF that is strictly increasing (except when it equals 0 or 1). This can be seen in the CDF plot below, based on the continuous example above.



For $0 < F(x) < 1$, the CDF is a one-to-one function of x. This implies it has an inverse function which produces unique percentiles of X.

$$\pi_q = F^{-1}(q),\ \text{for } 0 < q < 1$$

2. If X is **discrete**, its CDF plot will be a step graph, as shown in the discrete example above. Since the CDF is no longer a one-to-one function, the percentiles of X have to be calculated using the full definition of percentile stated at the beginning of this section.

3. For both discrete and continuous distributions, use the **CDF** to calculate percentiles.

Example 1.0.3.1

Suppose X is a random variable that has the following density function:

$$f(x) = \frac{1}{5}\, e^{-x\,/\,5}, \qquad x > 0$$

Determine the 60th percentile of X.

Solution

Since X is continuous, solve for $\pi_{0.6}$ such that $F(\pi_{0.6}) = 0.6$.

$$\begin{aligned} F(\pi_{0.6}) &= \int_0^{\pi_{0.6}} \frac{1}{5}\, e^{-x\,/\,5}\, \mathrm{d}x \\ &= \left[-e^{-x\,/\,5}\right]_0^{\pi_{0.6}} \\ &= 1 - e^{-\pi_{0.6}\,/\,5} \end{aligned}$$

$$0.6 = 1 - e^{-\pi_{0.6}/5}$$

$$e^{-\pi_{0.6}/5} = 0.4$$

$$\begin{aligned}\pi_{0.6} &= -5\ln(0.4)\\ &= \mathbf{4.5815}\end{aligned}$$

■

Mode

The *mode* is the value of the random variable that maximizes the PMF or the PDF. In other words, the mode is the critical point of the **global maximum** of the PMF/PDF.

For **continuous** random variables, the mode is easy to calculate. Maximizing the PDF means calculating the first derivative of $f(x)$ and equating it to 0. Then, solve for x, which is the mode.

$$f'(x) = 0 \quad \Rightarrow \quad x = \text{mode}$$

Coach's Remarks

The approach above computes the critical points of local maxima/minima of the probability function, but not the global maximum. To locate the global maximum, the second derivative test might be required; this is shown below in Example 1.0.3.2.

Also, since it is possible that the mode is an endpoint of the domain, the endpoints might need to be evaluated.

However, for most exam questions, the simple approach given above will be sufficient for continuous distributions.

While most distributions only have one mode, it is possible for a random variable to have multiple modes. This occurs when more than one value of the random variable produces the maximum PDF/PMF value.

In summary, for both discrete and continuous distributions, calculate the mode(s) by maximizing the **PMF**/**PDF**.

Example 1.0.3.2

A continuous random variable X has the following PDF:

$$f(x) = 0.75\left[1 - (x-1)^2\right], \qquad 0 < x < 2$$

Determine the mode of X.

Solution

Determine the derivative of $f(x)$.

$$\begin{aligned} f'(x) &= \frac{\mathrm{d}}{\mathrm{d}x}\left(0.75\left[1 - (x-1)^2\right]\right) \\ &= -1.5(x-1) \end{aligned}$$

Equate the derivative to 0, and solve for x.

$$-1.5(x-1)=0$$

$$x=\mathbf{1}$$

Thus, the mode of X is 1.

■

Coach's Remarks

The following step is usually not needed for the exam, but if you would like to verify that $x=1$ is a local maximum (and not a local minimum), you can apply the second derivative test.

$$\begin{aligned} f''(x) &= \frac{\mathrm{d}}{\mathrm{d}x}[-1.5(x-1)] \\ &= -1.5 < 0 \end{aligned}$$

A negative second derivative indicates the critical point is a local maximum. This confirms the mode of X is 1.

1.0.4 Moments

30m

The *expected value*, or *mean*, of a random variable is its average value. In general, the expected value of a random variable X can be calculated using one of the two formulas below:

$$\begin{aligned} \mathrm{E}[X] &= \sum_{\text{all } x} x \cdot p(x) && \text{(discrete)} \\ &= \int_{-\infty}^{\infty} x \cdot f(x)\,\mathrm{d}x && \text{(continuous)} \end{aligned}$$

The formula for $\mathrm{E}[X]$ can be generalized by replacing the random variable X with some function $g(X)$. Then, X would be a special case of $g(X)$. The expected value formulas now become:

$$\begin{aligned} \mathrm{E}[g(X)] &= \sum_{\text{all } x} g(x) \cdot p(x) && \text{(discrete)} \\ &= \int_{-\infty}^{\infty} g(x) \cdot f(x)\,\mathrm{d}x && \text{(continuous)} \end{aligned} \tag{1.0.4.1}$$

There is an alternative method to calculate the expected values. This method only works if the random variable X is **non-negative**. For a function $g(X)$ where $g(0) = 0$, we have

$$\mathrm{E}[g(X)] = \int_{0}^{\infty} g'(x) \cdot S(x)\,\mathrm{d}x \tag{1.0.4.2}$$

Note that the lower limit of the integral should **always** be 0, regardless of the domain of X. The derivation of this method is provided in the appendix at the end of this section.

Coach's Remarks

Equation 1.0.4.2 above should only be used for **continuous** variables. Under certain conditions, the same formula also works for discrete variables. However, to keep it simple, **avoid** this method for discrete variables.

Expected values have three simple properties that are worth remembering. For a constant c,

1. $\mathrm{E}[c] = c$
2. $\mathrm{E}[c \cdot g(X)] = c \cdot \mathrm{E}[g(X)]$
3. $\mathrm{E}[g_1(X) + g_2(X) + \ldots + g_k(X)] = \mathrm{E}[g_1(X)] + \mathrm{E}[g_2(X)] + \ldots + \mathrm{E}[g_k(X)]$

Certain expected values are also known as *moments*. Generally, moments can be categorized into two types: raw moments and central moments.

Raw Moments

The k^{th} *raw moment* of X, or the k^{th} *moment* of X for short, is defined as

$$\mu'_k = \mathrm{E}\left[X^k\right]$$

Note that the 1st raw moment of X is the mean. For the sake of simplicity, the mean is usually denoted as μ.

$$\mu'_1 = \mathrm{E}[X] = \mu$$

Central Moments

The k^{th} *central moment* of X is defined as

$$\mu_k = \mathrm{E}\Big[(X-\mu)^k\Big]$$

where μ is the mean of X.

Coach's Remarks

It is important to distinguish between the notations for raw moments, central moments, and the mean.

- μ'_k (with prime symbol and subscript k): k^{th} raw moment
- μ_k (with subscript k): k^{th} central moment
- μ: mean, or alternative notation for the 1$^{\text{st}}$ raw moment

Time for a pop quiz! What is μ_1, the 1$^{\text{st}}$ central moment? As it turns out, it **always** equals 0.

$$\begin{aligned}\mu_1 &= \mathrm{E}\Big[(X-\mu)^1\Big] \\ &= \mathrm{E}[X] - \mu \\ &= \mu - \mu \\ &= 0\end{aligned}$$

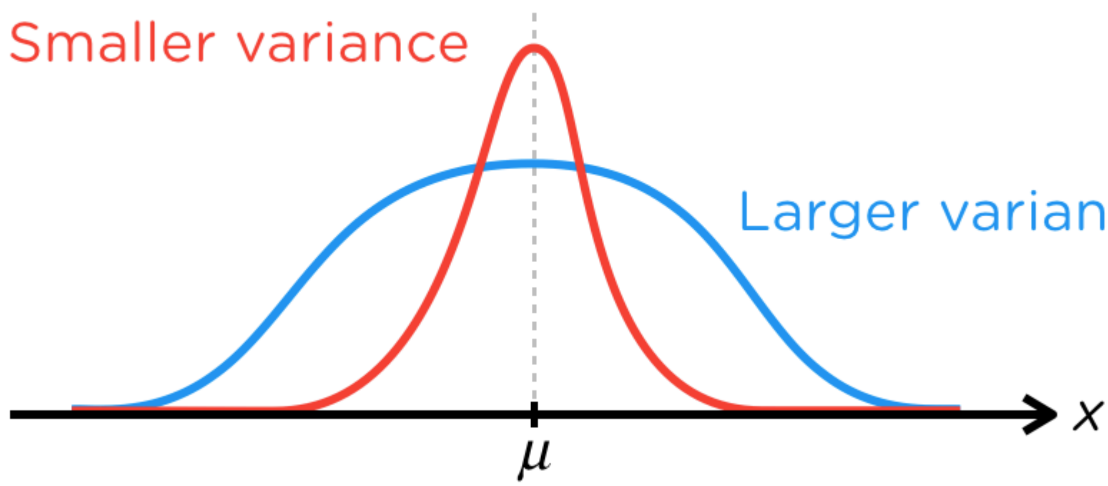
Variance

The most well-known central moment is the 2$^{\text{nd}}$ central moment, otherwise known as the *variance*. Notation-wise, the variance of X is $\mathrm{Var}[X]$, or σ^2.

$$\mu_2 = \mathrm{E}\left[(X-\mu)^2\right] = \mathrm{Var}[X] = \sigma^2$$

Variance measures how much the observations deviate from their mean. Mathematically, variance is the expected squared difference between the random variable and its mean.

- A **larger** variance means the observations are more **dispersed**, or more spread out from the mean.
- A **smaller** variance means the observations are **closer** to the mean.



Let's expand the expression for the 2nd central moment of X.

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

This gives us an alternative formula to calculate the variance. Since the 1st and 2nd raw moments are often easier to calculate, the formula above is our default variance formula. Further generalizing the formula using a function $g(X)$, we have

$$\mathrm{Var}[g(X)] = \mathrm{E}\left[g(X)^2\right] - \mathrm{E}[g(X)]^2 \tag{1.0.4.3}$$

Variance has four important properties that are useful for this exam. For constants a, b, and c,

1. $\mathrm{Var}[c] = 0$
2. $\mathrm{Var}[X + c] = \mathrm{Var}[X]$
3. $\mathrm{Var}[cX] = c^2\mathrm{Var}[X]$
4. $\mathrm{Var}[aX + bY] = a^2\mathrm{Var}[X] + b^2\mathrm{Var}[Y] + 2ab\mathrm{Cov}[X,\, Y]$, where $\mathrm{Cov}[X,\, Y] = \mathrm{E}[XY] - \mathrm{E}[X] \cdot \mathrm{E}[Y]$

The square root of the variance is the *standard deviation*, which is often denoted by σ. In the case of multiple random variables, subscripts will be added to σ to distinguish between the random variables.

$$\sigma = \sqrt{\mathrm{Var}[X]} \tag{1.0.4.4}$$

The *coefficient of variation* is the ratio of the standard deviation to the mean. In other words, it calculates the standard deviation per unit of mean.

$$CV = \frac{\sigma}{\mu} \tag{1.0.4.5}$$

Skewness

The 3rd central moment can be used to calculate the *skewness* of a distribution, which equals the 3rd central moment divided by the standard deviation cubed.

$$\text{Skewness} = \frac{\mu_3}{\sigma^3} \tag{1.0.4.6}$$

It is usually more practical to compute the numerator, μ_3, by expanding the expression for

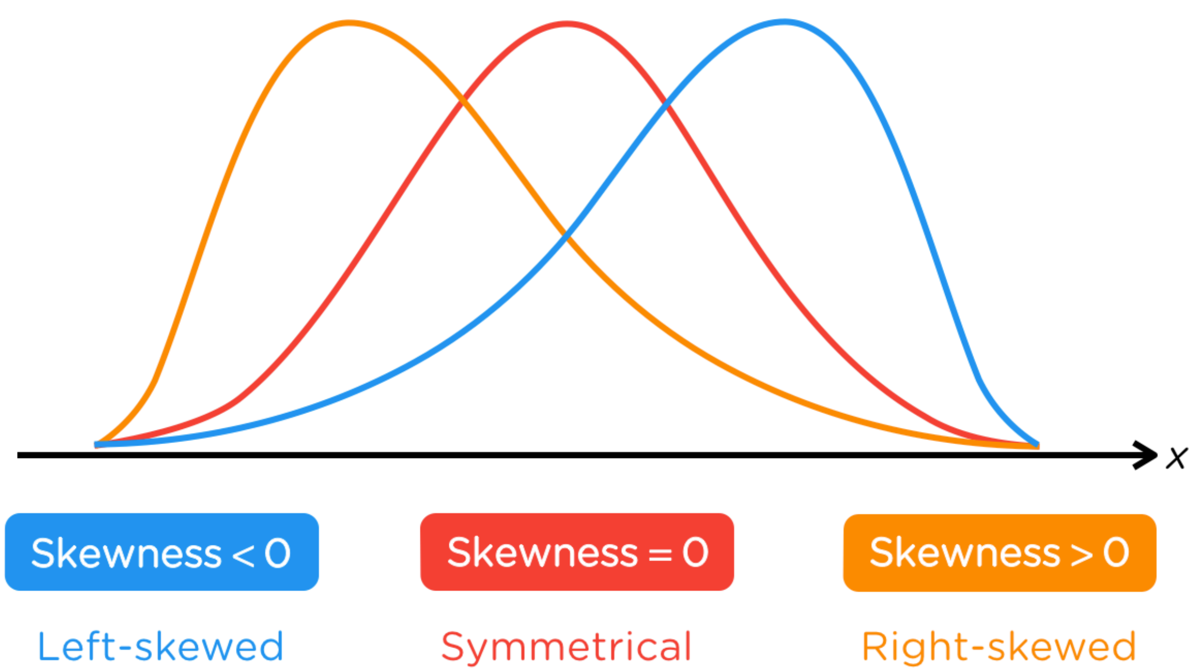
It is usually more practical to compute the numerator, μ_3, by expanding the expression for the 3rd central moment, similar to how we derived our default variance formula above.

$$\mu_3 = \mu_3' - 3\mu_2'\mu + 2\mu^3$$

The complete expansion can be found in the appendix at the end of this section. You should not memorize this equation. Instead, learn how to express a central moment in terms of raw moments.

Skewness measures how symmetric a distribution is relative to its mean.

- **Zero skewness** means the distribution is **perfectly symmetrical**. One popular example is the normal distribution.
- **Positive skewness** means the distribution is **right-skewed**, suggesting a longer right-tail. This implies smaller values are more likely to occur than larger values.
- **Negative skewness** means the distribution is **left-skewed**, suggesting a longer left-tail. This implies larger values are more likely to occur than smaller values.



Coach's Remarks

It is easy to confuse positive and negative skewness. Here is one way to help you remember.

- If a distribution has positive skewness, the endpoint of its longer tail will point towards the positive direction of the x-axis, i.e. point to the **right**. Therefore, "positively-skewed" also means "right-skewed".
- If a distribution has negative skewness, the endpoint of its longer tail will point towards the negative direction of the x-axis, i.e. point to the **left**. Therefore, "negatively-skewed" also means "left-skewed".

Kurtosis

The *kurtosis* of a distribution is the 4th central moment divided by the standard deviation to the 4th power.

$$\text{Kurtosis} = \frac{\mu_4}{\sigma^4} \qquad (1.0.4.7)$$

The numerator, μ_4, can be expressed using raw moments:

$$\mu_4 = \mu_4' - 4\mu_3'\mu + 6\mu_2'\mu^2 - 3\mu^4$$

The complete expansion is provided in the appendix at the end of this section.

Kurtosis measures the flatness of a distribution. It suggests the likelihood for a distribution to produce extreme values, or outliers. Here are a few kurtosis facts:

- The normal distribution has a kurtosis of 3.
- A distribution with kurtosis greater than 3 is more likely to produce outliers than the normal distribution, and vice versa.

Coach's Remarks

While it might not be directly tested, it is good to know that the coefficient of variation, skewness, and kurtosis are all **scale invariant**. In other words, if a random variable is multiplied by a positive factor, these three quantities will remain unchanged.

Example 1.0.4.1

A random variable X has the following density function:

$$f(x) = \frac{x}{8}, \qquad 0 \leq x \leq 4$$

Calculate

1. the coefficient of variation of X.
2. the skewness of X.
3. the kurtosis of X.

Solution to (1)

Calculate the coefficient of variation as follows:

$$
\begin{aligned}
\mathrm{E}[X] &= \int_0^4 x \cdot \frac{x}{8}\,\mathrm{d}x \\
&= \left[\frac{x^3}{3(8)}\right]_0^4 \\
&= \frac{8}{3}
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{E}\left[X^2\right] &= \int_0^4 x^2 \cdot \frac{x}{8}\,\mathrm{d}x \\
&= \left[\frac{x^4}{4(8)}\right]_0^4 \\
&= 8
\end{aligned}
$$

$$
\mathrm{Var}[X] = 8 - \left(\frac{8}{3}\right)^2 = \frac{8}{9}
$$

$$
\begin{aligned}
CV &= \frac{\sqrt{\mathrm{Var}[X]}}{\mathrm{E}[X]} \\
&= \frac{\sqrt{8/9}}{8/3} \\
&= \mathbf{0.3536}
\end{aligned}
$$

$\blacksquare$

Solution to (2)

All quantities needed to calculate the skewness have been calculated above, except the 3rd central moment.

$$\begin{aligned} \mathrm{E}\left[X^3\right] &= \int_0^4 x^3 \cdot \frac{x}{8}\,\mathrm{d}x \\ &= \left[\frac{x^5}{5(8)}\right]_0^4 \\ &= 25.6 \end{aligned}$$

$$\begin{aligned} \mu_3 &= \mu_3' - 3\mu_2'\mu + 2\mu^3 \\ &= 25.6 - 3(8)\left(\frac{8}{3}\right) + 2\left(\frac{8}{3}\right)^3 \\ &= -0.4741 \end{aligned}$$

$$\begin{aligned} \text{Skewness} &= \frac{-0.4741}{\left(\sqrt{8/9}\right)^3} \\ &= \mathbf{-0.5657} \end{aligned}$$

Since the skewness is negative, X is left-skewed.

■

Solution to (3)

To determine the kurtosis, calculate the 4th moment. All other required quantities have been calculated above.

$$\begin{aligned}\mathrm{E}[X^4] &= \int_0^4 x^4 \cdot \frac{x}{8}\,\mathrm{d}x \\ &= \left[\frac{x^6}{6(8)}\right]_0^4 \\ &= 85.333\end{aligned}$$

$$\begin{aligned}\mu_4 &= \mu_4' - 4\mu_3'\mu + 6\mu_2'\mu^2 - 3\mu^4 \\ &= 85.333 - 4(25.6)\left(\frac{8}{3}\right) + 6(8)\left(\frac{8}{3}\right)^2 - 3\left(\frac{8}{3}\right)^4 \\ &= 1.8963\end{aligned}$$

$$\begin{aligned}\text{Kurtosis} &= \frac{1.8963}{\left(\sqrt{8/9}\right)^4} \\ &= \mathbf{2.4}\end{aligned}$$

Since the kurtosis is less than 3, X is less likely to produce outliers than the normal distribution.

■

Example 1.0.4.2

The annual claim size of an insured has the following probability distribution:

Claim size	Probability
100	0.3
200	0.2
250	0.5

Calculate the skewness of the annual claim size.

Solution

Let X represent the annual claim size.

$$\text{Skewness} = \frac{\mu_3}{\sigma^3}$$

$$\begin{aligned}\mu &= \mathrm{E}[X] \\ &= 0.3(100) + 0.2(200) + 0.5(250) \\ &= 195\end{aligned}$$

$$\begin{aligned}\sigma^2 &= \mathrm{E}\left[(X-195)^2\right] \\ &= 0.3(100-195)^2 + 0.2(200-195)^2 + 0.5(250-195)^2 \\ &= 4{,}225\end{aligned}$$

$$\begin{aligned}\mu_3 &= \mathrm{E}\left[(X-195)^3\right] \\ &= 0.3(100-195)^3 + 0.2(200-195)^3 + 0.5(250-195)^3\end{aligned}$$

$$
\begin{aligned}
&= 0.3(100 - 195)^3 + 0.2(200 - 195)^3 + 0.5(250 - 195)^3 \\
&= -174{,}000
\end{aligned}
$$

$$
\begin{aligned}
\text{Skewness} &= \frac{\mu_3}{\sigma^3} \\
&= \frac{-174{,}000}{\left(\sqrt{4{,}225}\right)^3} \\
&= \mathbf{-0.6336}
\end{aligned}
$$

$\blacksquare$

Coach's Remarks

Notice that in this example, it is easier to calculate the 3rd central moment directly rather than in terms of raw moments. This is because X is a discrete random variable with only 3 possible values.

In addition, the final answer of -0.6336 makes sense because a negative skewness implies a left-skewed distribution, which suggests that larger values of the random variable are more likely to occur. This agrees with the annual claim size distribution which states that the largest claim size, 250, has the largest probability. Although knowing this concept does not help us determine the final answer, it can help us check our answer for reasonableness and catch mistakes.

1.0.5 Moment and Probability Generating Functions

15m

Moment Generating Function

The *moment generating function (MGF)* of a random variable X is denoted as $M_X(z)$. It is defined as

$$M_X(z) = \mathrm{E}\left[e^{zX}\right] \tag{1.0.5.1}$$

The subscript X may be dropped when the random variable is assumed or unspecified. Since the MGF is an expected value, Equation 1.0.4.1 can be used to determine $M_X(z)$.

$$\begin{aligned} M_X(z) &= \mathrm{E}\left[e^{zX}\right] \\ &= \sum_{\text{all } x} e^{zx} \cdot p(x) && \text{(discrete)} \\ &= \int_{-\infty}^{\infty} e^{zx} \cdot f(x)\,\mathrm{d}x && \text{(continuous)} \end{aligned}$$

As the name implies, MGFs generate **moments**. To generate the n^{th} moment, determine the n^{th} derivative of the MGF with respect to z, then evaluate the derivative at $z = 0$.

$$M_X^{(n)}(0) = \mathrm{E}[X^n] \tag{1.0.5.2}$$

Note that the superscript (n) in Equation 1.0.5.2 above denotes the n^{th} derivative.

For instance,

$$M_X^{(1)}(0) = \frac{\mathrm{d}}{\mathrm{d}z} M_X(z)\bigg|_{z=0} = \mathrm{E}[X]$$

$$M_X^{(2)}(0) = \frac{\mathrm{d}^2}{\mathrm{d}z^2} M_X(z)\bigg|_{z=0} = \mathrm{E}\left[X^2\right]$$

For any random variable (discrete or continuous), the MGF itself evaluated at $z = 0$ must equal 1.

$$M_X(0) = \mathrm{E}\left[e^{(0)X}\right] = E[1] = 1$$

Probability Generating Function

The *probability generating function (PGF)* of a random variable X is denoted as $P_X(z)$. It is defined as

$$P_X(z) = \mathrm{E}\left[z^X\right] \tag{1.0.5.3}$$

Again, the subscript X can be dropped if the random variable is assumed or unspecified.

Coach's Remarks

Do not confuse the PMF notation with the PGF notation. The PMF is denoted as a lowercase p, while the PGF is denoted as a capital P. Remember that they represent very different quantities despite looking similar. The former is a **probability**; the latter is the **expected value** of a specific function of X.

- PMF: $p_X(x) = \Pr(X = x)$
- PGF: $P_X(z) = \mathrm{E}\left[z^X\right]$

Similar to the MGF, Equation 1.0.4.1 can be applied to determine the PGF of a random variable.

The PGF is named such because it can be used to generate probabilities for a discrete distribution. To determine the probability at n, first determine the n^{th} derivative of the PGF with respect to z, evaluate the derivative at $z=0$, and divide the result by $n!$.

$$\frac{P_X^{(n)}(0)}{n!} = p(n)$$

However, this is not an efficient way to calculate probabilities, especially for common discrete distributions with PMF formulas given in the exam table. Also, the n^{th} derivative of the PGF can be tedious to determine. Therefore, you will not likely need to use the formula above for most exam questions.

Regardless, the PGF has another useful property:

$$P_X^{(n)}(1) = \mathrm{E}[X(X-1)\ldots(X-n+1)] \qquad (1.0.5.4)$$

Note that the n^{th} derivative of the PGF is evaluated at 1, **not** 0, for this formula.

The "$\ldots$" in the formula above represents a continuing pattern following the first 2 terms for a given value of n. In the product, the second term is 1 less than the first term, the third term is 1 less than the second term, and so on. This continues until the term $X-n+1$.

For instance,

- When $n=1$, the product starts from X to $X-1+1=X$. Therefore,

$$P_X^{(1)}(1) = \frac{\mathrm{d}}{\mathrm{d}z} P_X(z)\bigg|_{z=1} = \mathrm{E}[X]$$

- When $n=2$, the product starts from X to $X-2+1=X-1$. Therefore,

$$P_X^{(2)}(1) = \frac{\mathrm{d}^2}{\mathrm{d}z^2} P_X(z) \bigg|_{z=1} = \mathrm{E}[X(X-1)]$$

- When $n = 3$, the product starts from X to $X - 3 + 1 = X - 2$. Therefore,

$$P_X^{(3)}(1) = \frac{\mathrm{d}^3}{\mathrm{d}z^3} P_X(z) \bigg|_{z=1} = \mathrm{E}[X(X-1)(X-2)]$$

Coach's Remarks

As the names imply, the MGF and the PGF are **functions**. These functions are in terms of z, **not** x. While x represents the value of a random variable, z has no direct meaning or connection with the random variable. It is just a variable for the generating functions, which can be substituted with values to produce useful results, such as probabilities and moments, as discussed above.

There is a one-to-one relationship between a random variable and its MGF or PGF. Said differently, if two random variables have the same MGF or PGF, they **must** also have the same distribution. This property is extremely useful when proving two distributions are the same. In many cases, it is easier to prove the equality of two MGFs/PGFs than to prove the equality of two probability functions.

Example 1.0.5.1

For a random variable X, you are given:

$$P_X(z) = e^{4(z-1)}$$

Calculate

1. the expected value of 3^X.
2. the variance of X.

Solution to (1)

Recall a PGF is defined as $P_X(z) = \mathrm{E}\left[z^X\right]$. Therefore, to calculate the expected value of 3^X, evaluate the PGF at $z = 3$.

$$\begin{aligned}\mathrm{E}\left[3^X\right] &= P_X(3) \\ &= e^{4(3-1)} \\ &= \mathbf{2{,}980.96}\end{aligned}$$

■

Solution to (2)

To determine the variance of X:

Compute the first and second moments of X using the PGF property.

$$P_X^{(1)}(z) = \frac{\mathrm{d}}{\mathrm{d}z} P_X(z) = 4e^{4(z-1)}$$

$$\begin{aligned}\mathrm{E}[X] &= P_X^{(1)}(1) \\ &= 4e^{4(1-1)} \\ &= 4\end{aligned}$$

$$P_X^{(2)}(z) = \frac{\mathrm{d}}{\mathrm{d}z} P_X^{(1)}(z) = 16e^{4(z-1)}$$

$$\begin{aligned}\mathrm{E}[X(X-1)] &= P_X^{(2)}(1) \\ &= 16e^{4(1-1)} \\ &= 16\end{aligned}$$

$$\mathrm{E}\left[X^2\right] - \mathrm{E}[X] = 16$$

$$\begin{aligned}\mathrm{E}\left[X^2\right] &= 4 + 16 \\ &= 20\end{aligned}$$

$$\begin{aligned}\mathrm{Var}[X] &= \mathrm{E}\left[X^2\right] - \mathrm{E}[X]^2 \\ &= 20 - 4^2 \\ &= \mathbf{4}\end{aligned}$$

■

Coach's Remarks

The PGF in this question belongs to the Poisson distribution. Recognizing the distribution of a random variable from its PGF or MGF can be a huge time-saver. We could have found the mean and variance within seconds, just by observing the PGF.

1.0.6 Conditional Distributions and Independence

30m

Conditional Distribution

A *conditional probability* is the probability that an event occurs, **given** that another event has occurred. In probability notation, the probability that event A occurs given that event B has occurred is:

$$\Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)}$$

The numerator can be converted to:

$$\Pr(A \cap B) = \Pr(B \mid A) \cdot \Pr(A)$$

Combining the first and second equations above produces *Bayes' Theorem*.

$$\Pr(A \mid B) = \frac{\Pr(B \mid A) \Pr(A)}{\Pr(B)} \tag{1.0.6.1}$$

The concept of conditional probabilities can be extended to random variables to derive *conditional distributions*. If a random variable X is conditioned on values between j and k, then its conditional PDF is

$$f_{X \mid j < X < k}(x) = \frac{f_X(x)}{\Pr(j < X < k)}, \text{ where } j < x < k \tag{1.0.6.2}$$

Using the conditional PDF, the following can be calculated:

- $\Pr(X \leq x \mid j < X < k) = \int_j^x f_{X \mid j < X < k}(t) \, \mathrm{d}t$

- $\mathrm{E}[X \mid j < X < k] = \int_j^k x \cdot f_{X \mid j < X < k}(x)\,\mathrm{d}x$

For a discrete random variable X, replace the integrals and the PDFs in the equations above with sums and PMFs, respectively.

A random variable can also be conditioned on **another** random variable. For example, if a continuous random variable X is conditioned on another random variable Y, then this new conditional random variable is denoted $(X \mid Y)$. Its conditional probability and conditional mean can be determined as follows:

- Conditional probability: $\Pr(X \le c \mid Y = y) = \int_{-\infty}^{c} f_{X|Y}(x \mid y)\,\mathrm{d}x$
- Conditional mean: $\mathrm{E}[X \mid Y = y] = \int_{-\infty}^{\infty} x \cdot f_{X|Y}(x \mid y)\,\mathrm{d}x$

Again, if $(X \mid Y)$ is discrete, replace the integrals and the PDFs in the equations above with sums and PMFs, respectively.

The concept of conditional distributions is very important in relation to the three theorems below:

The Law of Total Probability

The *Law of Total Probability* states that if events A_1, A_2, ..., A_n are mutually exclusive and form a partition of a sample space, then

$$\begin{aligned}\Pr(B) &= \sum_{i=1}^{n} \Pr(B \cap A_i) \\ &= \sum_{i=1}^{n} \left[\Pr(B \,|\, A_i) \cdot \Pr(A_i)\right] \qquad (1.0.6.3)\end{aligned}$$

The same concept can be applied to determine the unconditional PMF of a discrete random variable X from a conditional distribution. In general,

$$\Pr(X = x) = \mathrm{E}_Y[\Pr(X = x \mid Y)] \tag{1.0.6.4}$$

Loosely speaking, the random variable Y takes the place of A_i in Equation 1.0.6.3. The subscript Y indicates that Y is the random variable for the "averaging". In other words,

- if Y is **discrete**, then

$$\Pr(X = x) = \sum_{\text{all } y} [\Pr(X = x \mid Y = y) \cdot \Pr(Y = y)]$$

- if Y is **continuous**, then

$$\Pr(X = x) = \int_{-\infty}^{\infty} \Pr(X = x \mid Y = y) \cdot f_Y(y)\,\mathrm{d}y$$

Note that X is **discrete** in both cases above. While this formula can be generalized to determine the unconditional PDF of a continuous random variable, the discrete form is more commonly used.

In addition, the Law of Total Probability can be used to determine the unconditional CDF and survival function of a random variable.

$$F_X(x) = \mathrm{E}_Y[\Pr(X \le x \mid Y)]$$

$$S_X(x) = \mathrm{E}_Y[\Pr(X > x \mid Y)]$$

The Law of Total Expectation

The *Law of Total Expectation* states that:

$$\mathrm{E}_X[X] = \mathrm{E}_Y[\mathrm{E}_X[X \mid Y]] \tag{1.0.6.5}$$

Keep in mind that the inner expectation, $\mathrm{E}_X[X \mid Y]$, is a function of Y, not a constant. Remember that $(X \mid Y)$ is in terms of both X and Y. If Y changes, the conditional distribution will change, and so will the conditional mean.

More generally, for any function $g(X)$,

$$\mathrm{E}_X[g(X)] = \mathrm{E}_Y[\mathrm{E}_X[g(X) \mid Y]]$$

Since the MGF and PGF are also expected values of some functions of X, the Law of Total Expectation can be applied to them, too.

The Law of Total Variance

The *Law of Total Variance* states that:

$$\mathrm{Var}_X[X] = \mathrm{E}_Y[\mathrm{Var}_X[X \mid Y]] + \mathrm{Var}_Y[\mathrm{E}_X[X \mid Y]] \tag{1.0.6.6}$$

The proof of this formula is lengthy, so it is included in the appendix at the end of this section.

Coach's Remarks

Many students assume that the unconditional variance equals the average of the conditional variances. This is incorrect because the variance of the conditional mean also needs to be included in the calculation.

$$\mathrm{Var}_X[X] = \mathrm{E}_Y[\mathrm{Var}_X[X \mid Y]] + \mathrm{Var}_Y[\mathrm{E}_X[X \mid Y]]$$

$$\neq \mathrm{E}_Y[\mathrm{Var}_X[X \mid Y]]$$

Let's apply these theorems in the examples below.

Example 1.0.6.1

For a health insurance policy, insureds are classified into two classes: smoker and non-smoker. Of the pool of insureds, 40% are smokers and 60% are non-smokers.

The means and variances of the annual claim size for the insureds are:

Class	Mean	Variance
Smoker	600	1,000
Non-smoker	450	800

Determine

1. the expected annual claim size of a randomly selected insured.
2. the variance of the annual claim size of a randomly selected insured.

Solution to (1)

Let X be the annual claim size.

The question asks for the unconditional mean and variance. The means and variances given in the tables are conditional on class.

Class	$\mathrm{E}[X \mid \text{Class}]$	$\mathrm{Var}[X \mid \text{Class}]$	$\Pr(\text{Class})$
Smoker	600	1,000	0.4
Non-smoker	450	800	0.6

Because we are calculating the unconditional mean and variance from conditional means and variances, Equations 1.0.6.5 and 1.0.6.6 should be used.

To determine the expected annual claim size, apply the Law of Total Expectation (Equation 1.0.6.5).

$$\begin{aligned}\mathrm{E}[X] &= \mathrm{E}[\mathrm{E}[X \mid \text{Class}]] \\ &= \sum_{\text{all classes}} \mathrm{E}[X \mid \text{Class}] \cdot \Pr(\text{Class}) \\ &= 600(0.4) + 450(0.6) \\ &= \mathbf{510}\end{aligned}$$

■

Solution to (2)

To determine the variance of the annual claim size, apply the Law of Total Variance (Equation 1.0.6.6).

$$\mathrm{Var}[X] = \mathrm{E}[\mathrm{Var}[X \mid \text{Class}]] + \mathrm{Var}[\mathrm{E}[X \mid \text{Class}]]$$

Determine each component individually.

Calculate $\mathrm{E}[\mathrm{Var}[X \mid \text{Class}]]$

- Calculate $\text{E}[\text{Var}[X \mid \text{Class}]]$.

$$\begin{aligned}\text{E}[\text{Var}[X \mid \text{Class}]] &= \sum_{\text{all classes}} \text{Var}[X \mid \text{Class}] \cdot \Pr(\text{Class}) \\ &= 1{,}000(0.4) + 800(0.6) \\ &= 880\end{aligned}$$

- Calculate $\text{Var}[\text{E}[X \mid \text{Class}]]$.

$$\text{Var}[\text{E}[X \mid \text{Class}]] = \text{E}\left[\text{E}[X \mid \text{Class}]^2\right] - \text{E}[\text{E}[X \mid \text{Class}]]^2$$

$$\begin{aligned}\text{E}\left[\text{E}[X \mid \text{Class}]^2\right] &= \sum_{\text{all classes}} \text{E}[X \mid \text{Class}]^2 \cdot \Pr(\text{Class}) \\ &= 600^2(0.4) + 450^2(0.6) \\ &= 265{,}500\end{aligned}$$

$$\text{E}[\text{E}[X \mid \text{Class}]] = 510$$

$$\begin{aligned}\text{Var}[\text{E}[X \mid \text{Class}]] &= 265{,}500 - 510^2 \\ &= 5{,}400\end{aligned}$$

Therefore,

$$\begin{aligned}\text{Var}[X] &= 880 + 5{,}400 \\ &= \mathbf{6{,}280}\end{aligned}$$

■

Coach's Remarks

Instead of using the Law of Total Variance, $\mathrm{Var}[X]$ can also be calculated using $\mathrm{E}\left[X^2\right] - \mathrm{E}[X]^2$, where the first and second moments of X are computed using the Law of Total Expectation.

$$\mathrm{E}\left[X^k\right] = \mathrm{E}\left[\mathrm{E}\left[X^k \mid \text{Class}\right]\right]$$

Both methods will produce the same result.

Example 1.0.6.2

The claim size of a homeowner's policy is distributed as follows:

Claim Size	Probability
50	0.3
100	$0.7 - q$
150	q

q varies by insured. For each insured, q has a distribution with the following density function:

$$f(q) = \frac{q}{0.175}, \qquad 0.1 < q < 0.6$$

Calculate

1. the expected claim size of a randomly selected insured.
2. the probability that a randomly selected insured has a claim of 100.

Solution to (1)

Let X be the claim size.

While not stated explicitly, the table in the question actually describes the distribution of $(X \mid q)$, **not** X. This is because the probabilities of the claim size depend on the values of q, where q varies. This fact is important to ensure correct application of the formulas.

To determine the expected claim size:

From Equation 1.0.6.5,

$$\mathrm{E}_X[X] = \mathrm{E}_q[\mathrm{E}_X[X \mid q]]$$

$$\begin{aligned}\mathrm{E}_X[X \mid q] &= \sum_{\text{all } x} x \cdot \Pr(X = x \mid q) \\ &= 50(0.3) + 100(0.7 - q) + 150q \\ &= 85 + 50q\end{aligned}$$

$$\begin{aligned}\mathrm{E}_X[X] &= \mathrm{E}_q[85 + 50q] \\ &= 85 + 50\mathrm{E}_q[q] \\ &= 85 + 50\int_{0.1}^{0.6} q \cdot \frac{q}{0.175}\,\mathrm{d}q \\ &= 85 + 50\int_{0.1}^{0.6} \frac{q^2}{0.175}\,\mathrm{d}q\end{aligned}$$

$$\begin{aligned} &= 85 + 50\left[\frac{q^3}{0.525}\right]_{0.1}^{0.6} \\ &= 85 + 50\left(\frac{43}{105}\right) \\ &= \mathbf{105.48} \end{aligned}$$

■

Solution to (2)

To determine the probability of getting a claim of 100:

From Equation 1.0.6.4,

$$\begin{aligned} \Pr(X = 100) &= \mathrm{E}_q[\Pr(X = 100 \mid q)] \\ &= \mathrm{E}_q[0.7 - q] \\ &= 0.7 - \mathrm{E}_q[q] \\ &= 0.7 - \frac{43}{105} \\ &= \mathbf{0.2905} \end{aligned}$$

■

Coach's Remarks

The three properties of expected values discussed in Section 1.0.4 were applied in the solution above. Alternatively, we may use Equation 1.0.4.1:

$$E_q[85+50q] = \int_{0.1}^{0.6} (85+50q) \cdot \frac{q}{0.175}\,\mathrm{d}q$$

$$E_q[0.7-q] = \int_{0.1}^{0.6} (0.7-q) \cdot \frac{q}{0.175}\,\mathrm{d}q$$

While the final answers will be the same, the calculations will be more tedious. Therefore, you are encouraged to apply the properties of expected values to simplify the calculations as much as possible.

Independence

If events A and B are *independent*, then the occurrence of A will not affect the probability of the occurrence of B, and vice versa. In mathematical terms,

$$\Pr(A\,|\,B) = \Pr(A)$$

$$\Pr(B\,|\,A) = \Pr(B)$$

Recall that in general,

$$\Pr(A \cap B) = \Pr(A \,|\, B) \cdot \Pr(B)$$

Therefore, with independence,

$$\Pr(A \cap B) = \Pr(A) \cdot \Pr(B) \tag{1.0.6.7}$$

The same concept can also be applied to random variables. If X and Y are independent, then the following holds true.

- Discrete X and Y:

$$\Pr(X = x \mid Y = y) = \Pr(X = x)$$

$$\Pr(X = x,\, Y = y) = \Pr(X = x) \cdot \Pr(Y = y)$$

- Continuous X and Y:

$$f_{X|Y}(x \mid y) = f_X(x)$$

$$f_{X,\,Y}(x,\, y) = f_X(x) \cdot f_Y(y)$$

Independence is a major concept for this exam because it has a very important application. Specifically, if X and Y are independent, then:

$$\mathrm{E}[g(X) \cdot h(Y)] = \mathrm{E}[g(X)] \cdot \mathrm{E}[h(Y)] \tag{1.0.6.8}$$

This special property also applies to MGFs and PGFs because they are expected values. For example, if X and Y are independent, then:

$$\begin{aligned} M_{X+Y}(z) &= \mathrm{E}\left[e^{z(X+Y)}\right] \\ &= \mathrm{E}\left[e^{zX} \cdot e^{zY}\right] \\ &= \mathrm{E}\left[e^{zX}\right] \cdot \mathrm{E}\left[e^{zY}\right] \qquad \text{(independence)} \\ &= M_X(z) \cdot M_Y(z) \end{aligned}$$

The following relationship can be derived in a similar manner.

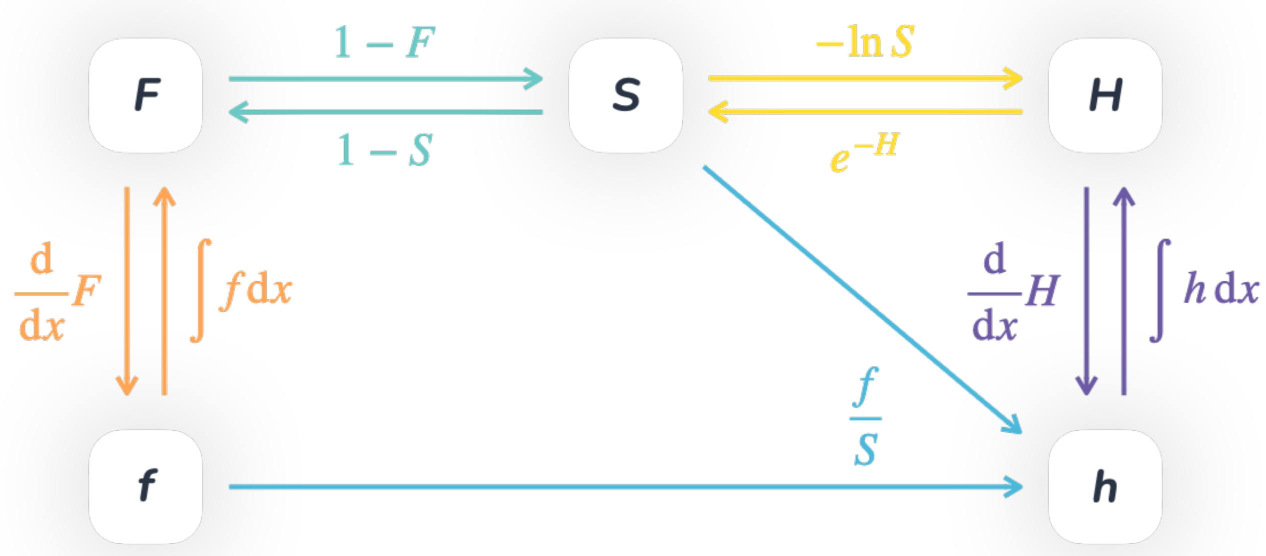
$$P_{X+Y}(z) = P_X(z) \cdot P_Y(z)$$

1.0 Summary

10m

Functions

This chart summarizes the relationships between the various functions of a continuous random variable.



Percentiles

The $100q^{\text{th}}$ percentile of a random variable X is any value π_q that satisfies these two properties:

- $\Pr(X < \pi_q) \leq q$
- $\Pr(X \leq \pi_q) \geq q$

If X is continuous, the $100q^{\text{th}}$ percentile is the value π_q such that $F(\pi_q) = q$.

Mode

The mode of a random variable is the value that produces the largest PMF or PDF.

If the random variable X is continuous, then

$$f'(x) = 0 \quad \Rightarrow \quad x = \text{mode}$$

Moments

- The k^{th} raw moment of X is defined as

$$\mu_k' = \mathrm{E}\left[X^k\right]$$

- The 1^{st} raw moment is the mean, which is usually denoted by μ.

$$\begin{aligned}\mu &= \mathrm{E}[X] \\ &= \sum_{\text{all } x} x \cdot p(x) && \text{(discrete)} \\ &= \int_{-\infty}^{\infty} x \cdot f(x)\,\mathrm{d}x && \text{(continuous)}\end{aligned}$$

- The k^{th} central moment of X is defined as

$$\mu_k = \mathrm{E}\left[(X-\mu)^k\right]$$

- The 2^{nd} central moment is the variance, which is usually denoted by $\mathrm{Var}[X]$ or σ^2.

$$\begin{aligned}\sigma^2 &= \mathrm{Var}[X] \\ &= \mathrm{E}\left[(X-\mu)^2\right] \\ &= \mathrm{E}\left[X^2\right] - \mathrm{E}[X]^2\end{aligned}$$

- The standard deviation, denoted σ, is the square root of the variance.

$$\sigma = \sqrt{\mathrm{Var}[X]}$$

- The coefficient of variation is the ratio of the standard deviation to the mean.

$$CV = \frac{\sigma}{\mu}$$

- The skewness of a distribution is:

$$\text{Skewness} = \frac{\mu_3}{\sigma^3}$$

- The kurtosis of a distribution is:

$$\text{Kurtosis} = \frac{\mu_4}{\sigma^4}$$

Moment Generating Function (MGF)

- The MGF of a random variable X is denoted as $M_X(z)$.

$$M_X(z) = \mathrm{E}\left[e^{zX}\right]$$

- All MGFs have the following property:

$$M_X^{(n)}(0) = \mathrm{E}[X^n]$$

Probability Generating Function (PGF)

- The PGF of a random variable X is denoted as $P_X(z)$.

$$P_X(z) = \mathrm{E}\left[z^X\right]$$

- All PGFs have the following property:

$$P_X^{(n)}(1) = \mathrm{E}[X(X-1)\ldots(X-n+1)]$$

Conditional Distribution

- Bayes' Theorem:

$$\Pr(A \,|\, B) = \frac{\Pr(B \,|\, A)\Pr(A)}{\Pr(B)}$$

- Based on the Law of Total Probability,

$$\Pr(X = x) = \mathrm{E}_Y[\Pr(X = x \mid Y)]$$

- Based on the Law of Total Expectation,

$$\mathrm{E}_X[X] = \mathrm{E}_Y[\mathrm{E}_X[X \mid Y]]$$

- Based on the Law of Total Variance,

$$\mathrm{Var}_X[X] = \mathrm{E}_Y[\mathrm{Var}_X[X \mid Y]] + \mathrm{Var}_Y[\mathrm{E}_X[X \mid Y]]$$

Independence

Events A and B are independent if and only if the following equation is true:

$$\Pr(A \cap B) = \Pr(A) \cdot \Pr(B)$$

If random variables X and Y are independent, then

$$\Pr(X = x,\, Y = y) = \Pr(X = x) \cdot \Pr(Y = y)$$

$$f_{X,Y}(x,\, y) = f_X(x) \cdot f_Y(y)$$

$$\mathrm{E}[g(X) \cdot h(Y)] = \mathrm{E}[g(X)] \cdot \mathrm{E}[h(Y)]$$

Appendix

10m

Expected Value - Survival Function Method

In general, the expected value of a function X, i.e. $g(X)$, is

$$\begin{aligned}
\mathrm{E}\left[g(X)\right] &= \int_0^\infty g(x) \cdot f_X(x)\,\mathrm{d}x \\
&= \left[-g(x) \cdot S_X(x)\right]\Big|_0^\infty - \int_0^\infty g'(x) \cdot \left[-S_X(x)\right]\mathrm{d}x \\
&= \Bigg[-g(\infty) \cdot \underbrace{S_X(\infty)}_{0} + \underbrace{g(0)}_{0} \cdot S_X(0)\Bigg] + \int_0^\infty g'(x) \cdot S_X(x)\,\mathrm{d}x \\
&= \int_0^\infty g'(x) \cdot S_X(x)\,\mathrm{d}x
\end{aligned}$$

Note that in order to simplify to the final line, two requirements need to be fulfilled:

1. The upper bound of the integral needs to be the upper bound of X, e.g. $S_X(\infty) = 0$ in the equation above.
2. The function evaluated at the lower bound needs to be 0, e.g. $g(0) = 0$ in the equation above.

If the two requirements are not fulfilled, we need to adjust the formula accordingly. Below are a few common usages of the method throughout this exam:

$$\begin{aligned}
\mathrm{E}\left[X\right] &= \int_0^\infty x \cdot f_X(x)\,\mathrm{d}x \\
&= \left[-x \cdot S_X(x)\right]\Big|_0^\infty - \int_0^\infty 1 \cdot \left[-S_X(x)\right]\mathrm{d}x \\
&= \int_0^\infty S_X(x)\,\mathrm{d}x
\end{aligned}$$

$$= \int_{\mathbf{0}} \boldsymbol{S_X(x)}\,\mathrm{d}\boldsymbol{x}$$

$$\begin{aligned}
\mathrm{E}\left[X \wedge u\right] &= \int_0^\infty \min(x,\, u) \cdot f_X(x)\,\mathrm{d}x \\
&= \int_0^u x \cdot f_X(x)\,\mathrm{d}x + \int_u^\infty u \cdot f_X(x)\,\mathrm{d}x \\
&= \left\{ \left[-x \cdot S_X(x)\right]\Big|_0^u - \int_0^u 1 \cdot \left[-S_X(x)\right]\mathrm{d}x \right\} + u \cdot S_X(u) \\
&= \left\{ -u \cdot S_X(u) + \int_0^u S_X(x)\,\mathrm{d}x \right\} + u \cdot S_X(u) \\
&= \int_{\mathbf{0}}^{\boldsymbol{u}} \boldsymbol{S_X(x)}\,\mathrm{d}\boldsymbol{x}
\end{aligned}$$

$$\begin{aligned}
\mathrm{E}\left[(X-d)_+\right] &= \int_0^\infty \max(x-d,\, 0) \cdot f_X(x)\,\mathrm{d}x \\
&= \int_0^d 0 \cdot f_X(x)\,\mathrm{d}x + \int_d^\infty (x-d) \cdot f_X(x)\,\mathrm{d}x \\
&= 0 + \left\{ \left[-(x-d) \cdot S_X(x)\right]\Big|_d^\infty - \int_d^\infty 1 \cdot \left[-S_X(x)\right]\mathrm{d}x \right\} \\
&= \int_{\boldsymbol{d}}^{\boldsymbol{\infty}} \boldsymbol{S_X(x)}\,\mathrm{d}\boldsymbol{x}
\end{aligned}$$

3rd Central Moment

Assume

- $\mu_k = \mathrm{E}\left[(X-\mu)^k\right]$
- $\mu'_k = \mathrm{E}\left[X^k\right]$

Then,

$$\begin{aligned}\mu_3 &= \mathrm{E}\left[(X-\mu)^3\right]\\ &= \mathrm{E}\left[X^3 - 3X^2\mu + 3X\mu^2 - \mu^3\right]\\ &= \mathrm{E}\left[X^3\right] - 3\mu\mathrm{E}\left[X^2\right] + 3\mu^2\mathrm{E}[X] - \mu^3\\ &= \mu'_3 - 3\mu\left(\mu'_2\right) + 3\mu^2(\mu) - \mu^3\\ &= \mu'_3 - 3\mu\left(\mu'_2\right) + 3\mu^3 - \mu^3\\ &= \mu'_3 - 3\mu'_2\mu + 2\mu^3\end{aligned}$$

4th Central Moment

Assume

- $\mu_k = \mathrm{E}\left[(X-\mu)^k\right]$
- $\mu'_k = \mathrm{E}\left[X^k\right]$

Then,

$$\begin{aligned}\mu_4 &= \mathrm{E}\left[(X-\mu)^4\right]\\ &= \mathrm{E}\left[X^4 - 4X^3\mu + 6X^2\mu^2 - 4X\mu^3 + \mu^4\right]\\ &= \mathrm{E}\left[X^4\right] - 4\mu\mathrm{E}\left[X^3\right] + 6\mu^2\mathrm{E}\left[X^2\right] - 4\mu^3\mathrm{E}[X] + \mu^4\\ &= \mu'_4 - 4\mu\left(\mu'_3\right) + 6\mu^2\left(\mu'_2\right) - 4\mu^3(\mu) + \mu^4\end{aligned}$$

$$\begin{aligned}
&= \mu_4' - 4\mu\left(\mu_3'\right) + 6\mu^2\left(\mu_2'\right) - 4\mu^4 + \mu^4 \\
&= \mu_4' - 4\mu_3'\mu + 6\mu_2'\mu^2 - 3\mu^4
\end{aligned}$$

The Law of Total Variance

$$\mathrm{Var}_X[X] = \mathrm{E}_X\left[X^2\right] - \mathrm{E}_X[X]^2$$

Apply the Law of Total Expectation.

$$\mathrm{E}_X[X] = \mathrm{E}_Y[\mathrm{E}_X[X \mid Y]]$$

$$\begin{aligned}
\mathrm{E}_X\left[X^2\right] &= \mathrm{E}_Y\left[\mathrm{E}_X\left[X^2 \mid Y\right]\right] \\
&= \mathrm{E}_Y\left[\mathrm{Var}_X[X \mid Y] + \mathrm{E}_X[X \mid Y]^2\right] \\
&= \mathrm{E}_Y[\mathrm{Var}_X[X \mid Y]] + \mathrm{E}_Y\left[\mathrm{E}_X[X \mid Y]^2\right]
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathrm{Var}_X[X] &= \mathrm{E}_X\left[X^2\right] - \mathrm{E}_X[X]^2 \\
&= \mathrm{E}_Y[\mathrm{Var}_X[X \mid Y]] + \mathrm{E}_Y\left[\mathrm{E}_X[X \mid Y]^2\right] - \mathrm{E}_Y[\mathrm{E}_X[X \mid Y]]^2
\end{aligned}$$

Recall that $\mathrm{Var}[g(X)] = \mathrm{E}\left[g(X)^2\right] - \mathrm{E}[g(X)]^2$. Thus, we can translate the second half of the equation above to:

$$\mathrm{E}_Y\left[\mathrm{E}_X[X \mid Y]^2\right] - \mathrm{E}_Y[\mathrm{E}_X[X \mid Y]]^2 = \mathrm{Var}_Y[\mathrm{E}_X\left[X \mid Y\right]]$$

Thus,

$$\mathrm{Var}_X[X] = \mathrm{E}_Y[\mathrm{Var}_X[X \mid Y]] + \mathrm{Var}_Y[\mathrm{E}_X[X \mid Y]]$$

1.1.0 Overview

5m

Insurance operates by pooling risk together across policies. Because of this, insurance companies are interested in two main components of losses: claim *severity* and claim *frequency*. Claim severity is the size of a loss, while claim frequency is the number of losses for a block of insurance policies over a set period of time. Frequency and severity are typically modeled separately. For frequency, a non-negative discrete random variable is generally used. For severity, a continuous random variable is most common. Then, the *aggregate loss* is found by summing the loss sizes across all losses.

We will begin this section by looking at the exam table for this exam and the following common claim severity distributions:

- Pareto
- Beta
- Uniform
- Normal
- Lognormal
- Gamma
- Exponential
- Weibull

Then, we will consider important properties of the exponential distribution, which will be useful in later subsections. Next, we will learn about two greedy algorithms. We will end by examining ways to create new distributions through mixtures, transformations, and splicing.

1.1.1 Exam Tables

10m

During the exam, you will have access to an exam table containing information to assist you in solving exam problems. Here is a link to the exam table.

What does the exam table include?

The *Akaike Information Criterion* (AIC) and the *Bayesian Information Criterion* (BIC) formulas may vary depending on the source, so the exam table includes formulas to use for AIC and BIC, unless other formulas are explicitly stated in a problem. AIC and BIC are typically used to compare the fit of models.

The *standard normal distribution table* lists $\Pr(Z \le z)$ for select values of z, where Z has a standard normal distribution. Below the normal distribution table are commonly used standard normal percentiles. This table is useful when solving problems related to the normal distribution.

The *Illustrative Life Table* lists relevant values for calculations involving the probability of survival and death and the payments of cash flows upon survival or death of human lives.

The *Interest Functions* table provides interest function values at $i = 0.06$ for different compounding periods.

Appendix A contains the continuous distributions. Section A.1 defines the **incomplete gamma function** and **incomplete beta function**. These functions are needed to evaluate the CDF of the gamma and beta distributions, respectively.

It is in your best interest to be familiar with the way the continuous distributions are organized. Knowing where the common distributions are will minimize the amount of time spent looking for them during the exam.

- From Section A.2 to A.3, the continuous distributions are sorted by the number of parameters in descending order, starting with distributions with three parameters, then two, then one.
- This pattern stops at Section A.4 where distributions that don't fit into any of the last three sections are listed. Only the lognormal, inverse Gaussian, and single-parameter Pareto distributions are important.

- Section A.5 lists distributions with finite support, i.e. the possible values are on a bounded interval. Only the beta distribution is important.

Despite the large number of continuous distributions in Appendix A, you do not need to know all of them. However, you need to be familiar with the commonly tested distributions to be able to use the exam table efficiently.

The exam table includes the following four basic functions/formulas for every continuous distribution (except log-t): the **PDF**, the **CDF**, the **raw moments**, and the **limited moments**. The **mode** formulas are also included for every continuous distribution except inverse Gaussian, log-t, generalized beta, and beta.

Coach's Remarks

For certain continuous distributions, the exam table provides formulas to calculate the k^{th} moment "if k is an integer". This only refers to **positive** integers. These formulas do not work if k is negative.

Appendix B contains the $(a,\, b,\, 0)$ class of discrete distributions. For each distribution, you are provided with the **PMF**, $\boldsymbol{a}$ and $\boldsymbol{b}$, the **mean**, the **variance**, and the **PGF**.

Coach’s Remarks

For the distributions listed in Appendices A and B, unless stated otherwise, their support is $0 < x < \infty$ (restricted to integers if the distribution is discrete). In fact, only a few distributions deviate from this general rule:

- A.4.1.4 Single-Parameter Pareto: $x > \theta$
- A.5.1.2 Beta: $0 < x < \theta$ (including generalized beta distribution)
- B.2.1.3 Binomial: $k = 0,\, 1,\, \ldots,\, m$

Coach’s Remarks

In some formulas, a specific notation is used to represent the product of an arithmetic sequence. For example, in Appendix B.2.1.4 (negative binomial distribution), the numerator of the PMF contains the term:

$$r(r+1)\ldots(r+k-1)$$

This notation describes the product of an arithmetic sequence that:

- Starts from r,
- Increments by 1 at each step,
- Ends at $r+k-1$.

Thus, for different values of k:

- When $k=1$, the sequence consists of a single term r, so the product is simply r.
- When $k=2$, the sequence includes r and $r+1$, yielding the product $r(r+1)$.
- When $k=3$, the sequence extends to r, $r+1$, $r+2$, giving the product $r(r+1)(r+2)$.
- This pattern continues accordingly.

The *t-distribution table* lists the percentiles corresponding to the two-sided tail areas of the t-distribution for specific degrees of freedom and probabilities.

The *F-distribution table* lists the the percentiles corresponding to upper-tail areas of the

F-distribution for specific numerator and denominator degrees of freedom and survival probabilities.

The *chi-square distribution table* lists the chi-square values for select percentiles corresponding to P (cumulative probability) and degrees of freedom.

1.1.2 Pareto

15m

Pareto

The Pareto distribution was originally used to study income distributions. It can be derived as a mixture of exponential random variables (shown in Section 1.1.9). On the exam, "Pareto" refers to *two-parameter Pareto* unless specified otherwise.

Let X follow a Pareto distribution with parameters α and θ, i.e.

$$X \sim \text{Pareto}\,(\alpha,\, \theta)$$

Then, X has the following PDF:

$$f(x) = \frac{\alpha\theta^{\alpha}}{(x+\theta)^{\alpha+1}}, \qquad x \geq 0$$

To recognize a Pareto distribution, note that the numerator of the PDF will be a constant and the denominator will have the term "x plus a constant" raised to a positive power.

$$f(x) = \frac{c}{(x+\theta)^{\alpha+1}}$$

The hazard function for Pareto decreases with x for any values of the parameters α and θ.

The Pareto distribution does not have an MGF, but the expected value can be found by evaluating the moments formula given in the exam table at $k = 1$.

$$\mathrm{E}[X] = \frac{\theta}{\alpha - 1}$$

The variance can be calculated from the first and second moments or using this shortcut.

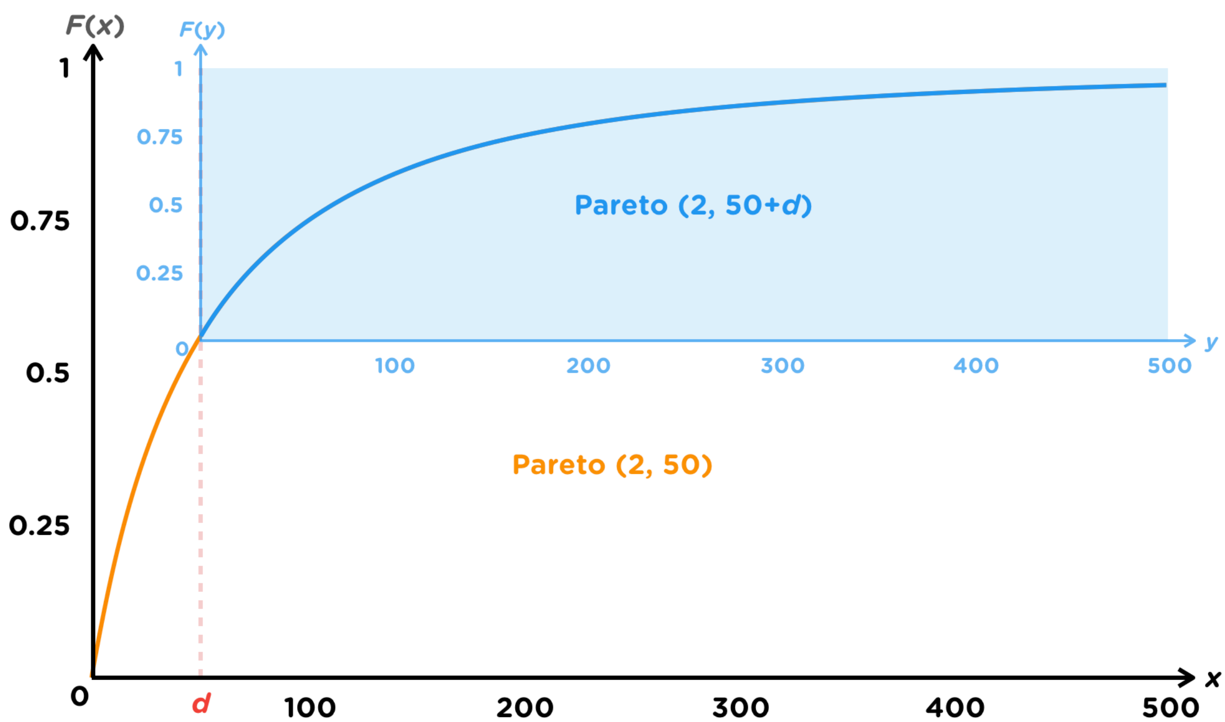
$$\mathrm{Var}[X] = \mathrm{E}[X]^2 \left(\frac{\alpha}{\alpha - 2} \right)$$

Note that a Pareto distribution only has k moments, where $k < \alpha$. For example, when $\alpha \leq 2$, the second moment does not exist. In that situation, the variance also does not exist since it depends on the second moment.

The Pareto distribution has a special property that is worth remembering. If $X \sim \text{Pareto}\ (\alpha,\, \theta)$ and $Y = X - d \mid X > d$, then

$$Y \sim \text{Pareto}\ (\alpha,\, \theta + d)$$

This is illustrated in the graph below.



Knowing this concept will be useful for later topics, e.g. deductibles.

Example 1.1.2.1

For each accident, loss size follows a Pareto distribution with parameters $\alpha = 3$ and $\theta = 500$.

Insurance only covers the portion of loss in excess of 200.

Determine the expected amount the insurance will pay for an accident, given the loss is greater than 200.

Solution

Let X represent loss size.

$$X \sim \text{Pareto}\ (3,\, 500)$$

The insurance payment is $X - 200$ if X exceeds 200. The expected payment given the loss is above 200 is

$$\text{E}[X - 200 \mid X > 200]$$

Apply the Pareto shortcut.

$$X - d \mid X > d \sim \text{Pareto}\ (\alpha,\, \theta + d)$$

$$X - 200 \mid X > 200 \sim \text{Pareto}\ (3,\, 700)$$

Therefore,

$$\begin{aligned}\mathrm{E}[X-200 \mid X>200] &= \frac{700}{3-1} \\ &= \mathbf{350}\end{aligned}$$

$\blacksquare$

Inverse Pareto

As the name suggests, an *inverse Pareto* distribution is an inverted Pareto distribution.

Specifically, if $Y \sim \text{Pareto}\left(\alpha,\, \theta^*\right)$ and $X = Y^{-1}$, then

$$X \sim \text{Inverse Pareto}\left(\tau=\alpha,\, \theta=\theta^{*-1}\right)$$

with the PDF

$$f(x) = \frac{\tau\theta x^{\tau-1}}{(x+\theta)^{\tau+1}}, \qquad x \geq 0$$

The proof of the inversion is included in the appendix at the end of this section.

An inverse Pareto's PDF resembles a Pareto's PDF, except the inverse Pareto's PDF has an $x^{\tau-1}$ term in its numerator.

$$f(x) = \frac{c \cdot x^{\tau-1}}{(x+\theta)^{\tau+1}}$$

Coach's Remarks

The same inversion methodology applies to all distributions on the exam table that have an inverse counterpart. The parameter θ is parameterized such that when X follows a certain distribution, X^{-1} follows the corresponding inverted distribution with the same parameters except theta becomes θ^{-1}. Another example would be the gamma and inverse gamma distributions.

Single-Parameter Pareto

Let X follow a *single-parameter Pareto* distribution with parameters α and θ, i.e.

$$X \sim \text{S-P Pareto}\ (\alpha,\, \theta)$$

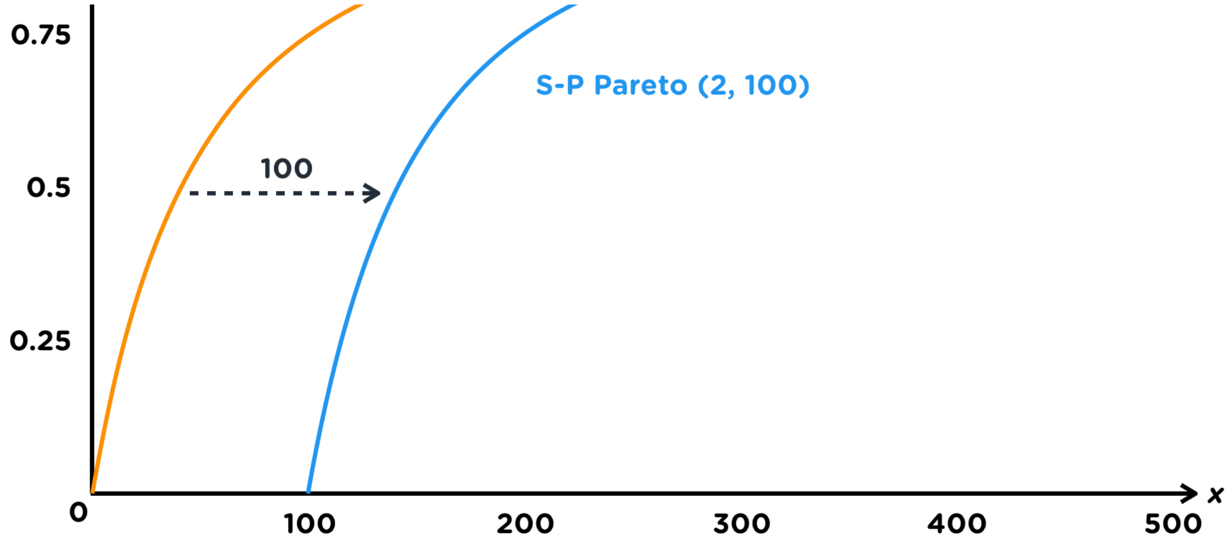
Then, X has the following PDF:

$$f(x) = \frac{\alpha\theta^{\alpha}}{x^{\alpha+1}}, \qquad x > \theta$$

A single-parameter Pareto distribution is recognizable from its domain, which starts with a positive constant, whereas most other distributions have domains that start with 0.

A single-parameter Pareto distribution is a Pareto distribution shifted by θ. For $Y \sim \text{Pareto}\ (\alpha,\, \theta)$ and $X = Y + \theta$,

$$X \sim \text{S-P Pareto}\ (\alpha,\, \theta)$$



Thus, assuming the same parameters, the single-parameter Pareto mean is greater than the Pareto mean by θ

$$\begin{aligned} \mathrm{E}[X] &= \mathrm{E}[Y+\theta] \\ &= \mathrm{E}[Y]+\theta \\ &= \frac{\theta}{\alpha-1}+\theta \\ &= \frac{\alpha\theta}{\alpha-1} \end{aligned}$$

and its variance is equal to the Pareto variance.

$$\begin{aligned} \mathrm{Var}[X] &= \mathrm{Var}[Y+\theta] \\ &= \mathrm{Var}[Y] \\ &= \mathrm{E}[Y]^2\left(\frac{\alpha}{\alpha-2}\right) \\ &= \left(\frac{\theta}{\alpha-1}\right)^2\left(\frac{\alpha}{\alpha-2}\right) \end{aligned}$$

Coach's Remarks

θ is the amount by which the Pareto distribution is shifted. Thus, it is not a true parameter and must be determined in advance. The only true parameter is α, hence the name single-parameter Pareto.

For $X \sim \text{S-P Pareto}\ (\alpha,\, \theta)$ and $d > \theta$, then

$$X \mid X > d \sim \text{S-P Pareto}\ (\alpha,\, d)$$

And since a single-parameter Pareto is a shifted Pareto, reversing the shift results in

$$X - d \mid X > d \sim \text{Pareto}\ (\alpha,\, d)$$

1.1.3 Beta and Uniform

10m

Beta

Let X follow a *beta* distribution with parameters a, b, and θ, i.e.

$$X \sim \text{Beta}\,(a,\, b,\, \theta)$$

Then, X has the following PDF:

$$f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\left(\frac{x}{\theta}\right)^{a}\left(1-\frac{x}{\theta}\right)^{b-1}\frac{1}{x}, \qquad 0 \le x \le \theta$$

A beta distribution's PDF is recognizable from its finite support and its purely polynomial terms, i.e. no negative powers of x, and no exponential or logarithmic terms.

Coach's Remarks

Many students forget the $\frac{1}{x}$ term at the end even though the PDF is available on the exam table. Don't be one of them.

The mean and variance are

$$\text{E}[X] = \frac{a}{a+b}\cdot\theta$$

$$\mathrm{Var}[X] = \frac{ab}{(a+b)^2\,(a+b+1)} \cdot \theta^2$$

Coach's Remarks

Most questions that use the beta distribution will set θ to 1. In that case, the PDF simplifies to

$$f(x) = c \cdot x^{a-1}(1-x)^{b-1}, \qquad 0 \le x \le 1$$

which makes it easier to integrate and recognize as beta.

Example 1.1.3.1

A random variable X has the following PDF:

$$f(x) = 105x^4(1-x)^2, \qquad 0 \le x \le 1$$

Determine the second moment of X.

Solution

Because the PDF has finite support and only consists of polynomial terms, it belongs to a beta distribution. The support is from 0 to 1, which means $\theta = 1$.

The PDF of a beta distribution with $\theta = 1$ is in the form of $c \cdot x^{a-1}(1-x)^{b-1}$. Compare this to the PDF given to deduce $a = 5$ and $b = 3$.

$$X \sim \text{Beta}\,(5,\,3,\,1)$$

Look up beta's moments formula in the exam table. We can use the simplified version since k is an integer in this example:

$$\text{E}\left[X^k\right] = \frac{\theta^k a\,(a+1)\cdots(a+k-1)}{(a+b)\,(a+b+1)\cdots(a+b+k-1)}$$

For the second moment, $k = 2$. So, the last term in the numerator is $a + 2 - 1 = a + 1$, and the last term in the denominator is $a + b + 2 - 1 = a + b + 1$. Altogether, the second moment is

$$\begin{aligned}\text{E}\left[X^2\right] &= \frac{\theta^2 a\,(a+1)}{(a+b)\,(a+b+1)} \\ &= \frac{1^2\,(5)\,(5+1)}{(5+3)\,(5+3+1)} \\ &= \frac{\mathbf{5}}{\mathbf{12}}\end{aligned}$$

■

Alternative Solution

Alternatively, you can solve using first principles.

$$\begin{aligned}
\mathrm{E}\big[X^2\big] &= \int_{-\infty}^{\infty} x^2 \cdot f(x)\,\mathrm{d}x \\
&= \int_0^1 x^2 \cdot 105\,x^4(1-x)^2\,\mathrm{d}x \\
&= 105 \cdot \int_0^1 x^6\left(1-2x+x^2\right)\mathrm{d}x \\
&= 105 \cdot \int_0^1 \left(x^6-2x^7+x^8\right)\mathrm{d}x \\
&= 105 \cdot \left[\frac{x^7}{7}-\frac{2x^8}{8}+\frac{x^9}{9}\right]_0^1 \\
&= \mathbf{\frac{5}{12}}
\end{aligned}$$

■

Uniform

A *uniform* distribution has a constant density. Let X follow a uniform distribution on the interval $[a,\, b]$, i.e.

$$X \sim \mathrm{Uniform}\ (a,\, b)$$

Then, X has the following PDF:

$$f(x) = \frac{1}{b-a}, \qquad a \le x \le b$$

The mean is the **midpoint** of the interval.

$$\mathrm{E}[X] = \frac{a+b}{2}$$

The variance is

$$\mathrm{Var}[X] = \frac{(a-b)^2}{12}$$

The second raw moment can be easily calculated by adding the variance to the square of the mean. However, some students prefer this shortcut:

$$\mathrm{E}\left[X^2\right] = \frac{a^2+ab+b^2}{3}$$

Coach's Remarks

A uniform distribution on the interval $[0,\,\theta]$ is equivalent to a beta distribution with parameters $a=b=1$ and θ.

Assume X is uniformly distributed on the interval $[a,\,b]$. Then, X given it is greater than d, where $a<d<b$, will be uniformly distributed on the interval $[d,\,b]$.

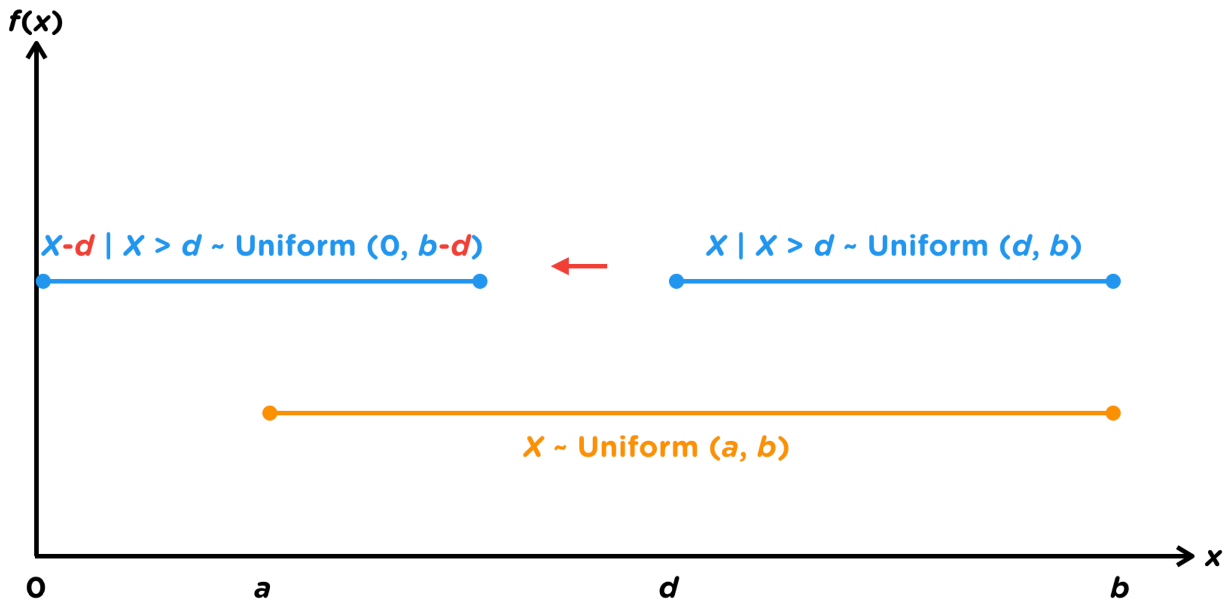
$$X \sim \mathrm{Uniform}\,(a,\,b)$$

$$X \mid X > d \sim \text{Uniform}\,(d,\, b)$$

Shifting X leftwards by d will shift the endpoints of the interval by the same amount. Therefore,

$$X - d \mid X > d \sim \text{Uniform}\,(0,\, b - d)$$

The figure below illustrates the transition from X to $X \mid X > d$ to $X - d \mid X > d$.



1.1.4 Normal and Lognormal

20m

Normal

Let X follow a *normal* distribution with mean μ and variance σ^2, i.e.

$$X \sim \text{Normal}\left(\mu, \sigma^2\right)$$

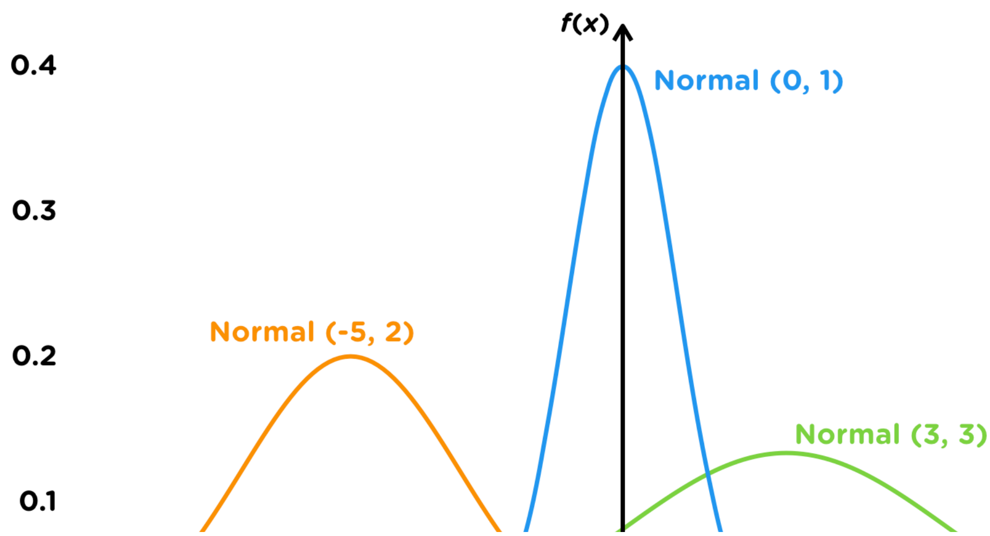
Then, X has the following PDF:

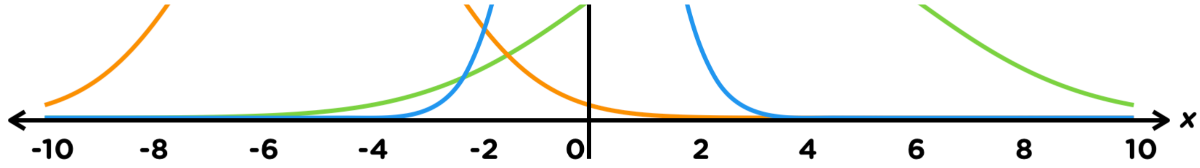
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \qquad -\infty < x < \infty$$

The normal distribution is unique in that its parameters are its mean and variance.

$$\text{E}[X] = \mu$$

$$\text{Var}[X] = \sigma^2$$





The blue curve in the middle is the PDF of a *standard* normal distribution, which has mean 0 and variance 1 and is usually denoted as Z.

$$Z \sim \text{Normal}\,(0, 1)$$

For a standard normal distribution, represent the CDF as $\Phi(z)$.

$$\Phi(z) = \Pr(Z \le z)$$

In addition, denote the $100q^{\text{th}}$ percentile of the standard normal distribution as z_q.

$$z_q = \Phi^{-1}(q)$$

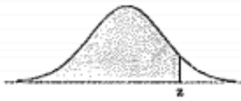
Understanding the standard normal distribution is critical in solving all normal distribution problems, including related distributions such as the lognormal distribution. A standard normal distribution table will be provided on the exam, so it's important to understand how to use it.

Consider the following example:

Z is a standard normal random variable.

1. Calculate $\Pr(Z \le 1.55)$.
2. Calculate $\Pr(Z \le -1.55)$.

Tables of the Normal Distribution



Probability Content from -∞ to Z

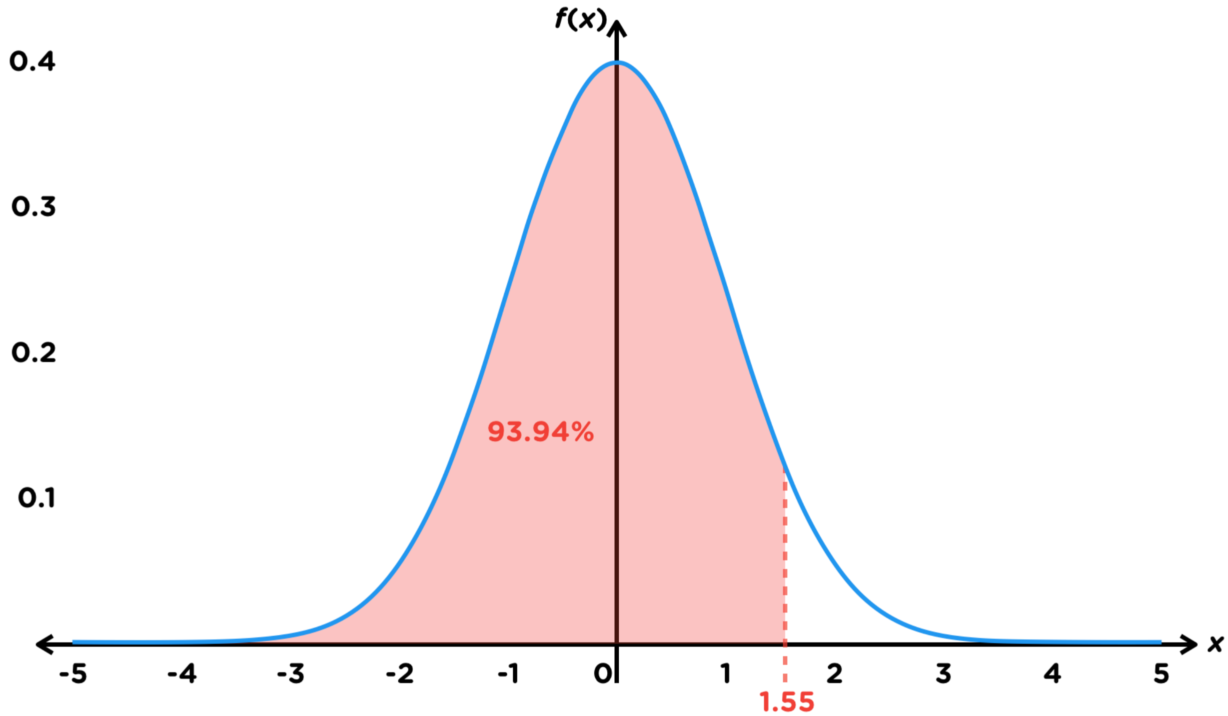
Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

Values of z for selected values of Pr(Z<z)

z	0.842	1.036	1.282	1.645	1.960	2.326	2.576
Pr(Z<z)	0.800	0.850	0.900	0.950	0.975	0.990	0.995

From the normal distribution table above, the intersection of the "1.5" row and the "0.05" column is 0.9394. This means the standard normal's CDF evaluated at 1.55 is 0.9394. A graphical representation of this area is shown in the graph below.

$$\begin{aligned}\Phi(1.55) &= \Pr(Z \le 1.55) \\ &= \mathbf{0.9394}\end{aligned}$$



In other words, the 93.94th percentile of the standard normal distribution is 1.55.

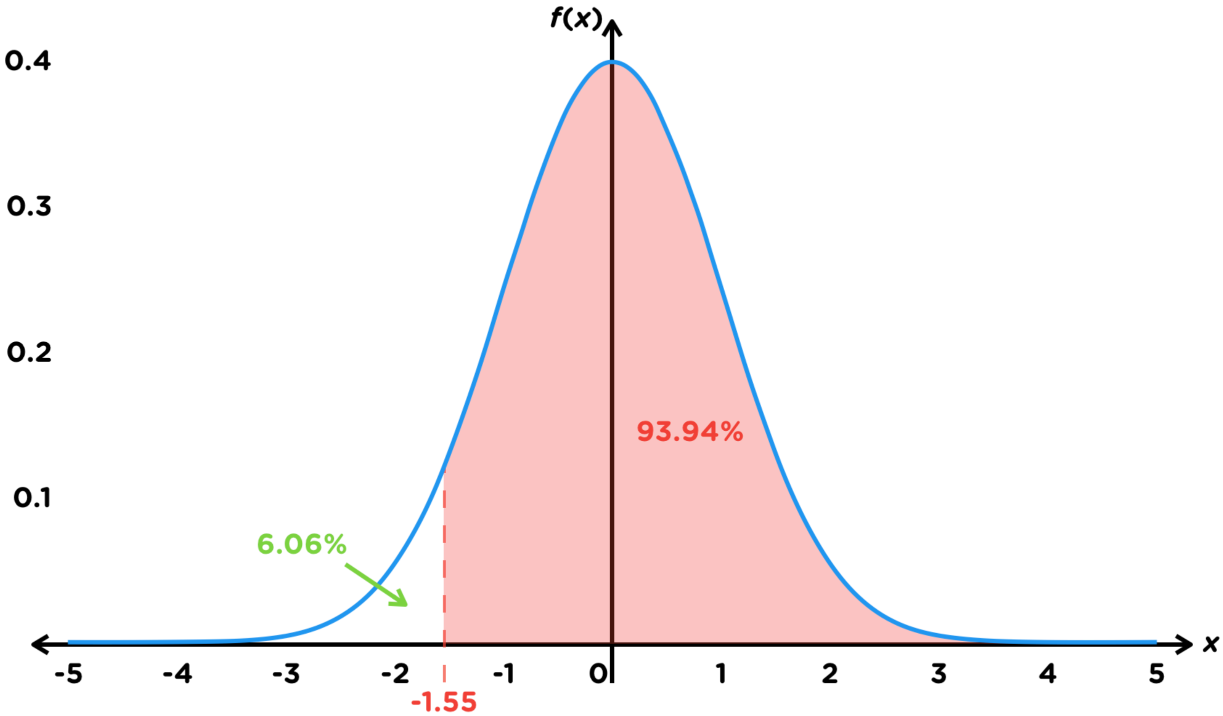
$$\begin{aligned} z_{0.9394} &= \Phi^{-1}(0.9394) \\ &= 1.55 \end{aligned}$$

Also, due to the symmetry of the standard normal distribution (see graph below),

$$\Pr(Z \geq -1.55) = \Pr(Z \leq 1.55) = 0.9394$$

Thus,

$$\begin{aligned} \Phi(-1.55) &= \Pr(Z \leq -1.55) \\ &= 1 - \Pr(Z > -1.55) \\ &= 1 - \Pr(Z \leq 1.55) \\ &= 1 - \Phi(1.55) \\ &= \mathbf{0.0606} \end{aligned}$$



In conclusion:

- $\Phi(z) = \Pr(Z \le z)$
- $z_q = \Phi^{-1}(q)$
- $\Phi(-z) = 1 - \Phi(z)$

The CDF of any normal distribution can be calculated by standardizing the normal random variable.

$$Z = \frac{X - \mu}{\sigma} \quad \Rightarrow \quad X = \mu + Z\sigma$$

$$\begin{aligned}\Pr(X \le x) &= \Pr(\mu + Z\sigma \le x) \\ &= \Pr\left(Z \le \frac{x - \mu}{\sigma}\right) \\ &= \Phi\left(\frac{x - \mu}{\sigma}\right)\end{aligned}$$

$$^{-}\left(\quad \sigma \quad \right)$$

The same logic applies to percentiles. Let x_q be the $100q^{\text{th}}$ percentile of a normal distribution with mean μ and variance σ^2.

$$x_q = \mu + z_q \sigma$$

Consider the following example:

X follows a normal distribution with mean 3 and standard deviation 2.

1. Calculate the 97th percentile of X.
2. Calculate $\Pr(X \le 5)$.

Tables of the Normal Distribution

Probability Content from -∞ to Z

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441

1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

Values of z for selected values of Pr(Z<z)

z	0.842	1.036	1.282	1.645	1.960	2.326	2.576
Pr(Z<z)	0.800	0.850	0.900	0.950	0.975	0.990	0.995

From the normal distribution table above, the intersection of the "1.8" row and the "0.08" column is 0.9699. This means the standard normal's CDF evaluated at 1.88 is 0.9699, which is closest to the 0.97 we need.

$$z_{0.97} = 1.88$$

Coach's Remarks

Students should not interpolate the z values to match the desired probability, unless explicitly told to do so in a problem. Just pick the z value with the closest probability to the desired probability.

Reverse the standardization to get the percentile for X.

$$\begin{aligned} x_{0.97} &= \mu + z_{0.97}\sigma \\ &= 3 + 1.88\,(2) \\ &= \mathbf{6.76} \end{aligned}$$

$$= 0.70$$

To calculate the probability, standardize X and refer to the normal distribution table.

$$\begin{aligned}\Pr(X \leq 5) &= \Pr\left(\frac{X-\mu}{\sigma} \leq \frac{5-3}{2}\right)\\ &= \Pr(Z \leq 1)\\ &= \Phi(1)\\ &= \mathbf{0.8413}\end{aligned}$$

Tables of the Normal Distribution

Probability Content from -∞ to Z

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

Values of z for selected values of Pr(Z<z)

z	0.842	1.036	1.282	1.645	1.960	2.326	2.576
Pr(Z<z)	0.800	0.850	0.900	0.950	0.975	0.990	0.995

Lognormal

A *lognormal* distribution is a log-transformed normal distribution.

For $Y \sim \text{Normal}\left(\mu, \sigma^2\right)$ and $X = e^Y$,

$$X \sim \text{Lognormal}\left(\mu, \sigma^2\right)$$

A lognormal distribution with parameters μ and σ^2 has the following PDF:

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right), \qquad x > 0$$

A lognormal distribution's PDF resembles a normal distribution's PDF, with the most prominent difference being $\ln x$ in the exponent in place of x.

Note that unlike the normal distribution, the μ and σ^2 of the lognormal distribution are **not** its mean and variance. They are just parameters.

The lognormal mean and variance are as follows. The variance can be calculated from the first and second moments or using the shortcut formula below.

$$\text{E}[X] = e^{\mu + 0.5\sigma^2}$$

$$\text{Var}[X] = \text{E}[X]^2\left(e^{\sigma^2} - 1\right)$$

The CDF of a lognormal distribution can be calculated by log-transforming the normal

The CDF of a lognormal distribution can be calculated by log transforming the normal random variable:

$$X = e^Y \quad \Rightarrow \quad Y = \ln X$$

$$\begin{aligned}
\Pr(X \le x) &= \Pr\left(e^Y \le x\right) \\
&= \Pr(Y \le \ln x) \\
&= \Pr\left(Z \le \frac{\ln x - \mu}{\sigma}\right) \\
&= \Phi\left(\frac{\ln x - \mu}{\sigma}\right)
\end{aligned}$$

Note that X is the lognormal random variable, Y is the normal random variable (mean μ and variance σ^2), and Z is the standard normal random variable.

Percentiles can be calculated intuitively using what we know about normal distribution percentiles:

1. Start with the standard normal's $100q^{\text{th}}$ percentile.

$$z_q$$

2. Convert to the corresponding $100q^{\text{th}}$ percentile of a non-standard normal random variable.

$$y_q = \mu + z_q\sigma$$

3. Convert the normal variable to its lognormal counterpart.

$$x_q = e^{y_q} = e^{\mu + z_q\sigma}$$

Consider the following example:

X is a lognormal random variable with parameters $\mu = 0.4$ and $\sigma^2 = 0.25$.

1. Calculate the 95$^{\text{th}}$ percentile of X.
2. Calculate $\Pr(X \le 2)$

Look up the 95$^{\text{th}}$ standard normal percentile from the normal distribution table.

Values of z for selected values of Pr(Z<z)

z	0.842	1.036	1.282	1.645	1.960	2.326	2.576
Pr(Z<z)	0.800	0.850	0.900	0.950	0.975	0.990	0.995

$$z_{0.95} = 1.645$$

Coach's Remarks

Note that the smaller table underneath the normal distribution table provides the z values for common percentiles. The z values here are accurate to three decimal places, compared to two decimal places in the full-sized table. You **should** use the three decimal z value as much as possible.

Reverse the standardization to calculate the 95$^{\text{th}}$ normal percentile.

$$0.4 + 1.645\sqrt{0.25} = 1.2225$$

$$y_{0.95} = 0.4 + 1.645\sqrt{0.25} = 1.2225$$

The 95th lognormal percentile is

$$x_{0.95} = e^{1.2225} = \mathbf{3.3957}$$

To calculate probability, convert the lognormal random variable to a normal, and then standardize it.

$$\begin{aligned}\Pr(X \le 2) &= \Pr(\ln X \le \ln 2) \\ &= \Pr\left(\frac{\ln X - \mu}{\sigma} \le \frac{\ln 2 - 0.4}{\sqrt{0.25}}\right) \\ &= \Pr(Z \le 0.59) \\ &= \Phi(0.59) \\ &= \mathbf{0.7224}\end{aligned}$$

Sum of Normal Random Variables

The sum of **independent** normal random variables is a new normal random variable with mean and variance equal to the sum of the individual means and variances, respectively.

For example, assume Y_1 and Y_2 are two independent normal random variables.

$$Y_1 \sim \text{Normal}\left(\mu_1, \sigma_1^2\right)$$

$$Y_2 \sim \text{Normal}\left(\mu_2, \sigma_2^2\right)$$

Define Y as the sum.

$$Y = Y_1 + Y_2$$

Y follows a normal distribution with the following mean and variance.

$$Y \sim \text{Normal}\left(\mu_1 + \mu_2,\, \sigma_1^2 + \sigma_2^2\right)$$

Product of Lognormal Random Variables

Continuing from above, let $X_1 = e^{Y_1}$ and $X_2 = e^{Y_2}$ be two independent lognormal random variables.

$$X_1 \sim \text{Lognormal}\left(\mu_1,\, \sigma_1^2\right)$$

$$X_2 \sim \text{Lognormal}\left(\mu_2,\, \sigma_2^2\right)$$

Define X as the product.

$$\begin{aligned} X &= X_1 \cdot X_2 \\ &= e^{Y_1} \cdot e^{Y_2} \\ &= e^{Y_1 + Y_2} \\ &= e^{Y} \end{aligned}$$

X follows a lognormal distribution with these new parameters.

$$X \sim \text{Lognormal}\left(\mu_1 + \mu_2,\, \sigma_1^2 + \sigma_2^2\right)$$

In conclusion, the **product** of independent lognormal random variables is a new lognormal random variable where μ and σ^2 are the sum of the individual μ's and σ^2's, respectively.

1.1.5 Gamma, Exponential, and Weibull

20m

Gamma

Let X follow a *gamma* distribution with parameters α and θ, i.e.

$$X \sim \text{Gamma}\,(\alpha,\, \theta)$$

Then, X has the following PDF:

$$f(x) = \frac{(x\,/\,\theta)^{\alpha} e^{-x\,/\,\theta}}{x \cdot \Gamma(\alpha)}, \qquad x > 0$$

Note that the gamma function is interpreted as $\Gamma(n) = (n-1) \cdot \Gamma(n-1)$ for positive values of n. So, we can simplify $\Gamma(n)$ to $(n-1)!$ for positive **integer** values of n. A gamma distribution with a positive integer value of α is also called an *Erlang* distribution.

A gamma distribution is recognizable from its PDF having the term e raised to a negative multiple of x, multiplied by x raised to a positive power.

$$f(x) = c \cdot x^{\alpha-1} e^{-\,x\,/\,\theta}$$

There is no closed form for the hazard function of X, but the hazard function decreases with x if $\alpha < 1$ and increases with x if $\alpha > 1$.

The mean and variance are

$$\mathrm{E}[X] = \alpha\theta$$

$$\mathrm{Var}[X] = \alpha\theta^2$$

Because the expression of the gamma CDF changes based on the value of α, the CDF is expressed in a general manner using the incomplete gamma function, which is defined at the beginning of Appendix A of the exam table.

$$F(x) = \Gamma\left(\alpha;\ \frac{x}{\theta}\right)$$

Coach's Remarks

Students tend to confuse the gamma function, $\Gamma(n)$, with the incomplete gamma function, $\Gamma(\alpha;\ x)$. Note that the latter has two inputs while the former only has one.

Furthermore, both of the functions can be evaluated easily using Excel:

- $\Gamma(n)$ = `GAMMA(n)`
- $\Gamma\left(\alpha;\ \frac{x}{\theta}\right)$ = `GAMMA.DIST(x, α, θ, TRUE)`
  - or, $\Gamma(\alpha;\ x)$ = `GAMMA.DIST(x, α, 1, TRUE)`

Evaluating an incomplete gamma function using the integral provided by the exam table is messy and algebraically intensive.

$$\Gamma(\alpha;\ x) = \frac{1}{\Gamma(\alpha)} \int_0^x t^{\alpha-1} e^{-t}\,\mathrm{d}t, \qquad \alpha > 0,\ x > 0$$

Fortunately, there is a shortcut when α is a **positive integer**.

To evaluate an incomplete gamma function $\Gamma(\alpha;\ x)$, let N be a Poisson random variable with mean x, i.e. the second input of the function. Then

$$\Gamma(\alpha;\, x) = 1 - \Pr(N < \alpha)$$

The derivation of the Poisson shortcut is included in the appendix at the end of this section.

Coach's Remarks

The PMF of the Poisson distribution is

$$p(n) = \frac{e^{-\lambda}\lambda^n}{n!}, \qquad n = 0,\, 1,\, 2, \ldots$$

The shortcut is better illustrated with an example.

Example 1.1.5.1

X follows a gamma distribution with parameters $\alpha = 2$ and $\theta = 5$.

$$f(x) = \frac{x\, e^{-x\,/\,5}}{25}, \qquad x > 0$$

Calculate $F(20)$.

Solution

$F_X(20) = \Gamma\left(2;\ \dfrac{20}{5}\right) = \Gamma(2;\ 4)$. So, let $N \sim \text{Poisson}\ (\lambda = 4)$. Therefore,

$$\begin{aligned} F_X(20) &= \Gamma(2;\ 4) \\ &= 1 - \Pr(N < 2) \\ &= 1 - [p_N(0) + p_N(1)] \\ &= 1 - \left(e^{-4} + 4\,e^{-4}\right) \\ &= \mathbf{0.9084} \end{aligned}$$

Alternatively, using first principles,

$$\begin{aligned} F_X(20) &= \int_0^{20} \frac{x\,e^{-x\,/\,5}}{25}\,\mathrm{d}x \\ &= \frac{1}{25} \cdot \left[-x \cdot 5e^{-x\,/\,5} - 25e^{-x\,/\,5}\right]_0^{20} \\ &= \mathbf{0.9084} \end{aligned}$$

Alternatively, using Excel, = `GAMMA.DIST(20, 2, 5, TRUE)` = **0.9084**.

■

The sum of **independent** gamma random variables, given that they have the **same** value of θ, is also a new gamma random variable.

Assume we have n gamma random variables that are independent of each other.

$$X_1 \sim \text{Gamma}\ (\alpha_1,\ \theta)$$

$$X_2 \sim \text{Gamma}\,(\alpha_2,\, \theta)$$

$$\vdots$$

$$X_n \sim \text{Gamma}\,(\alpha_n,\, \theta)$$

Let X be the sum of the gamma random variables.

$$X = \sum_{i=1}^{n} X_i$$

Then, X follows a gamma distribution with α equal to the sum of the parameters α_i for each of the n gamma random variables and with θ remaining the same.

$$X \sim \text{Gamma}\left(\sum_{i=1}^{n} \alpha_i,\, \theta\right)$$

Example 1.1.5.2

You are given the following information about a sample, X_1, X_2, X_3, and X_4:

- X_i's are all mutually independent.
- $X_i \sim \text{Gamma}\,(0.5, \theta)$ for $i = 1, 2, 3, 4$
- $Y = \sum_{i=1}^{4} X_i$

Calculate the probability that Y is less than its mean.

Solution

We are given that each X_i follows a gamma distribution with parameters $\alpha = 0.5$ and θ. Note that the sum of gamma random variables is also a gamma random variable with $\alpha^* = \sum_{i=1}^{n} \alpha_i$ and θ.

Hence, $Y = \sum_{i=1}^{4} X_i$ follows a gamma distribution with parameters $\alpha = 4 \cdot 0.5 = 2$ and θ. The mean of Y is $\alpha\theta = 2\theta$. So, we want to calculate $\Pr(Y < 2\theta)$, or $F_Y(2\theta)$.

Since α is an integer, we can use the Poisson shortcut to calculate $F_Y(2\theta) = \Gamma\left(2;\ \frac{2\theta}{\theta}\right) = \Gamma(2;\ 2)$. Let N be a Poisson random variable with $\lambda = 2$. Then,

$$\begin{aligned} F_Y(2\theta) &= 1 - \Pr(N < \alpha) \\ &= 1 - \Pr(N < 2) \\ &= 1 - [p_N(0) + p_N(1)] \\ &= 1 - e^{-2} - 2e^{-2} \\ &= \mathbf{0.5940} \end{aligned}$$

■

Exponential

The *exponential* distribution is often used to measure the time between the occurrence of events. So, it is related to the Poisson distribution, which counts the number of occurrences in a time interval. If the time until the next event is exponential with mean θ, then the number of occurrences of that event in a unit time interval is Poisson with mean $\frac{1}{\theta}$, and vice versa.

The exponential distribution is a special case of the gamma distribution, i.e. it is a gamma distribution with $\alpha = 1$. The PDF, mean, and variance simplify to:

$$f(x) = \frac{e^{-x/\theta}}{\theta}, \qquad x \geq 0$$

$$\mathrm{E}[X] = \theta$$

$$\mathrm{Var}[X] = \theta^2$$

The exponential distribution is highly tested on this exam, so we will explore more of its properties in the next subsection.

Weibull

The *Weibull* distribution is a transformed exponential distribution. For $Y \sim \text{Exponential}\,(\mu)$ and $X = Y^{1/\tau}$,

$$X \sim \text{Weibull}\,\left(\theta = \mu^{1/\tau}, \tau\right)$$

and has PDF

$$f(x) = \frac{\tau(x/\theta)^{\tau} e^{-(x/\theta)^{\tau}}}{x}, \qquad x > 0$$

Thus, when $\tau = 1$, the Weibull distribution is equivalent to the exponential distribution.

When $\theta = 1$ instead, the Weibull distribution is called the *standard Weibull*.

The Weibull PDF is similar to the gamma PDF, except the $x\,/\,\theta$ in the exponent is raised to a positive power. Use this fact to differentiate between a Weibull PDF and a gamma PDF.

$$f(x) = c \cdot x^{\tau-1} e^{-(x\,/\,\theta)^{\tau}}$$

The hazard function of X increases with x if $\tau > 1$ and decreases with x if $\tau < 1$.

The mean and variance are

$$\mathrm{E}[X] = \theta \cdot \Gamma\left(1 + \frac{1}{\tau}\right)$$

$$\mathrm{Var}[X] = \theta^2 \cdot \Gamma\left(1 + \frac{2}{\tau}\right) - \left[\theta \cdot \Gamma\left(1 + \frac{1}{\tau}\right)\right]^2$$

Inverse Counterparts

The inverted counterparts of these distributions can be derived based on the same logic used to derive inverse Pareto from Pareto.

Inverse Gamma

For $Y \sim \text{Gamma}\ (\alpha,\, \theta)$ and $X = Y^{-1}$,

$$X \sim \text{Inverse Gamma}\ (\alpha,\, \theta^{-1})$$

Inverse Exponential

For $Y \sim \text{Exponential}\,(\theta)$ and $X = Y^{-1}$,

$$X \sim \text{Inverse Exponential}\,\left(\theta^{-1}\right)$$

Inverse Weibull

For $Y \sim \text{Weibull}\,(\theta, \tau)$ and $X = Y^{-1}$,

$$X \sim \text{Inverse Weibull}\,\left(\theta^{-1}, \tau\right)$$

The proofs of the inversions are included in the appendix at the end of this section.

1.1.6 Properties of the Exponential Distribution

20m

The exponential distribution has many important properties, which makes it a commonly used severity distribution. These properties will be particularly useful when studying Poisson processes later in this section.

Hazard Function

Recall the probability density function of an exponential distribution with mean θ.

$$f(x) = \frac{1}{\theta} e^{-x/\theta}$$

We can express this in terms of λ, where $\lambda = \frac{1}{\theta}$.

$$f(x) = \lambda e^{-\lambda x}$$

Both variations are used for different reasons. We will refer to θ as the mean of an exponential distribution and λ as the *rate* of an exponential distribution. They are reciprocals of each other.

Recall that the failure rate function, or hazard function, is the ratio of the probability density function to the survival function.

$$h(x) = \frac{f(x)}{S(x)} = \frac{f(x)}{1 - F(x)}$$

The exponential distribution is the only continuous distribution with a constant failure rate function:

$$h(x) = \frac{\lambda e^{-\lambda x}}{e^{-\lambda x}} = \lambda$$

Memoryless Property

One of the most important properties of the exponential distribution is the memoryless property, or the lack of memory property. The exam will likely have one to two questions on this property, so make sure to understand it well. Let X be an exponential random variable. The memoryless property states that the probability of $X > t + s$ given $X > t$ is equal to the probability of $X > s$.

$$\Pr(X > t + s \mid X > t) = \Pr(X > s) \tag{1.1.6.1}$$

This can be proven as follows:

$$\begin{aligned}\Pr(X > t + s \mid X > t) &= \frac{\Pr(X > t + s \cap X > t)}{\Pr(X > t)} \\ &= \frac{\Pr(X > t + s)}{\Pr(X > t)} \\ &= \frac{e^{-\lambda(t+s)}}{e^{-\lambda t}} \\ &= \frac{e^{-\lambda t} e^{-\lambda s}}{e^{-\lambda t}} \\ &= e^{-\lambda s} \\ &= \Pr(X > s)\end{aligned}$$

So, the memoryless property loosely implies that what happened in the past does not affect what will happen in the future. For example, if the arrival time of a bus is an exponential random variable, no matter how long you have been waiting for a bus, the probability that you will wait another x minutes is the same as the probability of waiting x minutes from the start.

From a mathematical standpoint, the memoryless property means $(X - d \mid X > d)$ will follow the same distribution as X, where d is a positive constant.

Thus, for $X \sim \text{Exponential } (\theta)$,

$$X - d \mid X > d \sim \text{Exponential}\ (\theta)$$

Based on the memoryless property, the following are also true:

- $\text{E}[X - d \mid X > d] = \text{E}[X] = \theta$
- $\text{Var}[X - d \mid X > d] = \text{Var}[X] = \theta^2$
- $\Pr(X - d \leq a \mid X > d) = \Pr(X \leq a)$

Example 1.1.6.1

The time until the next bus arrives at a bus stop follows an exponential distribution with mean $\theta = 0.25$ hours. 25% of the buses that arrive at that bus stop are express buses while the rest are local buses.

Josh and Rachel ride the bus to work. Josh always takes the first bus that arrives, while Rachel always takes the first express bus. An express bus takes ten minutes to get to the office, while a local bus takes twenty minutes.

Josh has been waiting at the bus stop for four minutes before Rachel arrives.

Calculate Josh's expected total time to get to the office.

Solution

Josh has been waiting at the bus stop for four minutes by the time Rachel arrives. Does this affect his remaining expected wait time for a bus? Due to the memoryless property of exponential random variables, the amount of time that he has already waited does not affect his expected wait time thereafter.

Josh takes the first bus, local or express. This means his expected wait time for a bus is $\theta = 0.25$ hours, or 15 minutes. Since he has already waited four minutes, his expected total wait time is 19 minutes.

Now, how do we calculate the expected time he spends riding the bus? There are two possibilities:

- he spends 10 minutes riding an express bus, or
- he spends 20 minutes riding a local bus.

The probability that he rides an express bus is 0.25, and the probability that he rides a local bus is 0.75. Therefore, he is expected to spend

$$0.25\,(10) + 0.75\,(20) = 17.5$$

minutes on a bus. Thus, his expected total time to the office is $19 + 17.5 = \mathbf{36.5}$ minutes.

■

Joint Probabilities

Let X_1 and X_2 be independent exponential random variables with rates λ_1 and λ_2, respectively. The probability that X_1 is less than X_2 is:

$$\Pr(X_1 < X_2) = \frac{\lambda_1}{\lambda_1 + \lambda_2} \tag{1.1.6.2}$$

The derivation of this is included in the appendix at the end of this section.

We can generalize this for independent exponential random variables X_1, X_2, ..., X_n with rates λ_1, λ_2, ..., λ_n. X_i is the minimum random variable with probability

$$\Pr[X_i = \min(X_1, X_2, \ldots, X_n)] = \frac{\lambda_i}{\sum_{j=1}^{n} \lambda_j}$$

The memoryless property can also be applied to find probabilities involving two independent exponential random variables, X_1 and X_2.

$$\Pr(X_1 - X_2 > a \mid X_1 > X_2) = \Pr(X_1 > a)$$

Minimum of Independent Exponential Random Variables

Let X_1, X_2, ..., X_n be independent exponential random variables with rates λ_1, λ_2, ..., λ_n, respectively.

Then, let Y denote the minimum of X_1, X_2, ..., X_n. The cumulative distribution function of Y can be found using order statistics (covered in Section 2.7).

$$F(y) = 1 - e^{-(\lambda_1 + \lambda_2 + \ldots + \lambda_n)y}$$

Determine from the cumulative distribution function that Y follows an exponential distribution with rate $\lambda_1 + \lambda_2 + \ldots + \lambda_n$.

Thus, the minimum of X_1, X_2, ..., X_n is an exponential random variable with rate $\lambda = \lambda_1 + \lambda_2 + \ldots + \lambda_n$. Note that these exponential random variables do not need to be identical, but they do need to be independent.

Furthermore, Y is independent from the rank ordering of the n random variables, i.e.

$$\Pr(X_1^* < \ldots < X_n^* \mid Y > y) = \Pr(X_1^* < \ldots < X_n^*)$$

where X_1^*, ..., X_n^* are placeholders for any arrangement of X_1, ..., X_n. This is a consequence of the memoryless property. So, the probability of a certain ranking of the n random variables is independent of the location of the minimum random variable.

Example 1.1.6.2

You are given the following:

- A system has three independent components.
- The components have exponential lifetimes with means 3, 4, and 5 years, respectively.
- In order for the whole system to function, all three components must be functioning.

Calculate

1. the variance of the lifetime of the system.
2. the probability that the system lasts at least 3 years.

Solution to (1)

Let Y be the lifetime of the system. If at least one of the three components fails, the whole system fails. Therefore, the lifetime of the system is the minimum of three exponentials. The minimum of three exponentials is an exponential random variable with rate:

$$\lambda = \frac{1}{3} + \frac{1}{4} + \frac{1}{5} = \frac{47}{60}$$

Therefore, the expected value of the lifetime of the system is $\theta = \frac{1}{\lambda} = \frac{60}{47}$, and the variance is:

$$\mathrm{Var}\left[Y\right] = \theta^2 = \mathbf{1.6297}$$

■

Solution to (2)

The probability that the system lasts at least 3 years is:

$$\Pr(Y \geq 3) = e^{-3(47)\,/\,60} = \mathbf{0.09537}$$

■

Alternative Solution to (2)

Let X_1, X_2, and X_3 be the lifetimes of the individual components. Then, the probability that the system lasts 3 years can be found as the probability that X_1, X_2, and X_3 are all greater than 3.

$$\begin{aligned}\Pr(Y \geq 3) &= \Pr[(X_1 \geq 3) \cap (X_2 \geq 3) \cap (X_3 \geq 3)] \\ &= \Pr(X_1 \geq 3) \cdot \Pr(X_2 \geq 3) \cdot \Pr(X_3 \geq 3) \\ &= e^{-3\,/\,3} \cdot e^{-3\,/\,4} \cdot e^{-3\,/\,5} \\ &= \mathbf{0.09537}\end{aligned}$$

■

Sum of Independent and Identical Exponential Random Variables

Let X_1, X_2, ..., X_n be independent and identical exponential random variables each with rate λ. The sum $Y = X_1 + X_2 + \ldots + X_n$ is a gamma random variable with parameters $\alpha = n$ and $\theta = \frac{1}{\lambda}$.

Recall that the exponential distribution is a special case of the gamma distribution with $\alpha = 1$.

1.1.7 Greedy Algorithms

25m

A *greedy algorithm* is an algorithmic problem-solving strategy that involves making optimal choices at each stage with the hope of approximating or reaching the optimal solution to the problem. It is considered greedy because at each stage, the best choice is made without accounting for future choices that could result in a better solution.

For some computational problems, it can be difficult and time-consuming to find the optimal solution. Greedy algorithms apply properties of the exponential distribution to help find a solution in less time; however, there is a trade-off of accuracy and precision.

In this subsection, we will discuss two greedy algorithms. To illustrate how these algorithms work, let's consider the following job assignment scenario.

You are a project manager, and you need to assign a group of employees to available jobs. You have five employees and five jobs.

Let $C_{i,j}$ be the cost of assigning employee i to job j, where both i and j take on integers from 1 to 5. The costs of each job assignment are given below:

$i \backslash j$	1	2	3	4	5
1	11	15	10	13	12
2	18	14	12	13	11
3	14	19	13	16	10
4	9	15	18	12	10
5	12	11	10	9	8

You want to assign these jobs in a way that minimizes the cost of the project.

Greedy Algorithm A

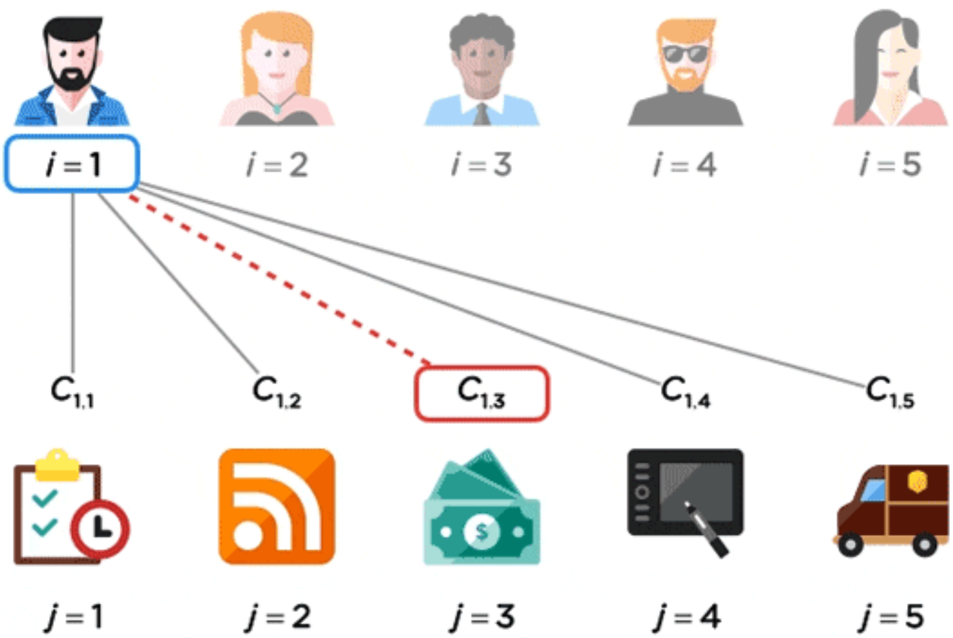
Let's begin by considering Greedy Algorithm A.

For $i = 1, 2, ..., n$:

1. Choose the job assignment with the lowest cost, i.e. $\min_j C_{i,j}$, among all $n - i + 1$ possible assignments.

2. Assign that job to that employee.
3. Remove that employee and that job from their respective sets.

Using this algorithm, we will first assign employee 1 to job 3. Then, assign employee 2 to job 5, and so on. The following animation illustrates this algorithm.



i \ j	1	2	3	4	5
1	11	15	10	13	12
2	18	14	12	13	11
3	14	19	13	16	10
4	9	15	18	12	10
5	12	11	10	9	8

Hence, we end up with this assignment that has a total cost of 58.

Employee	**Job**	**Cost**
1	3	10
2	5	11
3	1	14
4	4	12
5	2	11

Note that for the purposes of illustration, we assume that the actual cost for each job assignment pair is known. What happens if we do not know the exact cost for each assignment, but we only know that the costs follow an exponential distribution?

Let the cost of each job assignment, $C_{i,j}$, be independent and identically distributed exponential random variables with mean θ.

Calculate the expected total cost for the project using Greedy Algorithm A.

The cost for the first assignment is the minimum of n exponentials. Recall that the minimum of n exponential random variables is also an exponential random variable with a rate equal to the sum of the n exponential rates. Also, recall that the rate is the reciprocal of the mean.

Let θ_i be the expected cost of the i^{th} assignment. Thus, the cost for the first assignment follows an exponential distribution with mean

$$\theta_1 = \frac{1}{\frac{1}{\theta} + \frac{1}{\theta} + \ldots + \frac{1}{\theta}} = \frac{1}{\frac{n}{\theta}} = \frac{\theta}{n}$$

The cost for the second assignment is the minimum of $n-1$ exponentials. The expected cost of the second assignment is:

$$\theta_2 = \frac{1}{\frac{1}{\theta} + \frac{1}{\theta} + \ldots + \frac{1}{\theta}} = \frac{1}{\frac{n-1}{\theta}} = \frac{\theta}{n-1}$$

This pattern will continue for the remaining assignments, where the expected cost of the n^{th} assignment will be $\frac{\theta}{1}$. Ultimately, the expected total cost will be:

$$\begin{aligned}\mathrm{E}\left[\text{total cost}\right] &= \frac{\theta}{n} + \frac{\theta}{n-1} + \ldots + \frac{\theta}{1} \\ &= \theta \sum_{i=1}^{n} \frac{1}{i}\end{aligned} \tag{1.1.7.1}$$

Example 1.1.7.1 (Adapted from CAS S S2016 1)

Alice, Bob, and Chris are hired by a firm to drive three vehicles — a van, a truck, and a bus (only one driver is needed for each). Each employee has different skills and requires different amounts of training for each vehicle. The cost to train an employee i for vehicle j, $C_{i,\,j}$, is an independent exponential random variable with mean 150.

To minimize the total cost of training, the firm uses the following procedure to assign the employees to their vehicles:

- Alice is assigned to the vehicle which minimizes her training cost, $C_{Alice,\,j}$.
- Bob is then assigned the vehicle of the two remaining that minimizes $C_{Bob,\,j}$.
- Chris is then assigned the remaining vehicle.

Calculate

1. the expected cost of the last training assignment.
2. the firm's expected total cost of training these three employees.

Solution to (1)

Notice that this assignment uses Greedy Algorithm A. Since Chris is assigned the last vehicle, the cost of the training assignment is exponentially distributed with mean $\mathbf{150}$.

■

Solution to (2)

Using Equation 1.1.7.1, the firm's expected total cost is:

$$\begin{aligned}\mathrm{E}\left[\text{total cost}\right] &= 150\sum_{i=1}^{3}\frac{1}{i} \\ &= 150\left(\frac{1}{3}+\frac{1}{2}+\frac{1}{1}\right) \\ &= \mathbf{275}\end{aligned}$$

$\blacksquare$

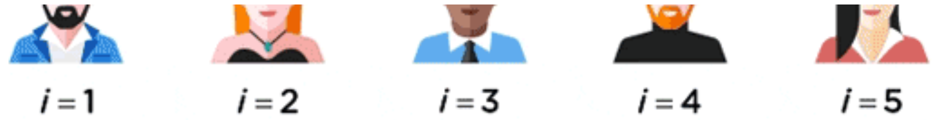
Greedy Algorithm B

Now, let's consider Greedy Algorithm B. Instead of looking at the possible jobs for one employee at a time, we look at all possible pairs of employees and jobs. The algorithm is as follows:

For $k = n^2$, $(n-1)^2$, ..., 1^2:

1. Choose the job assignment with the lowest cost, i.e. $\min\limits_{i,\,j} C_{i,\,j}$, among all k possible assignments.
2. Assign that job to that employee.
3. Remove that employee and that job from their respective sets.

Using this algorithm, we will first assign employee 5 to job 5. Then, assign employee 4 to job 1, and so on. The following animation illustrates this algorithm.



i \ j	1	2	3	4	5
1	11	15	10	13	12
2	18	14	12	13	11
3	14	19	13	16	10
4	9	15	18	12	10
5	12	11	10	9	8

Hence, we end up with this assignment that has a total cost of 59.

Employee	Job	Cost
5	5	8
4	1	9
1	3	10
2	4	13
3	2	19

Coach's Remarks

Notice that the best assignment, i.e. the one that results in the lowest cost, is:

Employee	Job	Cost
1	3	10
2	2	14
3	5	10
4	1	9
5	4	9

with a total cost of 52. Neither Greedy Algorithm A nor Greedy Algorithm B achieves this optimal solution. This illustrates the greedy nature of these two algorithms.

Similar to Greedy Algorithm A, assume that we do not know the actual cost of the job assignments. How do we calculate the expected total cost?

Let the cost of each job assignment, $C_{i,j}$, be independent and identically distributed exponential random variables with mean θ.

Calculate the expected total cost for the project using Greedy Algorithm B.

Because there are n employees and n jobs, there are n^2 possible combinations of employees and jobs. The cost for the first assignment is the minimum of n^2 exponentials. Its expected value is:

$$\frac{\theta}{n^2}$$

What is the expected cost for the second assignment? To solve for the expected cost of the second assignment, we can condition on the second assignment cost being greater than the first assignment cost. Due to the memoryless property, the difference between each remaining job assignment and the first assignment also follows an exponential distribution. Hence, the cost of assignment and the first assignment also follows an exponential distribution. Hence, the cost of the second assignment is the cost of the first job plus the minimum of $(n-1)^2$ exponentials. Thus, its expected value is:

$$\frac{\theta}{n^2}+\frac{\theta}{(n-1)^2}$$

Following the same pattern, the expected value of the third cost is:

$$\frac{\theta}{n^2}+\frac{\theta}{(n-1)^2}+\frac{\theta}{(n-2)^2}$$

A detailed derivation of the expected cost of the k^{th} assignment under Greedy Algorithm B is provided in the appendix at the end of this section.

Therefore, the expected total cost is:

$$\begin{aligned}\mathrm{E}[\text{total cost}] &= \frac{\theta}{n^2}+\left(\frac{\theta}{n^2}+\frac{\theta}{(n-1)^2}\right)+\ldots+\left(\frac{\theta}{n^2}+\frac{\theta}{(n-1)^2}+\ldots+\frac{\theta}{1^2}\right)\\ &= n\left(\frac{\theta}{n^2}\right)+(n-1)\left(\frac{\theta}{(n-1)^2}\right)+\ldots+(1)\left(\frac{\theta}{1^2}\right)\\ &= \frac{\theta}{n}+\frac{\theta}{n-1}+\ldots+\frac{\theta}{1}\\ &= \theta\sum_{i=1}^{n}\frac{1}{i}\end{aligned} \qquad (1$$

Notice that the expected cost of each assignment is different for both greedy algorithms. This is because for Greedy Algorithm A, **a different set of random variables** is analyzed at each step; for Greedy Algorithm B, the analysis is done for subsets of the **same set of random variables** at every step.

Coach's Remarks

Greedy Algorithm B results in the same expected total cost as Greedy Algorithm A. If there is an exam question on Greedy Algorithm B and you need to calculate the expected value of the total cost, you can apply Greedy Algorithm A and get the same result. Note that this is only true when calculating the expected **total** cost; it does not apply when calculating the expected costs of the individual job assignments.

Example 1.1.7.2

Skyler, Allysa, and Simon are three new employees at Company A. They are hired for three different tasks: content writer, videographer, and data analyst. The amount of hours (in hundreds) needed to train an employee for a job follows an exponential distribution with mean $\theta = 2$.

The company decides to minimize the total number of hours needed to train the three employees using the following approach:

- The company looks at all nine possible pairs of employee and job and chooses the pair with the lowest hours needed.
- After eliminating the first employee and the first job, the company looks at the remaining four pairs and chooses the pair with the lowest hours needed.
- The remaining pair is the last job assignment.

Calculate

1. the expected number of hours of the last job assignment.
2. the total expected number of hours for all three employees.

Solution to (1)

Notice that this assignment uses Greedy Algorithm B. The first assignment is a minimum of nine exponentials. The minimum of nine exponentials follows an exponential distribution with mean:

$$\frac{\theta}{n^2} = \frac{2}{3^2} = 0.2222$$

The expected number of hours of the second assignment is:

$$0.2222 + \frac{\theta}{(n-1)^2} = 0.2222 + \frac{2}{(3-1)^2} = 0.7222$$

The expected number of hours of the last assignment is:

$$0.7222 + \frac{\theta}{(n-2)^2} = 0.7222 + \frac{2}{(3-2)^2} = \mathbf{2.7222}$$

■

Solution to (2)

The total expected number of hours is the sum of the expected number of hours for all three employees.

$$0.2222 + 0.7222 + 2.7222 = \mathbf{3.6667}$$

We can also calculate the total expected number of hours using Equation 1.1.7.2.

$$\begin{aligned} \mathrm{E}[\text{total}] &= \theta \sum_{i=1}^{n} \frac{1}{i} \\ &= 2\left(\frac{1}{1} + \frac{1}{2} + \frac{1}{3}\right) \\ &= \mathbf{3.6667} \end{aligned}$$

■

1.1.8 Transformations

35m

Transformation involves creating a new distribution from a random variable (or variables) whose distribution is known. For example, if we know the distribution of X, we can use that to determine the distribution of W, where W is a function of X, i.e. $W = g(X)$.

Common transformations include scaling (multiplying X by a constant), the power transformation (raising X to a power), and the exponential transformation (taking the exponential of X).

In addition, there are three main methods of transformation: the CDF method, the PDF method, and the MGF method. These transformations will mainly apply to continuous distributions.

Scaling

All continuous distributions listed on the exam table (except for lognormal, inverse Gaussian, and log-t) are parameterized such that θ is a *scale parameter*. That means when a random variable is scaled, the new random variable will follow the same distribution with the same parameters except θ will be scaled based on the scaling factor.

For example, if $X \sim \text{Pareto}\ (\alpha,\, \theta)$ and c is a positive constant, then

$$cX \sim \text{Pareto}\ (\alpha,\, c\theta)$$

To scale a lognormal distribution, add the natural log of the scaling factor to μ.

$$X \sim \text{Lognormal}\ \left(\mu,\, \sigma^2\right)$$

$$\Downarrow$$

$$cX \sim \text{Lognormal}\ \left(\mu + \ln c,\, \sigma^2\right)$$

For all other distributions, use one of the methods below to find the distribution of a scaled random variable.

CDF Method

The CDF method can be used for transforming one variable to one variable, i.e. $W = g(X)$, or for transforming two variables to one variable, i.e. $W = g(X_1,\, X_2)$.

For the univariate case, follow these steps:

1. Using the equation of transformation, $W = g(X)$, express

$$F_W(w) = \Pr(W \le w) = \Pr[g(X) \le w]$$

2. Differentiate $F_W(w)$ to get $f_W(w)$ if required.

For the bivariate case, follow these steps:

1. Using the equation of transformation, $W = g(X_1, X_2)$, express

$$F_W(w) = \Pr(W \le w) = \Pr[g(X_1, X_2) \le w]$$

2. Calculate $F_W(w) = \Pr(W \le w)$ by integrating over the region of integration defined by the domain of $f_{X_1, X_2}(x_1, x_2)$ and $g(x_1, x_2) \le w$.
3. Differentiate $F_W(w)$ to get $f_W(w)$ if required.

Example 1.1.8.1

A device containing two key components fails when both components fail. The lifetimes, T_1 and T_2, of those two components are independent with common density function

$$f(t) = e^{-t}, \qquad t > 0$$

The cost, X, of operating the device until failure is $2T_1 + T_2$.

Calculate the density function of X where $x > 0$.

Solution

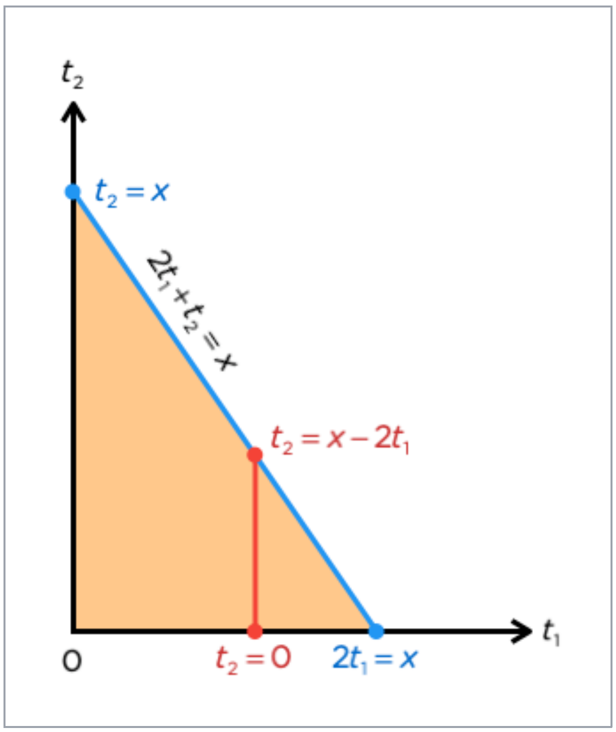
First, find $f_{T_1, T_2}(t_1, t_2)$. Because of the independence of T_1 and T_2, the joint density function is the product of the density functions of T_1 and T_2.

$$\begin{aligned} f_{T_1,T_2}(t_1, t_2) &= f_{T_1}(t_1) \cdot f_{T_2}(t_2) \\ &= e^{-t_1} \cdot e^{-t_2} \\ &= e^{-t_1-t_2}, \qquad t_1 > 0, \quad t_2 > 0 \end{aligned}$$

Next, find $F_X(x)$. Begin with

$$F_X(x) = \Pr(X \le x) = \Pr(2T_1 + T_2 \le x)$$

Then, sketch a diagram of the region described by the inequality $2t_1 + t_2 \le x$ and the boundaries $t_1 > 0$ and $t_2 > 0$, with t_1 and t_2 as the axes.



Since T_1 and T_2 are positive, this results in x being positive as well. Not only is x positive, but it is a positive constant. Notice that regardless of the exact value of x, the region of integration has the same bounds. Thus, we may arbitrarily assign x as any positive constant in sketching the diagram.

The region bounded by $t_1 > 0, t_2 > 0$, and $2t_1 + t_2 \le x$ is shown as the shaded region. The line $2t_1 + t_2 = x$ intersects the vertical axis at $t_2 = x$ and intersects the horizontal axis at $2t_1 = x$. The CDF of X is $\Pr(2T_1 + T_2 \le x)$. So, to calculate the CDF of X, we perform double integration over the shaded region.

In this example, we will integrate with respect to t_2 first, then t_1. Draw an arbitrary vertical line across the region of integration. The intersection points between the vertical line and the edges of the shaded region are the limits of the inner integration. The limits of the outer integration are the leftmost and rightmost values in the region, i.e., $t_1 = 0$ and $t_1 = \frac{x}{2}$

i.e. $t_1 = 0$ and $t_1 = \frac{x}{2}$.

$$
\begin{aligned}
F_X(x) &= \int_0^{x/2}\int_0^{x-2t_1} f_{T_1,T_2}(t_1,t_2)\,\mathrm{d}t_2\,\mathrm{d}t_1 \\
&= \int_0^{x/2}\int_0^{x-2t_1} e^{-t_1-t_2}\,\mathrm{d}t_2\,\mathrm{d}t_1 \\
&= \int_0^{x/2} \left[-e^{-t_1-t_2}\right]_0^{x-2t_1}\mathrm{d}t_1 \\
&= \int_0^{x/2} \left(-e^{-t_1-x+2t_1}+e^{-t_1}\right)\mathrm{d}t_1 \\
&= \int_0^{x/2} \left(e^{-t_1}-e^{-x+t_1}\right)\mathrm{d}t_1 \\
&= \left[-e^{-t_1}-e^{-x+t_1}\right]_0^{x/2} \\
&= -e^{-x/2}-e^{-x/2}+1+e^{-x} \\
&= 1-2e^{-x/2}+e^{-x}, \qquad x>0
\end{aligned}
$$

Finally, the density function is the derivative of the CDF.

$$
\begin{aligned}
f_X(x) &= \frac{\mathrm{d}}{\mathrm{d}x}F_X(x) \\
&= \frac{\mathrm{d}}{\mathrm{d}x}\left(1-2e^{-x/2}+e^{-x}\right) \\
&= -2e^{-x/2}\left(-\frac{1}{2}\right)-e^{-x} \\
&= \boldsymbol{e^{-x/2}-e^{-x}, \quad x>0}
\end{aligned}
$$

■

PDF Method

The PDF method can be used when transforming n random variables into n other variables, i.e. $W_1 = g_1(X_1, X_2, \ldots, X_n)$, ..., $W_n = g_n(X_1, X_2, \ldots, X_n)$, assuming it is a one-to-one transformation. In general, the steps are as follows.

1. Find the inverses of the equations of transformation, $x_1 = h_1(w_1, \ldots, w_n)$, ..., $x_n = h_n(w_1, \ldots, w_n)$.

2. Calculate the determinant of the Jacobian matrix, J.

$$J = \det \begin{bmatrix} \dfrac{\partial x_1}{\partial w_1} & \cdots & \dfrac{\partial x_1}{\partial w_n} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial x_n}{\partial w_1} & \cdots & \dfrac{\partial x_n}{\partial w_n} \end{bmatrix}$$

3. Find $f_{W_1,\ldots,W_n}(w_1, \ldots, w_n)$ using the following formula, for $J \neq 0$. Note that $|J|$ is the absolute value of the determinant.

$$f_{W_1,\ldots,W_n}(w_1, \ldots, w_n) = f_{X_1,\ldots,X_n}[h_1(w_1, \ldots, w_n), \ldots, h_n(w_1, \ldots, w_n)] \cdot |J| \quad (1$$

For the univariate case, this simplifies to the formula

$$f_W(w) = f_X\left[h(w)\right] \cdot \left| \frac{\mathrm{d}}{\mathrm{d}w} h(w) \right| \qquad (1.1.8.2)$$

where $h(w) = g^{-1}(w)$.

Example 1.1.8.2

A random variable X follows a distribution with the CDF

$$F_X(x) = \frac{x}{10}, \qquad 0 \leq x \leq 10$$

You are given another random variable, Y, where

$$Y = e^X$$

Determine the PDF of Y.

Solution

First, calculate the PDF of X.

$$\begin{aligned} f_X(x) &= \frac{\mathrm{d}}{\mathrm{d}x} F_X(x) \\ &= \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{x}{10}\right) \\ &= \frac{1}{10} \end{aligned}$$

Next, find the inverse of the equation of transformation.

$$Y = e^X \quad \Rightarrow \quad g^{-1}(Y) = X = \ln Y$$

Then, find the PDF of Y as $f_Y(y) = f_X\left[g^{-1}(y)\right] \cdot \left|\frac{\mathrm{d}}{\mathrm{d}y} g^{-1}(y)\right|$.

$$\begin{aligned} f_Y(y) &= f_X\left[g^{-1}(y)\right] \cdot \left|\frac{\mathrm{d}}{\mathrm{d}y} g^{-1}(y)\right| \\ &= f_X(\ln y) \cdot \left|\frac{\mathrm{d}}{\mathrm{d}y}(\ln y)\right| \\ &= \frac{1}{10} \cdot \left|\frac{1}{y}\right| \\ &= \boldsymbol{\frac{1}{10y}}, \qquad \boldsymbol{1 \le y \le e^{10}} \end{aligned}$$

The domain of $f_Y(y)$ is the result of the transformation of the domain of $f_X(x)$. The domain of $f_X(x)$ is $0 \le x \le 10$. So, substitute $x = \ln y$ in to get $0 \le \ln y \le 10$. This can be rewritten as $e^0 \le y \le e^{10}$, or equivalently, $1 \le y \le e^{10}$.

■

For the bivariate case, this simplifies to the formula

$$f_{W_1,W_2}(w_1, w_2) = f_{X_1,X_2}[h_1(w_1, w_2), h_2(w_1, w_2)] \cdot |J|, \qquad J \neq 0$$

where J is a two-by-two matrix.

$$J = \det \begin{bmatrix} \dfrac{\partial x_1}{\partial w_1} & \dfrac{\partial x_1}{\partial w_2} \\ \dfrac{\partial x_2}{\partial w_1} & \dfrac{\partial x_2}{\partial w_2} \end{bmatrix}$$

Coach's Remarks

In general, the determinant of a two-by-two matrix is

$$D = \det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc$$

In the bivariate case, we can extend the PDF method to include the transformation of two variables to one variable. Simply create a dummy variable for W_2 before finding the inverses of the equations of transformations. It is typically easiest to set W_2 equal to either X_1 or X_2. Then, after finding the joint PDF of the transformed variables, integrate $f_{W_1,W_2}(w_1, w_2)$ with respect to the dummy variable to get the marginal PDF of $f_{W_1}(w_1)$.

Let's redo Example 1.1.8.1 to see how this works.

Example 1.1.8.3

A device containing two key components fails when both components fail. The lifetimes, T_1 and

T_2, of those two components are independent with common density function

$$f(t) = e^{-t}, \qquad t > 0$$

The cost, X, of operating the device until failure is $2T_1 + T_2$.

Calculate the density function of X where $x > 0$.

Solution

First, we'll need $f_{T_1,T_2}(t_1, t_2)$. We calculated the joint density in Example 1.1.8.1 to be

$$f_{T_1,T_2}(t_1, t_2) = e^{-t_1-t_2}, \qquad t_1 > 0, \quad t_2 > 0$$

Since we are using the PDF method to transform two variables to one variable, introduce a dummy variable. We are given $X = 2T_1 + T_2$. We rename X to X_1 and introduce the dummy variable X_2. We define X_2 as $X_2 = T_2$. Therefore,

$$X_1 = 2T_1 + T_2$$

$$X_2 = T_2$$

Find the inverses of the equations of transformation.

$$X_1 = 2T_1 + T_2 \quad \Rightarrow \quad X_1 = 2T_1 + X_2 \quad \Rightarrow \quad T_1 = \frac{X_1 - X_2}{2} = h_1(X_1, X_2)$$

$$X_2 = T_2 \quad \Rightarrow \quad T_2 = X_2 = h_2(X_1, X_2)$$

For every possible pair (T_1, T_2), there is exactly one corresponding pair (X_1, X_2), and vice versa. Thus, the transformation is one-to-one.

Next, calculate $|J|$.

$$J = \det \begin{bmatrix} \dfrac{\partial t_1}{\partial x_1} & \dfrac{\partial t_1}{\partial x_2} \\ \dfrac{\partial t_2}{} & \dfrac{\partial t_2}{} \end{bmatrix}$$

$$\begin{aligned}
&\lfloor \partial x_1 \qquad \partial x_2 \rfloor \\
&= \det \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ 0 & 1 \end{bmatrix} \\
&= \frac{1}{2}(1) + \frac{1}{2}(0) \\
&= \frac{1}{2}
\end{aligned}$$

So,

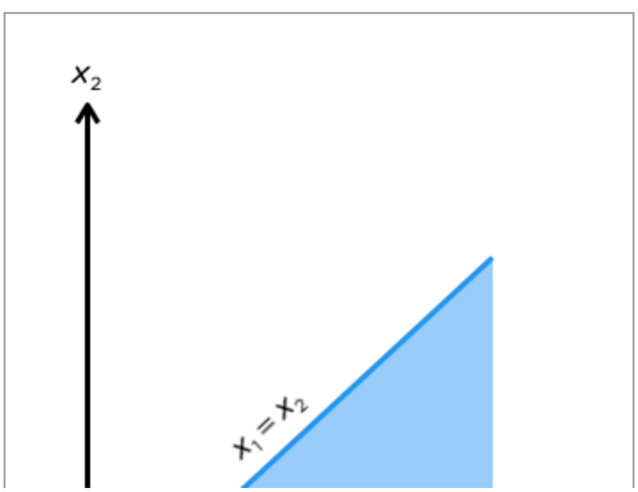
$$|J| = \left| \frac{1}{2} \right| = \frac{1}{2}$$

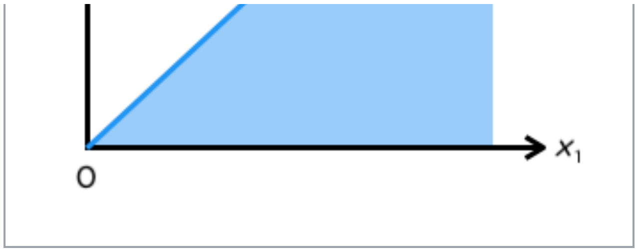
Then, find the joint density of X_1 and X_2.

$$\begin{aligned}
f_{X_1,X_2}(x_1, x_2) &= f_{T_1,T_2}[h_1(x_1, x_2), h_2(x_1, x_2)] \cdot |J| \\
&= f_{T_1,T_2}\left(\frac{x_1 - x_2}{2}, x_2\right) \cdot \frac{1}{2} \\
&= e^{-(x_1-x_2)/2-x_2} \cdot \frac{1}{2} \\
&= \frac{1}{2} e^{-x_1/2-x_2/2}, \qquad 0 < x_2 < x_1 < \infty
\end{aligned}$$

The domain of $f_{X_1,X_2}(x_1, x_2)$ is the result of the transformation of the domain of $f_{T_1,T_2}(t_1, t_2)$. The domain of $f_{T_1,T_2}(t_1, t_2)$ is $t_1 > 0$ and $t_2 > 0$.

Then, substitute $t_1 = \frac{x_1-x_2}{2}$ and $t_2 = x_2$ to get $\frac{x_1-x_2}{2} > 0$ and $x_2 > 0$. This can be written as $x_1 > x_2$ and $x_2 > 0$, or equivalently, $0 < x_2 < x_1 < \infty$.





Finally, find $f_{X_1}(x_1)$ by integrating the joint density function with respect to the dummy variable. Because we renamed X to X_1, this will result in the density function of X.

$$\begin{aligned} f_{X_1}(x_1) &= \int_0^{x_1} f_{X_1,X_2}(x_1,\, x_2)\,\mathrm{d}x_2 \\ &= \int_0^{x_1} \frac{1}{2} e^{-x_1/2 - x_2/2}\,\mathrm{d}x_2 \\ &= \frac{1}{2} e^{-x_1/2}\left[(-2)e^{-x_2/2}\right]_0^{x_1} \\ &= -e^{-x_1/2}\left(e^{-x_1/2} - 1\right) \\ &= e^{-x_1/2} - e^{-x_1}, \qquad x_1 > 0 \end{aligned}$$

Because $t_1 > 0, t_2 > 0$, and $x_1 = 2t_1 + t_2$, the domain is $x_1 > 0$. Rename X_1 to X, and notice that this matches our answer using the CDF method.

$$f_X(x) = \boldsymbol{e^{-x/2} - e^{-x}, \qquad x > 0}$$

$\blacksquare$

Recall that in order to use the PDF method, the transformation should be one-to-one. However, there are certain cases where the PDF method can be applied even if the transformation is not one-to-one. For example, assume we want to find the PDF of $W = X^2$, where X is valid on $-\infty < x < \infty$. Here, $-\infty < x < 0$ and $0 < x < \infty$ both map onto the same values of W. So, to use the PDF method, we need to separate the domain of X into disjoint sets, where $W = X^2$ is a one-to-one transformation for each set. This can be achieved with the two aforementioned disjoint sets, $-\infty < x < 0$ and $0 < x < \infty$. Then, determine the inverses of the equation of transformation for each disjoint set.

- For $-\infty < x < 0$, $x = h(w) = -\sqrt{w}$

- For $-\infty < x < 0, x = h(w) = -\sqrt{w}$.
- For $0 < x < \infty, x = h(w) = \sqrt{w}$.

Finally, calculate the PDF of W as

$$f_W(w) = f_X\left(-\sqrt{w}\right) \cdot \left|\frac{\mathrm{d}}{\mathrm{d}w}\left(-\sqrt{w}\right)\right| + f_X\left(\sqrt{w}\right) \cdot \left|\frac{\mathrm{d}}{\mathrm{d}w}\sqrt{w}\right|$$

We can generalize this for the transformation of n random variables into n other variables, where the transformation is not one-to-one. To make it work, we need to be able to separate the joint domain of X_1, ..., X_n into k mutually disjoint sets, such that the transformation for each set is onto the same domain and is one-to-one.

In that case, to determine the joint PDF of the transformed random variables, first follow the steps of the general PDF method for each disjoint set. Then, sum the joint PDFs from step 3 across all disjoint sets.

MGF Method

The MGF method works best when the transformed random variable is a linear function of n random variables, e.g. $W = X_1 + X_2 + X_3$.

Assume $W = g(X_1, X_2, \ldots, X_n)$ is the equation of transformation and $f(x_1, \ldots, x_n)$ is the joint PDF. Then, the steps for this method are as follows.

1. Find the MGF of W as

$$\mathrm{E}\left[e^{tW}\right] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{tg(x_1, \ldots, x_n)} f(x_1, \ldots, x_n)\,\mathrm{d}x_1 \cdots \mathrm{d}x_n$$

2. Recognize the MGF of W as belonging to a common distribution.

This method is not as useful since it relies on being able to recognize the MGF of W.

1.1.9 Mixtures

30m

Mixtures, or compound distributions, are derived from combining multiple distributions. There are two types of mixtures, discrete and continuous.

Since mixtures involve conditional distributions, we will frequently use the Law of Total Probability (Equation 1.0.6.4), the Law of Total Expectation (Equation 1.0.6.5), and the Law of Total Variance (Equation 1.0.6.6), so make sure to have a good understanding of these formulas.

Discrete Mixtures

A random variable Y is a *discrete mixture* of random variables X_1, X_2, ..., X_n if its probability function can be expressed as a weighted average of the probability functions of these n random variables.

$$\begin{aligned} f_Y(y) &= w_1 \cdot f_{X_1}(y) + w_2 \cdot f_{X_2}(y) + \ldots + w_n \cdot f_{X_n}(y) \\ &= \sum_{i=1}^{n} w_i \cdot f_{X_i}(y) \end{aligned} \tag{1.1.9.1}$$

In the equation above, w_i represents the *weight* of random variable X_i. Often, the weights can be treated as probabilities because they have the following properties:

- $\sum_{i=1}^{n} w_i = 1$
- $0 \leq w_i \leq 1$

Coach's Remarks

The $f_{X_i}(y)$'s in Equation 1.1.9.1 do not necessarily need to be continuous PDFs. They are generic probability functions because mixture distributions can be a combination of discrete and continuous distributions.

The following are a direct consequence of Equation 1.1.9.1.

$$F_Y(y) = \sum_{i=1}^{n} w_i \cdot F_{X_i}(y) \tag{1.1.9.2}$$

$$S_Y(y) = \sum_{i=1}^{n} w_i \cdot S_{X_i}(y)$$

$$\mathrm{E}\left[Y^k\right] = \sum_{i=1}^{n} w_i \cdot \mathrm{E}\left[X_i^k\right] \tag{1.1.9.3}$$

Note that the variance cannot be calculated as a weighted average. Instead, apply the Law of Total Variance.

Coach's Remarks

The Bernoulli shortcut is related to the Bernoulli distribution. It is a technique to quickly calculate the variance of a variable that has only **two** possible values.

In general, for a variable

$$X = \begin{cases} a, & \text{Probability} = q \\ b, & \text{Probability} = 1 - q \end{cases}$$

the variance of X can be calculated as

$$\mathrm{Var}[X] = (a - b)^2 q\,(1 - q)$$

This means when $n = 2$, the "variance of the expected value" component of the Law of Total Variance equals $\left(\mathrm{E}[X_1] - \mathrm{E}[X_2]\right)^2 \cdot w_1 \cdot w_2$.

Keep in mind that a mixture is **not** the same as a *linear combination* of random variables. Assume $n = 2$.

- Then, a mixture means that $f_Y(y) = w_1 \cdot f_{X_1}(y) + w_2 \cdot f_{X_2}(y)$. So, Y follows the distribution of X_1 for $100w_1\%$ of the time, and follows the distribution of X_2 for $100w_2\%$ of the time.
- A linear combination means that $Y = w_1X_1 + w_2X_2$. Here, Y follows neither the distribution of X_1 nor X_2. Instead, its exact value is dependent on the values of both X_1 and X_2. For a linear combination, w_1 and w_2 are not weights. Therefore, they can be any real number and do not have to add to 1.

Example 1.1.9.1

For an automobile insurance coverage issued by company XYZ, policyholders can be categorized into three types of risks: low, medium, and high. The annual loss amount has a mixed exponential distribution.

The average loss per year for each risk class is:

Risk	Average Loss per Year
Low	300
Medium	540
High	800

Suppose 40% of the insureds are low risk and 35% of the insureds are high risk.

For a randomly selected policyholder, calculate

1. the probability that the annual loss is less than 500.
2. the expected annual loss.
3. the standard deviation of the annual loss.

Solution to (1)

Let X be the annual loss size. It is a mixture of 3 exponential distributions, one for each risk class.

Let L, M, and H be low, medium, and high risk indicators, respectively.

Therefore,

- $(X \mid L) \sim \text{Exponential } (300)$, where $\Pr(L) = 0.4$
- $(X \mid M) \sim \text{Exponential } (540)$, where $\Pr(M) = 0.25$
- $(X \mid H) \sim \text{Exponential } (800)$, where $\Pr(H) = 0.35$

The probability that the annual loss is less than 500 is:

$$\begin{aligned}\Pr(X < 500) &= \mathrm{E}[\Pr(X < 500 \mid \text{Risk})] \\ &= 0.4 \Pr(X < 500 \mid L) + 0.25 \Pr(X < 500 \mid M) + 0.35 \Pr(X < 500 \mid H) \\ &= 0.4\left(1 - e^{-500\,/\,300}\right) + 0.25\left(1 - e^{-500\,/\,540}\right) + 0.35\left(1 - e^{-500\,/\,800}\right) \\ &= \mathbf{0.6381}\end{aligned}$$

■

Solution to (2)

The expected annual loss is

$$\begin{aligned}\mathrm{E}[X] &= \mathrm{E}[\mathrm{E}[X \mid \text{Risk}]] \\ &= 0.4\mathrm{E}[X \mid L] + 0.25\mathrm{E}[X \mid M] + 0.35\mathrm{E}[X \mid H] \\ &= 0.4(300) + 0.25(540) + 0.35(800)\end{aligned}$$

$$= \mathbf{535}$$

■

Solution to (3)

To calculate the standard deviation of the annual loss, the Law of Total Variance can be used. However, since $\mathrm{E}[X]$ has already been calculated above, it is more efficient to calculate the second moment of X and compute $\mathrm{E}\left[X^2\right] - \mathrm{E}[X]^2$.

Refer to the exam table for the second moment of the exponential distribution.

$$\begin{aligned}
\mathrm{E}\left[X^2\right] &= \mathrm{E}\left[\mathrm{E}\left[X^2 \mid \mathrm{Risk}\right]\right] \\
&= 0.4\mathrm{E}\left[X^2 \mid L\right] + 0.25\mathrm{E}\left[X^2 \mid M\right] + 0.35\mathrm{E}\left[X^2 \mid H\right] \\
&= 0.4\left(2 \cdot 300^2\right) + 0.25\left(2 \cdot 540^2\right) + 0.35\left(2 \cdot 800^2\right) \\
&= 665{,}800
\end{aligned}$$

$$\begin{aligned}
\mathrm{Var}[X] &= \mathrm{E}\left[X^2\right] - \mathrm{E}[X]^2 \\
&= 665{,}800 - 535^2 \\
&= 379{,}575
\end{aligned}$$

$$\sigma_X = \sqrt{\mathrm{Var}[X]} = \mathbf{616.097}$$

■

Alternative Solution to (3)

For completeness, let's find the the standard deviation of the annual loss using the Law of Total Variance.

$$\mathrm{Var}[X] = \mathrm{E}[\mathrm{Var}[X \mid \mathrm{Risk}]] + \mathrm{Var}[\mathrm{E}[X \mid \mathrm{Risk}]]$$

The first piece is the average of the individual variances.

$$\begin{aligned}\mathrm{E}[\mathrm{Var}[X \mid \mathrm{Risk}]] &= 0.4\mathrm{Var}[X \mid L] + 0.25\mathrm{Var}[X \mid M] + 0.35\mathrm{Var}[X \mid H] \\ &= 0.4 \cdot 300^2 + 0.25 \cdot 540^2 + 0.35 \cdot 800^2 \\ &= 332{,}900\end{aligned}$$

The second piece is the variance of the individual means. Recall that the variance is the average squared deviation from the mean of $\mathrm{E}[X \mid \mathrm{Risk}]$, which we previously found to equal 535. So, the second piece is:

$$\begin{aligned}\mathrm{Var}[\mathrm{E}[X \mid \mathrm{Risk}]] &= 0.4(\mathrm{E}[X \mid L] - 535)^2 + 0.25(\mathrm{E}[X \mid M] - 535)^2 + 0.35(\mathrm{E}[X \mid H] - 535 \\ &= 0.4(300 - 535)^2 + 0.25(540 - 535)^2 + 0.35(800 - 535)^2 \\ &= 46{,}675\end{aligned}$$

Thus,

$$\begin{aligned}\mathrm{Var}[X] &= \mathrm{E}[\mathrm{Var}[X \mid \mathrm{Risk}]] + \mathrm{Var}[\mathrm{E}[X \mid \mathrm{Risk}]] \\ &= 332{,}900 + 46{,}675 \\ &= 379{,}575\end{aligned}$$

$$\sigma_X = \sqrt{\mathrm{Var}[X]} = \mathbf{616.097}$$

Coach's Remarks

In the calculation of $\mathrm{E}\left[X^2\right]$ above, notice that it is the weighted average of the conditional second moments. This is not the same as the weighted average of the squared conditional first moments.

$$\begin{aligned}\mathrm{E}\left[X^2\right] &= \sum_{\text{all risks}} \Pr(\text{Risk}) \cdot \mathrm{E}\left[X^2 \mid \text{Risk}\right] \\ &\neq \sum_{\text{all risks}} \Pr(\text{Risk}) \cdot \mathrm{E}[X \mid \text{Risk}]^2\end{aligned}$$

Continuous Mixtures

A *continuous mixture* is conceptually similar to a discrete mixture. While a discrete mixture combines a countable number of distributions, a continuous mixture combines an uncountable number of distributions. Most continuous mixtures combine conditional distributions of Y, where the parameter follows a continuous distribution and therefore has an uncountable number of possible values.

For instance, if Y follows an exponential distribution with mean λ, where λ follows a Pareto distribution with parameters $\alpha = 3$ and $\theta = 30$, then Y, unconditionally, is a continuous mixture. In mathematical notation,

$$Y \mid \lambda \sim \text{Exponential}\ (\lambda)$$

$$\lambda \sim \text{Pareto}\ (3,\, 30)$$

Example 1.1.9.2

An actuarial student sits for a multiple-choice test. The number of questions answered correctly follows a binomial distribution with parameters $m = 30$ and q, where q varies by the difficulty of the test and is uniformly distributed between 0.7 and 0.9.

Calculate

1. the expected number of questions answered correctly.
2. the probability that the student answers more than 28 questions correctly.

Solution to (1)

Let X be the number of questions answered correctly.

$$X \mid q \sim \text{Binomial}\ (30,\, q)$$

$$q \sim \text{Uniform}\ (0.7,\, 0.9)$$

To determine the mean number of questions answered correctly, apply the Law of Total Expectation.

$$\begin{aligned} \mathrm{E}[X] &= \mathrm{E}_q[\mathrm{E}_X[X \mid q]] \\ &= \mathrm{E}_q[30q] \\ &= 30\mathrm{E}_q[q] \\ &= 30\left(\frac{0.7 + 0.9}{2}\right) \\ &= \mathbf{24} \end{aligned}$$

Solution to (2)

To calculate $\Pr(X > 28)$, first recall that the upper limit of a binomial random variable is its parameter m. In this case, $m = 30$. Therefore, to answer more than 28 questions correctly, the student will have to get 29 or 30 questions correct.

$$\Pr(X > 28) = \Pr(X = 29) + \Pr(X = 30)$$

Calculate $\Pr(X = 29)$ and $\Pr(X = 30)$ as follows:

$$\begin{aligned}
\Pr(X = 29) &= \mathrm{E}_q[\Pr(X = 29 \mid q)] \\
&= \mathrm{E}_q\left[\binom{30}{29} q^{29}(1-q)^1\right] \\
&= \mathrm{E}_q\left[30\left(q^{29} - q^{30}\right)\right] \\
&= 30 \int_{0.7}^{0.9} \left(q^{29} - q^{30}\right) \cdot \frac{1}{0.9 - 0.7}\,\mathrm{d}q \\
&= \frac{30}{0.2}\left[\frac{1}{30}q^{30} - \frac{1}{31}q^{31}\right]_{0.7}^{0.9} \\
&= 0.02731
\end{aligned}$$

$$\begin{aligned}
\Pr(X = 30) &= \mathrm{E}_q[\Pr(X = 30 \mid q)] \\
&= \mathrm{E}_q\left[\binom{30}{30} q^{30}(1-q)^0\right] \\
&= \mathrm{E}_q\left[q^{30}\right] \\
&= \int_{0.7}^{0.9} q^{30} \cdot \frac{1}{0.9 - 0.7}\,\mathrm{d}q \\
&\quad 1 \left[1 \quad {}^{31} \right]^{0.9}
\end{aligned}$$

$$\begin{aligned} &= \frac{}{0.2}\left[\frac{}{31}q^{\cdot}\right]_{0.7} \\ &= 0.00615 \end{aligned}$$

Therefore,

$$\begin{aligned} \Pr(X > 28) &= 0.02731 + 0.00615 \\ &= \mathbf{0.03346} \end{aligned}$$

■

POISSON-GAMMA MIXTURE

While there are hundreds of possible continuous mixtures, the *Poisson-gamma mixture* is a very special one because the mixed distribution is equivalent to a **negative binomial** distribution. Specifically, if

$$X \mid \lambda \sim \text{Poisson}\ (\lambda)$$

$$\lambda \sim \text{Gamma}\ (\alpha,\, \theta)$$

then

$$X \sim \text{Negative Binomial}\ (r = \alpha,\, \beta = \theta)$$

The proof of this relationship is provided in the appendix at the end of this section.

Let's work on an example to see how this relationship helps us solve exam problems.

Example 1.1.9.3

Example 1.1.5.5

Insurance company ABC releases a new insurance product. The company believes that the annual number of claims per insured for this product is Poisson distributed with parameter λ, where λ follows an exponential distribution with variance 16.

Determine the probability that an insured selected at random will file no more than 1 claim next year.

Solution

Let N be the annual number of claims.

Since the exponential distribution has a variance of θ^2, λ is exponentially distributed with parameter $\theta = \sqrt{16} = 4$. Recall that the exponential distribution is a special case of the gamma distribution with $\alpha = 1$. Therefore,

$$N \mid \lambda \sim \text{Poisson}\ (\lambda)$$

$$\lambda \sim \text{Gamma}\ (1,\, 4)$$

Then,

$$N \sim \text{Negative Binomial}\ (1,\, 4)$$

A negative binomial distribution with $r = 1$ is a geometric distribution. Therefore, we can apply the geometric formula from the exam table to calculate the probabilities.

$$\begin{aligned} \Pr(N \le 1) &= \Pr(N = 0) + \Pr(N = 1) \\ &= \frac{1}{1+4} + \frac{4}{(1+4)^{1+1}} \\ &= \mathbf{0.36} \end{aligned}$$

■

Coach's Remarks

Without knowing the negative binomial shortcut, the probability can still be calculated using the Law of Total Probability.

$$\begin{aligned}\Pr(N \le 1) &= \int_0^\infty \Pr(N \le 1 \mid \lambda) \cdot f_\lambda(\lambda) \; \mathrm{d}\lambda \\ &= \int_0^\infty \left(\frac{e^{-\lambda}\lambda^0}{0!} + \frac{e^{-\lambda}\lambda^1}{1!} \right) \cdot \frac{1}{4} e^{-\lambda/4} \; \mathrm{d}\lambda\end{aligned}$$

While first principles will always work, recognizing that a Poisson-gamma mixture is equivalent to a negative binomial distribution will remove the need to perform tedious calculations.

EXPONENTIAL-INVERSE GAMMA MIXTURE

The exponential-inverse gamma mixture is equivalent to a Pareto distribution. If

$$X \mid \lambda \sim \text{Exponential}\;(\lambda)$$

$$\lambda \sim \text{Inverse Gamma}\;(\alpha, \theta)$$

where λ is the mean of $X \mid \lambda$, then

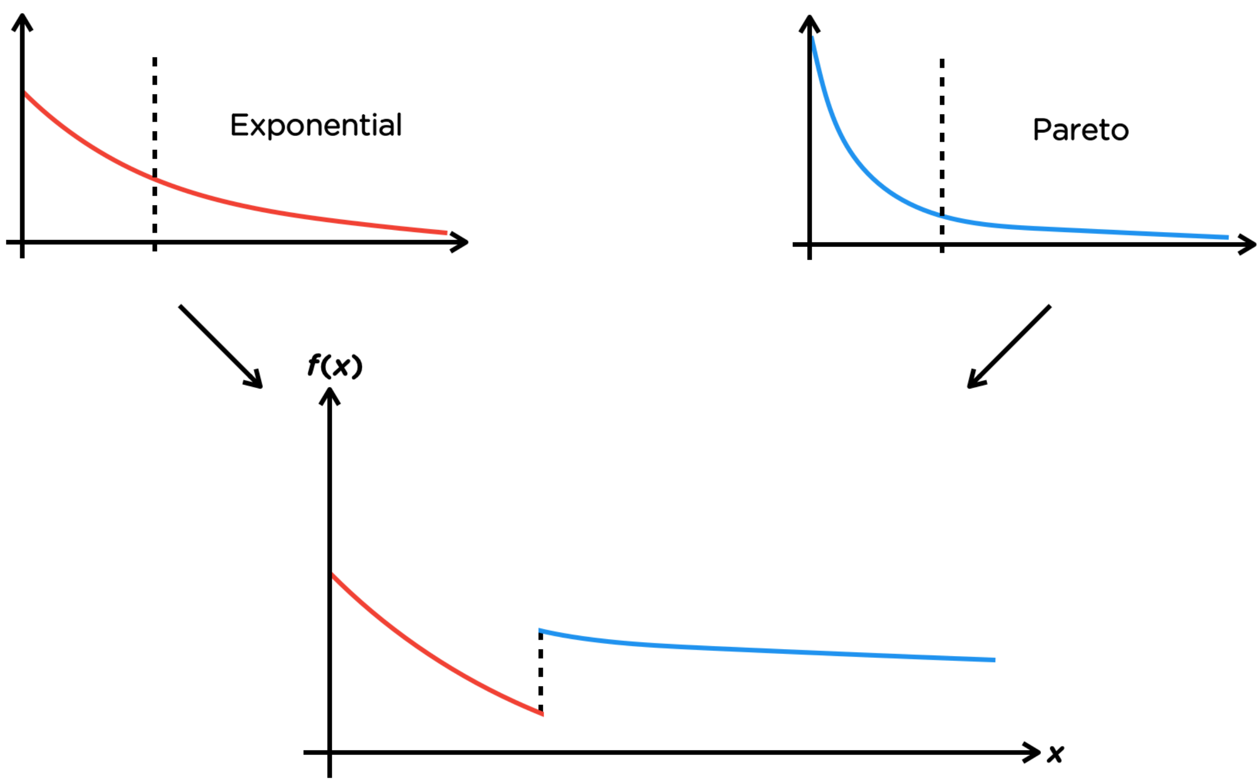
$$X \sim \text{Pareto}\;(\alpha, \theta)$$

1.1.10 Splices

20m

The idea of splicing is to combine different probability distributions over different intervals to produce a new distribution.

For example, Coaching Actuaries observes that its monthly revenue follows a unique pattern. For smaller revenues, the density function appears to be exponential; for larger revenues, the density function appears to be Pareto. Therefore, to create a distribution for the monthly revenue, Coaching Actuaries builds a model where smaller values correspond to an exponential distribution, while larger values correspond to a Pareto distribution. This is an example of a *spliced distribution*.



In general, a random variable Y has a spliced distribution if it has a probability function that can be expressed as:

$$f_Y(y) = \begin{cases} c_1 \cdot f_{X_1}(y), & a_0 < y < a_1 \\ c_2 \cdot f_{X_2}(y), & a_1 < y < a_2 \\ \vdots & \vdots \end{cases}$$

$$c_n \cdot f_{X_n}(y), \quad a_{n-1} < y < a_n$$

Note that the c_i's are not weights, so they do **not** have to sum to 1. However, these coefficients are non-negative because a valid PDF cannot be negative. The c_i's are chosen such that the probability function of Y integrates/sums to 1.

Coach's Remarks

A continuous density function is **not** necessary for a spliced distribution. To be clear, this is about the density function (continuous versus discontinuous); not the distribution (continuous versus discrete). The figure above shows a valid discontinuous density function, with discontinuity where the two distributions are joined. Do not assume a continuous density function for spliced distributions unless stated in the question.

Let's go over a couple examples to learn how to solve splicing problems.

Example 1.1.10.1

Keith models the monthly revenue of his business using a spliced distribution.

- The density function of revenue less than 2,000 is a multiple of the probability density function of an exponential distribution with mean 1,000.
- The density function of revenue above 2,000 is proportional to the probability density function of an exponential distribution with mean 3,000.

The density function of the spliced model is continuous.

Determine the probability that monthly revenue is greater than 3,000, according to Keith's model.

Solution

Let X be Keith's modeled monthly revenue.

First, set up the PDF of X.

$$f_X(x) = \begin{cases} c_1\left(\dfrac{1}{1{,}000}e^{-x\,/\,1{,}000}\right), & x < 2{,}000 \\ c_2\left(\dfrac{1}{3{,}000}e^{-x\,/\,3{,}000}\right), & x \geq 2{,}000 \end{cases}$$

Notice the exponential PDFs are multiplied by constants c_1 and c_2. Both "a multiple of" and "proportional to" in the question mean a constant factor should be applied to the density function of the splice.

In order for $f_X(x)$ to be a valid PDF, it must integrate to 1.

$$1 = \int_0^{2{,}000} c_1\left(\frac{1}{1{,}000}e^{-x\,/\,1{,}000}\right)\mathrm{d}x + \int_{2{,}000}^{\infty} c_2\left(\frac{1}{3{,}000}e^{-x\,/\,3{,}000}\right)\mathrm{d}x$$

$$\begin{aligned} 1 &= c_1\left[-e^{-x\,/\,1{,}000}\right]_0^{2{,}000} + c_2\left[-e^{-x\,/\,3{,}000}\right]_{2{,}000}^{\infty} \\ &= c_1\left(1 - e^{-2}\right) + c_2\left(e^{-2\,/\,3}\right) \end{aligned}$$

$$c_1 = \frac{1 - c_2\left(e^{-2\,/\,3}\right)}{1 - e^{-2}}$$

Since the density of the spliced distribution is continuous, both splices of the distribution must meet at the splicing point. In other words, the PDF at

$x = 2{,}000$ must be equal for both splices. Therefore,

$$c_1\left(\frac{1}{1{,}000}e^{-2{,}000\,/\,1{,}000}\right) = c_2\left(\frac{1}{3{,}000}e^{-2{,}000\,/\,3{,}000}\right)$$

Substitute the expression for c_1 and solve for c_2.

$$\frac{1 - c_2\left(e^{-2\,/\,3}\right)}{1 - e^{-2}}\left(\frac{1}{1{,}000}e^{-2}\right) = c_2\left(\frac{1}{3{,}000}e^{-2\,/\,3}\right)$$

$$3e^{-2}\left[1 - c_2 e^{-2\,/\,3}\right] = \left(1 - e^{-2}\right)c_2 e^{-2\,/\,3}$$

$$c_2 = \frac{3e^{-2}}{e^{-2\,/\,3}(1 - e^{-2} + 3e^{-2})} = 0.6223$$

The value of c_1 is not needed because $\Pr(X > 3{,}000)$ does not depend on c_1.

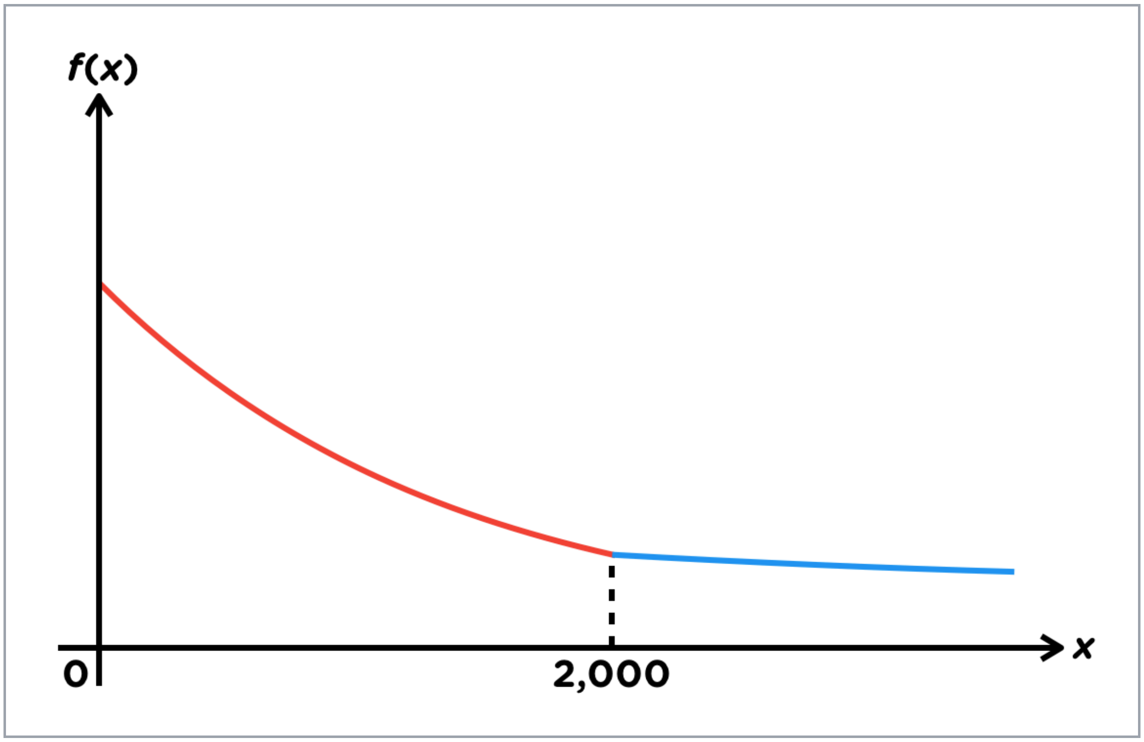
$$\begin{aligned}\Pr(X > 3{,}000) &= \int_{3{,}000}^{\infty} 0.6223\left(\frac{1}{3{,}000}e^{-x\,/\,3{,}000}\right)\mathrm{d}x \\ &= 0.6223\left[-e^{-x\,/\,3{,}000}\right]_{3{,}000}^{\infty} \\ &= 0.6223\left(e^{-1}\right) \\ &= \mathbf{0.2289}\end{aligned}$$

■

Coach's Remarks

Coach's Remarks

The problem specifies that the density function of the spliced model is continuous. Therefore, both functions "meet" at the splicing point and will produce the same value of $f(2{,}000)$, as shown in the graph below:



However, as previously mentioned, you should not assume this is true unless it is stated in the question. Do not make blind assumptions!

Example 1.1.10.2

Larry models his monthly expenses using a spliced distribution.

- Expenses less than 500 are uniformly distributed over $[0,\ 500]$.

- Expenses greater than 500 are modeled using a density function that is identical to the exponential PDF with parameter $\theta = 600$.

Calculate the expected value of Larry's monthly expenses.

Solution

Let X be Larry's modeled monthly expenses.

Set up the PDF of X using the information provided in the question.

$$f_X(x) = \begin{cases} c\left(\dfrac{1}{500}\right), & x < 500 \\ \dfrac{1}{600}e^{-x\,/\,600}, & x \geq 500 \end{cases}$$

For $x < 500$, notice that although the expenses are "uniform over $[0,\ 500]$", the PDF of X is $c\left(\frac{1}{500}\right)$ instead of just $\frac{1}{500}$. Without the constant c, the PDF of X from 0 to 500 would integrate to 1, leaving no room for the exponential splice.

For $x \geq 500$, the question specifies that the density function of X in this range is **identical** to the PDF of an exponential with mean 600. Therefore, there is no need to multiply the exponential PDF by a constant.

Recall that in order for $f_X(x)$ to be a valid PDF, it must integrate to 1. Therefore,

$$1 = \int_0^{500} c\left(\frac{1}{500}\right)\mathrm{d}x + \int_{500}^{\infty} \frac{1}{600}e^{-x\,/\,600}\,\mathrm{d}x$$

$$\begin{aligned} 1 &= c + \left[e^{-500\,/\,600}\right] \\ &= c + 0.4346 \end{aligned}$$

$$1 - 0.4346$$

$$\begin{aligned} c &= 1 - 0.4346 \\ &= 0.5654 \end{aligned}$$

Therefore,

$$f_X(x) = \begin{cases} 0.5654\left(\dfrac{1}{500}\right), & x < 500 \\ \dfrac{1}{600}e^{-x\,/\,600}, & x \geq 500 \end{cases}$$

$$\begin{aligned} \mathrm{E}[X] &= \int_0^{500} 0.5654\left(\frac{x}{500}\right)\mathrm{d}x + \int_{500}^{\infty} \frac{x}{600}e^{-x\,/\,600}\,\mathrm{d}x \\ &= \frac{0.5654}{500}\left[\frac{x^2}{2}\right]_0^{500} + \left[-xe^{-x\,/\,600} - 600e^{-x\,/\,600}\right]_{500}^{\infty} \\ &= 141.35 + 478.058 \\ &= \mathbf{619.408} \end{aligned}$$

■

Coach's Remarks

Notice the question could have left out the range of the uniformly distributed splice, i.e. the first bullet point could have been:

- Expenses less than 500 are uniformly distributed.

Since the word "uniform" implies a **constant** density function, we could have set up the PDF of X as:

$$f_X(x) = \begin{cases} k, & x < 500 \\ \dfrac{1}{600}e^{-x\,/\,600}, & x \geq 500 \end{cases}$$

$$\frac{\quad}{600}e^{-x/600}, \quad x \geq 500$$

where k is a constant, and it would not have affected the solution in any way. In fact, k is numerically equivalent to $c\left(\frac{1}{500}\right)$.

Coach's Remarks

As shown in the examples above, it is important to pay attention to the wording of the question regarding how the spliced distribution is constructed. Focus on terms that are commonly used for splicing questions, such as:

- The PDF ... is **proportional to** ...
- The PDF ... is **a multiple of** ...
- The PDF ... is **identical to** ...
- The density function of the spliced distribution is **continuous**.

1.1 Summary

15m

Common Distributions

PARETO

The S-P Pareto distribution is a Pareto distribution shifted rightwards by θ. Thus, its mean is θ greater than the Pareto's mean. They have the same variance.

BETA

A beta distribution with parameters $a = b = 1$ and θ is equivalent to a uniform distribution on the interval $[0,\ \theta]$.

NORMAL

Distribution	CDF	Mean	Variance	Percentile
Standard normal, Z	$\Phi(z)$	0	1	$z_q = \Phi^{-1}(q)$
Normal, Y	$\Phi\left(\dfrac{y-\mu}{\sigma}\right)$	μ	σ^2	$\mu + z_q\sigma$
Lognormal, X	$\Phi\left(\dfrac{\ln x-\mu}{\sigma}\right)$	$e^{\mu+0.5\sigma^2}$	$\mathrm{E}[X]^2\left(e^{\sigma^2}-1\right)$	$e^{\mu+z_q\sigma}$

A lognormal distribution is a log-transformed normal distribution:

$$X = e$$

GAMMA

- The exponential distribution is a special case of the gamma distribution with $\alpha = 1$.
- For $X \sim \text{Gamma}\,(\alpha, \theta)$ where α is a positive integer,

$$F_X(x) = 1 - \Pr(N < \alpha)$$

where $N \sim \text{Poisson}\left(\lambda = \dfrac{x}{\theta}\right)$.

INVERSE DISTRIBUTIONS

If a random variable X follows a certain distribution that has a parameter θ and an inverse counterpart, then $1\,/\,X$ follows the inverse counterpart with the same parameters, except θ is inverted.

Properties of Exponential Distribution

- An exponential distribution has a mean of θ and a constant failure rate of $\lambda = \frac{1}{\theta}$.
- The memoryless property states that for an exponential random variable X, the probability of $X > t + s$ given that $X > t$ is the same as the probability of $X > s$.
- Let X_1 and X_2 be independent exponential random variables with rate functions λ_1 and λ_2, respectively. Then:

$$\lambda$$

$$\Pr(X_1 < X_2) = \frac{\lambda_1}{\lambda_1 + \lambda_2}$$

- Let $X_1, X_2, \ldots, X_n$ be independent exponential random variables with rate functions $\lambda_1, \lambda_2, \ldots, \lambda_n$. Then, the minimum of $X_1, X_2, \ldots, X_n$ is an exponential random variable with rate function $\lambda = \lambda_1 + \lambda_2 + \cdots + \lambda_n$.

- Let $X_1, X_2, \ldots, X_n$ be independent and identical exponential random variables with rate function λ. Then, the sum $X_1 + X_2 + \cdots + X_n$ is a gamma random variable with parameters $\alpha = n$ and $\theta = \frac{1}{\lambda}$.

Greedy Algorithms

- Algorithm A

  For $i = 1$, 2, ..., n:

  1. Choose the job assignment with the lowest cost, i.e. $\min\limits_{j} C_{i,j}$, among all $n - i + 1$ possible assignments.
  2. Assign that job to that employee.
  3. Remove that employee and that job from their respective sets.

- Algorithm B

  For $k = n^2$, $(n-1)^2$, ..., 1^2:

  1. Choose the job assignment with the lowest cost, i.e. $\min\limits_{i,j} C_{i,j}$, among all k possible assignments.
  2. Assign that job to that employee.
  3. Remove that employee and that job from their respective sets.

- The **total** expected cost using either greedy algorithm for n employees and n jobs, where the cost for each pair is exponential with mean θ, is:

$$\mathrm{E}\left[\text{total cost}\right] = \theta \sum_{i=1}^{n} \frac{1}{i}$$

Transformation

SCALING

All continuous distributions listed on the exam table (except for lognormal, inverse Gaussian, and log-t) are parameterized such that θ is a scale parameter. So, the new random variable will follow the same distribution with the same parameters except θ will be scaled based on the scaling factor.

To scale a lognormal distribution, add the natural log of the scaling factor to μ.

CDF METHOD

For the univariate case, follow these steps:

1. Using the equation of transformation, $W = g(X)$, express $F_W(w) = \Pr(W \le w) = \Pr[g(X) \le w]$.
2. Differentiate $F_W(w)$ to get $f_W(w)$ if required.

For the bivariate case, follow these steps:

1. Using the equation of transformation, $W = g(X_1,\, X_2)$, express $F_W(w) = \Pr(W \le w) = \Pr[g(X_1,\, X_2) \le w]$.
2. Calculate $F_W(w) = \Pr(W \le w)$ by integrating over the region of integration defined by the domain of $f_{X_1,\, X_2}(x_1,\, x_2)$ and $g(x_1,\, x_2) \le w$.
3. Differentiate $F_W(w)$ to get $f_W(w)$ if required.

PDF METHOD

1. Find the inverses of the equations of transformation, $x_1 = h_1(w_1, \ldots, w_n)$, ..., $x_n = h_n(w_1, \ldots, w_n)$.

2. Calculate the determinant of the Jacobian matrix, J, and take its absolute value.

$$J = \begin{vmatrix} \dfrac{\partial x_1}{\partial w_1} & \cdots & \dfrac{\partial x_1}{\partial w_n} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial x_n}{\partial w_1} & \cdots & \dfrac{\partial x_n}{\partial w_n} \end{vmatrix}$$

3. Find $f_{W_1,\ldots,W_n}(w_1, \ldots, w_n)$ using the following formula, for $J \neq 0$.

$$f_{W_1,\ldots,W_n}(w_1, \ldots, w_n) = f_{X_1,\ldots,X_n}[h_1(w_1, \ldots, w_n), \ldots, h_n(w_1, \ldots, w_n)] \cdot |J|$$

MGF METHOD

1. Find the MGF of W as

$$\mathrm{E}\left[e^{tW}\right] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{tg(x_1,\ldots,x_n)} f(x_1, \ldots, x_n)\,\mathrm{d}x_1 \cdots \mathrm{d}x_n$$

2. Recognize the MGF of W as belonging to a common distribution.

Mixtures

A random variable Y is a discrete mixture of the random variables X_1, X_2, ..., X_n if its PDF is given by:

$$f_Y(y) = \sum_{i=1}^{n} w_i \cdot f_{X_i}(y),\ \text{where}\ \sum_{i=1}^{n} w_i = 1$$

The following equations are also true.

$$F_Y(y) = \sum_{i=1}^{n} w_i \cdot F_{X_i}(y)$$

$$S_Y(y) = \sum_{i=1}^{n} w_i \cdot S_{X_i}(y)$$

$$\mathrm{E}\left[Y^k\right] = \sum_{i=1}^{n} w_i \cdot \mathrm{E}\left[X_i^k\right]$$

A continuous mixture is defined by a probability function conditioned on a parameter that is not constant, but instead follows a continuous distribution. A popular continuous mixture is the Poisson-gamma mixture.

$$X \mid \lambda \sim \text{Poisson}\ (\lambda)$$

$$\lambda \sim \text{Gamma}\ (\alpha, \theta)$$

The unconditional distribution of X is negative binomial:

$$X \sim \text{Negative Binomial}\ (r = \alpha, \beta = \theta)$$

Splices

A random variable Y has a spliced distribution if its probability function can be expressed as:

$$f_Y(y) = \begin{cases} c_1 \cdot f_{X_1}(y), & a_0 < y < a_1 \\ c_2 \cdot f_{X_2}(y), & a_1 < y < a_2 \\ \vdots & \vdots \\ c_n \cdot f_{X_n}(y), & a_{n-1} < y < a_n \end{cases}$$

Note that the c_i's may not sum up to 1.

To calculate the c_i's,

- integrate/sum probability function and equate to 1
- set PDF at splicing point equal for both splices if density function is continuous

Appendix

25m

Inverse Distributions

PARETO AND INVERSE PARETO

For X following a Pareto distribution with parameters α and θ, and $Y = X^{-1}$, using the CDF transformation method,

$$\begin{aligned}F_Y(y) &= \Pr(Y \le y)\\&= \Pr\left(X^{-1} \le y\right)\\&= \Pr\left(X \ge y^{-1}\right)\\&= S_X\left(y^{-1}\right)\\&= \left(\frac{\theta}{y^{-1}+\theta}\right)^{\alpha}\\&= \left(\frac{\theta y}{1+\theta y}\right)^{\alpha}\\&= \left(\frac{y}{\theta^{-1}+y}\right)^{\alpha}\end{aligned}$$

which resembles the CDF of an inverse Pareto distribution with parameters α and θ^{-1}.

In conclusion,

$$X \sim \text{Pareto}\,(\alpha,\,\theta)$$

$$\Downarrow$$

$$X^{-1} \sim \text{Inverse Pareto}\,\left(\alpha,\,\theta^{-1}\right)$$

GAMMA AND INVERSE GAMMA

For X following a gamma distribution with parameters α and θ, and $Y = X^{-1}$,

$$
\begin{aligned}
F_Y(y) &= \Pr\left(Y \le y\right) \\
&= \Pr\left(X^{-1} \le y\right) \\
&= \Pr\left(X \ge y^{-1}\right) \\
&= S_X\left(y^{-1}\right) \\
&= 1 - \Gamma\left(\alpha;\ \frac{1}{\theta y}\right) \\
&= 1 - \Gamma\left(\alpha;\ \frac{\theta^{-1}}{y}\right)
\end{aligned}
$$

which resembles the CDF of an inverse gamma distribution with parameters α and θ^{-1}.

In conclusion,

$$X \sim \text{Gamma}\left(\alpha,\ \theta\right)$$

$$\Downarrow$$

$$X^{-1} \sim \text{Inverse Gamma}\left(\alpha,\ \theta^{-1}\right)$$

EXPONENTIAL AND INVERSE EXPONENTIAL

For X following an exponential distribution with mean θ, and $Y = X^{-1}$,

$$
\begin{aligned}
F_Y(y) &= \Pr\left(Y \le y\right) \\
&= \Pr\left(X^{-1} \le y\right) \\
&= \Pr\left(X \ge y^{-1}\right) \\
&= S_X\left(y^{-1}\right) \\
&= e^{-y^{-1}/\theta} \\
&= e^{-\theta^{-1}/y}
\end{aligned}
$$

$$= e^{-\theta^{-1}/y}$$

which resembles the CDF of an inverse exponential distribution with parameter θ^{-1}.

In conclusion,

$$X \sim \text{Exponential}\,(\theta)$$

$$\Downarrow$$

$$X^{-1} \sim \text{Inverse Exponential}\,\left(\theta^{-1}\right)$$

WEIBULL AND INVERSE WEIBULL

For X following a Weibull distribution with parameters θ and τ, and $Y = X^{-1}$,

$$\begin{aligned} F_Y(y) &= \Pr\,(Y \le y) \\ &= \Pr\left(X^{-1} \le y\right) \\ &= \Pr\left(X \ge y^{-1}\right) \\ &= S_X\left(y^{-1}\right) \\ &= e^{-\left(y^{-1}/\theta\right)^{\tau}} \\ &= e^{-\left(\theta^{-1}/y\right)^{\tau}} \end{aligned}$$

which resembles the CDF of an inverse Weibull distribution with parameters θ^{-1} and τ.

In conclusion,

$$X \sim \text{Weibull}\,(\theta,\,\tau)$$

$$\Downarrow$$

$$X^{-1} \sim \text{Inverse Weibull}\,\left(\theta^{-1},\,\tau\right)$$

Poisson Shortcut for Incomplete Gamma Function

For α is a positive integer,

$$\begin{aligned}
\Gamma(\alpha;\, x) &= \frac{1}{\Gamma(\alpha)} \int_0^x t^{\alpha-1} e^{-t}\, \mathrm{d}t \\
&= \frac{1}{(\alpha-1)!} \left[-t^{\alpha-1} e^{-t} - (\alpha-1) t^{\alpha-2} e^{-t} - \ldots - (\alpha-1)! e^{-t}\right]_0^x \\
&= \frac{1}{(\alpha-1)!} \left[-x^{\alpha-1} e^{-x} - (\alpha-1) x^{\alpha-2} e^{-x} - \ldots - (\alpha-1)! e^{-x} + (\alpha-1)!\right] \\
&= 1 - \left[\frac{x^{\alpha-1} e^{-x}}{(\alpha-1)!} + \frac{x^{\alpha-2} e^{-x}}{(\alpha-2)!} + \ldots + e^{-x}\right] \\
&= 1 - \Pr(N < \alpha)
\end{aligned}$$

where $N \sim \text{Poisson}\ (\lambda = x)$.

Exponential Joint Probabilities

Let X_1 and X_2 be independent exponential random variables with rates λ_1 and λ_2, respectively. To find the probability that X_1 is less than X_2, integrate the joint density function of X_1 and X_2. Because X_1 and X_2 are independent, the joint density function is simply the product of the individual density functions.

$$\begin{aligned}
\Pr\left(X_1 < X_2\right) &= \int_0^\infty \int_0^{x_2} \lambda_1 e^{-\lambda_1 x_1} \lambda_2 e^{-\lambda_2 x_2}\, \mathrm{d}x_1\, \mathrm{d}x_2 \\
&= \int_0^\infty \lambda_2 e^{-\lambda_2 x_2} \int_0^{x_2} \lambda_1 e^{-\lambda_1 x_1}\, \mathrm{d}x_1\, \mathrm{d}x_2
\end{aligned}$$

$$
\begin{aligned}
&= \int_0 \lambda_2 e \quad \int_0 \lambda_1 e \quad \mathrm{d}x_1 \, \mathrm{d}x_2 \\
&= \int_0^\infty \lambda_2 e^{-\lambda_2 x_2} \left(1 - e^{-\lambda_1 x_2}\right) \mathrm{d}x_2 \\
&= \int_0^\infty \lambda_2 e^{-\lambda_2 x_2} \, \mathrm{d}x_2 - \int_0^\infty \lambda_2 e^{-(\lambda_2 + \lambda_1) x_2} \, \mathrm{d}x_2 \\
&= 1 - \frac{\lambda_2}{\lambda_1 + \lambda_2} \\
&= \frac{\lambda_1}{\lambda_1 + \lambda_2}
\end{aligned}
$$

Expected Cost for Greedy Algorithm B

Let:

- $C_{i,j}$ be the cost of assigning employee i to job j,
- $C_{i,j}$ be independent and identically distributed exponential random variables with mean θ, and
- C_k be the cost of the k^{th} assignment under Greedy Algorithm B.

Under Greedy Algorithm B, the first assignment is the minimum of n^2 exponential random variables, which also follows an exponential distribution. Hence, the expected value of the first assignment is:

$$
C_1 = \min_{i,j} C_{i,j} \sim \text{Exponential}\left(\frac{\theta}{n^2}\right)
$$

$$
\mathrm{E}\left[C_1\right] = \frac{\theta}{n^2}
$$

After the first assignment is made, there are $(n-1)^2$ possible $C_{i,j}$'s under consideration for the second assignment, all of which must be greater than C_1. Due to the memoryless property, the difference between each remaining cost and C_1 also follows an exponential distribution:

distribution:

$$C_{i,j} - C_1 \mid C_{i,j} > C_1 \sim \text{Exponential}(\theta)$$

It should be no surprise that the minimum of those $(n-1)^2$ possible differences also follows an exponential distribution:

$$\min_{i,j \notin A_1} (C_{i,j} - C_1 \mid C_{i,j} > C_1) \sim \text{Exponential}\left(\frac{\theta}{(n-1)^2}\right)$$

Note that A_k represents the set of $\{i, j\}$ pairs chosen up to the k^{th} assignment.

Since we choose the minimum of the remaining $(n-1)^2$ costs for the second assignment, we can express the second assignment as:

$$C_2 = C_1 + \min_{i,j \notin A_1} (C_{i,j} - C_1 \mid C_{i,j} > C_1)$$

Hence, the expected value of the second assignment is:

$$\begin{aligned}
\mathrm{E}[C_2] &= \mathrm{E}[C_1] + \mathrm{E}\left[\min_{i,j \notin A_1} (C_{i,j} - C_1 \mid C_{i,j} > C_1)\right] \\
&= \frac{\theta}{n^2} + \frac{\theta}{(n-1)^2}
\end{aligned}$$

We can generalize this result for the k^{th} assignment:

$$\begin{aligned}
\mathrm{E}[C_k] &= \mathrm{E}[C_{k-1}] + \mathrm{E}\left[\min_{i,j \notin A_{k-1}} (C_{i,j} - C_{k-1} \mid C_{i,j} > C_{k-1})\right], \qquad k = 2, 3, \ldots, n \\
&= \left[\frac{\theta}{n^2} + \frac{\theta}{(n-1)^2} + \ldots + \frac{\theta}{(n-k+2)^2}\right] + \frac{\theta}{(n-k+1)^2}
\end{aligned}$$

Poisson-Gamma Mixture

Let:

- $N \mid \lambda \sim \text{Poisson}\ (\lambda)$

$$\Pr(N = n \mid \Lambda = \lambda) = \frac{e^{-\lambda}\lambda^n}{n!}$$

- $\Lambda \sim \text{Gamma}\ (\alpha,\, \theta)$

$$f_\Lambda(\lambda) = \frac{(\lambda/\theta)^\alpha e^{-\lambda/\theta}}{\lambda \cdot \Gamma(\alpha)} = \frac{\lambda^{\alpha-1}e^{-\lambda/\theta}}{\theta^\alpha \cdot \Gamma(\alpha)}$$

Solve for the unconditional probability of $N = n$:

$$\begin{aligned}
\Pr(N = n) &= \mathrm{E}_\Lambda[\Pr(N = n \mid \Lambda)] \\
&= \int_0^\infty \Pr(N = n \mid \Lambda = \lambda) \cdot f_\Lambda(\lambda)\, \mathrm{d}\lambda \\
&= \int_0^\infty \frac{e^{-\lambda}\lambda^n}{n!} \cdot \frac{\lambda^{\alpha-1}e^{-\lambda/\theta}}{\theta^\alpha \cdot \Gamma(\alpha)}\, \mathrm{d}\lambda \\
&= \frac{1}{n!} \cdot \frac{1}{\theta^\alpha \cdot \Gamma(\alpha)} \int_0^\infty \lambda^{n+\alpha-1} e^{-\lambda[(1+\theta)/\theta]}\, \mathrm{d}\lambda
\end{aligned}$$

Focus on the integral:

$$\begin{aligned}
\int_0^\infty \lambda^{n+\alpha-1} e^{-\lambda[(1+\theta)/\theta]}\, \mathrm{d}\lambda &= \int_0^\infty \frac{\lambda^{n+\alpha} e^{-\lambda/[\theta/(1+\theta)]}}{\lambda} \cdot \frac{[\theta/(1+\theta)]^{n+\alpha}}{[\theta/(1+\theta)]^{n+\alpha}} \cdot \frac{\Gamma(n+\alpha)}{\Gamma(n+\alpha)}\, \mathrm{d}\lambda \\
&= [\theta/(1+\theta)]^{n+\alpha} \cdot \Gamma(n+\alpha) \int_0^\infty \frac{\lambda^{n+\alpha} e^{-\lambda/[\theta/(1+\theta)]}}{\lambda \cdot [\theta/(1+\theta)]^{n+\alpha} \cdot \Gamma(n+\alpha)}
\end{aligned}$$

$$
\begin{aligned}
&= [\theta/(1+\theta)]^{n+\alpha} \cdot \Gamma(n+\alpha) \cdot 1 \\
&= \frac{\theta^{n+\alpha}}{(1+\theta)^{n+\alpha}} \cdot \Gamma(n+\alpha)
\end{aligned}
$$

Note that $\displaystyle\int_0^\infty \frac{\lambda^{n+\alpha} e^{-\lambda/[\theta/(1+\theta)]}}{\lambda \cdot [\theta/(1+\theta)]^{n+\alpha} \cdot \Gamma(n+\alpha)} \, \mathrm{d}\lambda = 1$ because the integrand is a gamma PDF with parameters $n+\alpha$ and $[\theta/(1+\theta)]$ integrated from 0 to infinity.

Therefore,

$$
\begin{aligned}
\Pr(N=n) &= \frac{1}{n!} \cdot \frac{1}{\theta^\alpha \cdot \Gamma(\alpha)} \cdot \frac{\theta^{n+\alpha}}{(1+\theta)^{n+\alpha}} \cdot \Gamma(n+\alpha) \\
&= \frac{1}{n!} \cdot \frac{1}{(\alpha-1)!} \cdot \frac{\theta^{n}}{(1+\theta)^{n+\alpha}} \cdot (n+\alpha-1)! \\
&= \frac{(n+\alpha-1)!}{n! \cdot (\alpha-1)!} \cdot \left(\frac{\theta}{1+\theta}\right)^n \left(\frac{1}{1+\theta}\right)^\alpha \\
&= \binom{n+\alpha-1}{n} \cdot \left(\frac{\theta}{1+\theta}\right)^n \left(\frac{1}{1+\theta}\right)^\alpha
\end{aligned}
$$

Notice that the expression above has the same form as a negative binomial PMF:

$$
\Pr(N=n) = \binom{n+r-1}{n} \left(\frac{\beta}{1+\beta}\right)^n \left(\frac{1}{1+\beta}\right)^r
$$

Therefore,

$$
N \sim \text{Negative Binomial}\,(r=\alpha,\, \beta=\theta)
$$

Negative Binomial PMF

Let N be a negative binomial random variable that counts the number of "failures" before the r^{th} "success"". With p being the probability of success of a trial, the probability of having n failures before the r^{th} success is

$$p_n = \binom{n+r-1}{r-1} p^r (1-p)^n, \qquad n = 0,\, 1,\, 2,\, \ldots$$

If we let $p = \dfrac{1}{1+\beta}$, then we can rewrite the PMF as

$$\begin{aligned} p_n &= \frac{(n+r-1)!}{(r-1)!\,n!} \left(\frac{1}{1+\beta}\right)^r \left(\frac{\beta}{1+\beta}\right)^n \\ &= \begin{cases} \dfrac{1}{(1+\beta)^r}, & n = 0 \\ \dfrac{r\,(r+1)\ldots(n+r-1)}{n!} \cdot \dfrac{\beta^n}{(1+\beta)^{n+r}}, & n = 1,\, 2,\, \ldots \end{cases} \end{aligned}$$

In this parameterization, $\beta = \dfrac{1-p}{p}$ represents the odds of failure.

Furthermore, the mean and variance are

$$\begin{aligned} \mathrm{E}\left[N\right] &= r\left(\frac{1}{p} - 1\right) \\ &= r\,(1+\beta-1) \\ &= r\beta \end{aligned}$$

$$\begin{aligned} \mathrm{Var}\left[N\right] &= r\left(\frac{1-p}{p^2}\right) \\ &= r\left(\frac{1-p}{p} \cdot \frac{1}{p}\right) \\ &= r\left[\beta \cdot (1+\beta)\right] \\ &= r\beta(1+\beta) \end{aligned}$$

1.2.0 Overview

5m

Most of the time, insurance policies do not cover full loss amounts from the ground up because complete coverage is often expensive and could therefore be difficult to sell. In addition, policyholders may take more risks if they know any losses they incur will be fully covered by insurance. There are a few common measures an insurer can implement to modify its coverage, limit its losses, and sell a more competitive product.

- *Policy limit* — to cap exposure for large claims.
- *Deductible* — to only cover claims over a certain amount.
- *Coinsurance* — to share a percentage of claims with the policyholders.

All 3 measures pass a portion of the *ground-up losses* to the policyholders. Ground-up loss refers to the entire loss amount incurred by the insured before any coverage modification is applied.

In short, claim payments often are not equal to ground-up losses. Thus, this section focuses on calculating the expected payment for a policy that implements the measures above.

Note that throughout this section, the term "payment" refers to the **insurer's** payment, unless otherwise specified. That is because an actuary, in most cases, represents the insurer in an insurance transaction.

Below is an overview of what will be covered in this section:

- **Policy limits**
- **Deductibles**
    - Ordinary deductibles
    - Loss elimination ratio
    - Payment per loss vs. payment per payment
    - Impact of deductibles on claim frequency
    - Franchise deductibles
- **Coinsurance**

1.2.1 Policy Limits

10m

The policy limit is the maximum amount the **insurer** will pay for a single loss.

Let X represent the loss variable. Note that X can never go below zero, as negative values do not make sense in the context of loss amounts. Then, for insurance with a policy limit u, the insurance payment is defined as

$$X \wedge u = \begin{cases} X, & X < u \\ u, & X \geq u \end{cases}$$

which implies the payment amount is the loss amount capped at u. In other words, the insurer pays whichever is lower: the loss or the policy limit. $(X \wedge u)$ is also called the *limited loss variable*.

The average payment amount in this case is called the *limited expected value*.

$$\mathrm{E}[X \wedge u]$$

For losses **below** the policy limit, i.e. $X < u$, the payment amount is the loss amount, so the average payment amount for this portion is:

$$\int_0^u x \cdot f(x)\,\mathrm{d}x$$

For losses **above** the policy limit, i.e. $X \geq u$, the payment amount is the policy limit, so the average payment amount for this portion is:

$$\int_u^\infty u \cdot f(x)\,\mathrm{d}x = u \cdot S(u)$$

Thus, the average payment amount is:

$$\mathrm{E}[X \wedge u] = \int_0^u x \cdot f(x)\,\mathrm{d}x + u \cdot S(u)$$

This can be extended to the k^{th} moment of the limited loss variable:

$$\mathrm{E}\Big[(X \wedge u)^k\Big] = \int_0^u x^k \cdot f(x)\,\mathrm{d}x + u^k \cdot S(u) \qquad (1.2.1.1)$$

Using Equation 1.0.4.2, i.e. the survival function method, the limited loss moments can also be expressed as

$$\mathrm{E}[X \wedge u] = \int_0^u S(x)\,\mathrm{d}x$$

$$\mathrm{E}\Big[(X \wedge u)^k\Big] = \int_0^u k\,x^{k-1} S(x)\,\mathrm{d}x \qquad (1.2.1.2)$$

Coach's Remarks

Claim size usually has a continuous distribution; thus, we will only introduce the formulas in continuous form. In the rare case where losses are discretely distributed, substitute sums for the integrals. For example:

$$\mathrm{E}\Big[(X \wedge u)^k\Big] = \left[\sum_{x \le u} x^k \cdot p(x)\right] + u^k \cdot S(u)$$

Note that while claim size is typically continuous, claim count often has a discrete distribution.

Example 1.2.1.1

Claim size for a medical insurance policy follows a Pareto distribution with parameters $\alpha = 5$ and $\theta = 1{,}000$.

The medical insurance has a policy limit of 500.

Determine the expected insurance payment for a claim.

Solution

Let X represent the claim size.

$$X \sim \text{Pareto}\ (5,\ 1{,}000)$$

The Pareto distribution's limited expected value formula is given in the exam table. Look up the formula and plug in the parameters and policy limit to calculate the answer.

$$\begin{aligned}\mathrm{E}[X \wedge 500] &= \frac{1{,}000}{5-1}\left[1 - \left(\frac{1{,}000}{500 + 1{,}000}\right)^{5-1}\right] \\ &\approx \mathbf{200.62}\end{aligned}$$

■

Example 1.2.1.2

For an auto insurance policy, claim amounts follow a distribution with the CDF:

$$F(x) = 1 - 0.6e^{-0.01x} - 0.4e^{-0.002x}$$

The auto insurance has a policy limit of 200.

Calculate the insurance company's expected payment for one claim.

Solution

$$\begin{aligned}
\mathrm{E}[X \wedge u] &= \int_0^u S(x)\,\mathrm{d}x \\
&= \int_0^{200} \left(0.6e^{-0.01x} + 0.4e^{-0.002x}\right) \mathrm{d}x \\
&= \left[-\frac{0.6}{0.01}e^{-0.01x} - \frac{0.4}{0.002}e^{-0.002x}\right]_0^{200} \\
&= \mathbf{117.82}
\end{aligned}$$

■

Alternative Solution

There is an alternative that does not require integration. Recognize that X is a mixture of two exponential distributions.

$$F(x) = 0.6\left(1 - e^{-0.01x}\right) + 0.4\left(1 - e^{-0.002x}\right)$$

$$w_1 = 0.6 \qquad \rightarrow \qquad X_1 \sim \text{Exponential } (100)$$

$$w_2 = 0.4 \qquad \rightarrow \qquad X_2 \sim \text{Exponential } (500)$$

Recall that for a mixture, raw moments can be calculated as the weighted average of the moments for each component. This concept extends to limited expected values as well. Note that the exponential distribution's limited expectation formula is provided in the exam table.

$$\begin{aligned} \mathrm{E}[X \wedge 200] &= w_1 \mathrm{E}[X_1 \wedge 200] + w_2 \mathrm{E}[X_2 \wedge 200] \\ &= 0.6\left[100\left(1 - e^{-200\,/\,100}\right)\right] + 0.4\left[500\left(1 - e^{-200\,/\,500}\right)\right] \\ &= \mathbf{117.82} \end{aligned}$$

■

1.2.2 Deductibles

45m

The deductible is the amount of each claim that the **policyholder** is responsible for paying before the insurer will pay anything on the claim. There are two types of deductibles: *ordinary deductibles* and *franchise deductibles*. If a question only says "deductible", it means **ordinary deductible**.

Ordinary Deductibles

Let X be the loss variable. For a policy with an ordinary deductible d, the **policyholder** (not **insurer**) pays the loss up to d.

$$X \wedge d = \begin{cases} X, & X < d \\ d, & X \geq d \end{cases}$$

Coach's Remarks

This is equivalent to what an **insurer** would pay for a policy with a policy limit d.

Thus, the **insurer** will be responsible for covering the remaining amount:

$$(X - d)_+ = \begin{cases} 0, & X \leq d \\ X - d, & X > d \end{cases}$$

Naturally, the contributions of the policyholder and insurer must add up to the full loss amount.

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

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

Coach's Remarks

$(X \wedge d)$ can be expressed as $\min(X,\, d)$. When X is smaller than d, the value is X. When X is greater than d, the value is capped at d.

$(X-d)_+$ can be expressed as $\max(X-d,\, 0)$. When $(X-d)$ is positive, i.e. greater than 0, the value is $(X-d)$. When $(X-d)$ is less than 0, the value is floored at 0.

Let's see why the relationships above hold by using the min and max expressions.

$$\begin{array}{ccccc}
\min(X,\, d) & + & \max(X,\, d) & = & X + d \\
\min(X,\, d) & + & \max(X,\, d) - d & = & X \\
\min(X,\, d) & + & \max(X - d,\, 0) & = & X \\
(X \wedge d) & + & (X - d)_+ & = & X
\end{array}$$

The first line is true because the minimum of any two quantities plus the maximum of the same two quantities will always equal the sum of the quantities.

Based on the above relationship, the following is also true.

$$\mathrm{E}[X \wedge d] + \mathrm{E}\left[(X-d)_+\right] = \mathrm{E}[X]$$

Thus, the expected insurance payment can be calculated as

$$\mathrm{E}\left[(X-d)_+\right] = \mathrm{E}[X] - \mathrm{E}[X \wedge d] \qquad (1.2.2.1)$$

This formula is **recommended** under most circumstances because the mean and limited expectation formulas for almost all distributions are provided in the exam table.

However, the formula does **not** apply to moments other than the first. Take the second moment for example, as seen below.

	X^2	$(X \wedge d)^2$	$X^2 - (X \wedge d)^2$	$(X-d)_+^2$
$X \le d$	X^2	X^2	0	0
$X > d$	X^2	d^2	$X^2 - d^2$	$X^2 - 2dX + d^2$

Notice that the values in column 4 are not equal to the values in column 5. Therefore,

$$\mathrm{E}\left[(X-d)_+^2\right] \neq \mathrm{E}\left[X^2\right] - \mathrm{E}\left[(X \wedge d)^2\right]$$

In cases where higher moments are needed, such as when calculating the variance, apply first principles instead, as shown below.

For losses **below** the deductible, i.e. $X \le d$, the insurer pays nothing, and the average payment amount for this portion is:

$$\int_0^d 0 \cdot f(x)\,\mathrm{d}x = 0$$

For the losses **above** the deductible, i.e. $X > d$, the insurance pays the loss amount minus the deductible, and the average payment amount for this portion is:

$$\int_d^\infty (x-d) \cdot f(x)\,\mathrm{d}x$$

Thus, the average payment amount is:

$$\mathrm{E}\left[(X-d)_+\right] = \int_d^\infty (x-d) \cdot f(x)\,\mathrm{d}x$$

This can be extended to the k^{th} moment of the payment variable:

$$\mathrm{E}\left[(X-d)_+^k\right] = \int_d^\infty (x-d)^k \cdot f(x)\,\mathrm{d}x \tag{1.2.2.2}$$

Likewise, using the survival function method,

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

$$\mathrm{E}\left[(X-d)_+^k\right] = \int_d^\infty k(x-d)^{k-1} S(x)\,\mathrm{d}x \tag{1.2.2.3}$$

Example 1.2.2.1

For an insurance policy:

- Losses follow an exponential distribution with mean 500.
- The policy has an ordinary deductible of 100 per loss.

Calculate the expected insurance payment per loss.

Solution

Define X to be the loss.

$$X \sim \text{Exponential}\,(500)$$

$$X \sim \text{Exponential}\ (500)$$

$$\mathrm{E}\left[(X-d)_+\right] = \mathrm{E}[X] - \mathrm{E}[X \wedge d]$$

The limited expected value of the exponential distribution is provided in the exam table.

$$\begin{aligned} \mathrm{E}[X \wedge 100] &= 500\left(1 - e^{-100\,/\,500}\right) \\ &= 90.63 \end{aligned}$$

Then, the expected insurance payment per loss is

$$\begin{aligned} \mathrm{E}\left[(X-100)_+\right] &= 500 - 90.63 \\ &= \mathbf{409.37} \end{aligned}$$

■

Coach's Remarks

Note that Equations 1.2.2.2 and 1.2.2.3 would produce the same result. Using Equation 1.2.2.3,

$$\begin{aligned} \mathrm{E}\left[(X-100)_+\right] &= \int_{100}^{\infty} S(x)\,\mathrm{d}x \\ &= \int_{100}^{\infty} e^{-x\,/\,500}\,\mathrm{d}x \\ &= \left[-500e^{-x\,/\,500}\right]_{100}^{\infty} \\ &= 500e^{-100\,/\,500} \end{aligned}$$

$$= \mathbf{409.37}$$

Loss Elimination Ratio

The *loss elimination ratio (LER)* measures how much the insurer saves by imposing an ordinary deductible.

So what does the insurer no longer pay, on average, after imposing an ordinary deductible? It is the portion of the loss that the policyholder has to pay, on average, i.e.

$$\mathrm{E}[X \wedge d]$$

Divide that by the average full loss amount, $\mathrm{E}[X]$, to compute the LER.

$$\mathrm{LER} = \frac{\mathrm{E}[X \wedge d]}{\mathrm{E}[X]} \tag{1.2.2.4}$$

Example 1.2.2.2

In the year 2017, claim amounts have the density function

$$f(x) = \frac{1}{10}, \qquad 0 < x < 10$$

In 2018, assume claims will follow the same distribution, and an ordinary deductible of 2 will be implemented.

Determine the loss elimination ratio in 2018.

Solution

$$\text{LER} = \frac{\text{E}[X \wedge d]}{\text{E}[X]}$$

Since the PDF is constant, we know X has a uniform distribution.

$$X \sim \text{Uniform }(0,\ 10)$$

Calculate the mean and limited expected value:

$$\begin{aligned}\text{E}[X] &= \frac{0+10}{2} \\ &= 5\end{aligned}$$

$$\begin{aligned}\text{E}[X \wedge 2] &= \int_0^2 \frac{x}{10}\,\text{d}x + 2 \cdot S(2) \\ &= \left[\frac{x^2}{20}\right]_0^2 + 2 \cdot \frac{10-2}{10} \\ &= 1.8\end{aligned}$$

The loss elimination ratio is

$$\begin{aligned}\text{LER} &= \frac{1.8}{5} \\ &= \mathbf{0.36}\end{aligned}$$

Payment per Loss vs. Payment per Payment

Up to this point, the expected payment we have been calculating is the *expected payment per loss*.

Most of the time, losses below the deductible are not reported since the policyholder would not be reimbursed anyway. Therefore, insurance companies often do not have sufficient information about losses below the deductible.

So, insurance companies are likely more interested in the *expected payment per payment*. To illustrate the difference between these two terms, consider the following example.

You are given the following losses:

$$2 \quad 3 \quad 7 \quad 9 \quad 14$$

An insurance policy has an ordinary deductible of 5.

Determine

1. the average payment **per loss**.
2. the average payment **per payment**.

The average payment per **loss** is simply the total payments divided by the number of **losses**.

$$\frac{\text{Total payment}}{\text{Number of losses}} = \frac{0+0+2+4+9}{5} = 3$$

Using the same logic, the average payment per **payment** is the total payments divided by the number of **payments**.

There are no payments for the first two losses since they are below the deductible. Thus, the number of payments is 3.

$$\frac{\text{Total payment}}{\text{Number of payments}} = \frac{2+4+9}{3} = 5$$

Coach's Remarks

Intuitively speaking, the former is the average payment **per loss incurred**, and the latter is the average payment **per payment made**. Since the number of payments will never exceed the number of losses, the average payment per payment will always be greater than or equal to the average payment per loss.

$$\mathrm{E}\left[Y^P\right] \geq \mathrm{E}\left[Y^L\right]$$

In general, we use Y^L to represent the payment per loss variable, and Y^P for the payment per payment variable. As a refresher, for a policy with an ordinary deductible, the expected payment per loss is

$$\mathrm{E}\left[Y^L\right] = \mathrm{E}\left[(X-d)_+\right]$$

As shown in the previous example, the expected payment per payment is the expected payment **given there is actually a payment**, which occurs when the loss is greater than the deductible. Therefore,

$$\begin{aligned}
\mathrm{E}\left[Y^P\right] &= \mathrm{E}\left[Y^L \mid Y^L > 0\right] \\
&= \mathrm{E}[X - d \mid X > d]
\end{aligned}$$

Recall from Section 1.0.6 that the conditional PDF is

$$f(x \mid X > d) = \frac{f(x)}{S(d)}, \qquad x > d$$

Using first principles, the expected payment per payment can be derived as follows:

$$\begin{aligned}
\mathrm{E}\left[Y^P\right] &= \mathrm{E}[X - d \mid X > d] \\
&= \int_d^\infty (x - d) \cdot f(x \mid X > d)\,\mathrm{d}x \\
&= \int_d^\infty (x - d) \cdot \frac{f(x)}{S(d)}\,\mathrm{d}x \\
&= \frac{1}{S(d)} \cdot \int_d^\infty (x - d) \cdot f(x)\,\mathrm{d}x \\
&= \frac{\mathrm{E}\left[(X - d)_+\right]}{S(d)} \\
&= \frac{\mathrm{E}\left[Y^L\right]}{S(d)}
\end{aligned}$$

Coach's Remarks

Many students mistake the following as the expected payment per payment.

$$\int_d^\infty (x-d)\cdot f(x)\,\mathrm{d}x = \mathrm{E}\left[Y^L\right] \neq \mathrm{E}\left[Y^P\right]$$

Remember that in order to calculate the expected payment per payment, we need to use the conditional PDF.

$$\int_d^\infty (x-d)\cdot \frac{f(x)}{S(d)}\,\mathrm{d}x = \mathrm{E}\left[Y^P\right]$$

In conclusion, an easier way to switch between the expected payment per payment and the expected payment per loss is to use the following equations.

$$\mathrm{E}\left[Y^P\right] = \frac{\mathrm{E}\left[Y^L\right]}{S(d)} \qquad \Leftrightarrow \qquad \mathrm{E}\left[Y^L\right] = \mathrm{E}\left[Y^P\right]\cdot S(d)$$

To remember this relationship, note that $\mathrm{E}\left[Y^P\right]$ will always be greater than or equal to $\mathrm{E}\left[Y^L\right]$. Because $S(d)\le 1$, in order to convert $\mathrm{E}\left[Y^L\right]$ to $\mathrm{E}\left[Y^P\right]$, we would need to divide by $S(d)$ to get a larger value.

This relationship is applicable to all higher-order moments.

$$\mathrm{E}\left[\left(Y^P\right)^k\right] = \frac{\mathrm{E}\left[\left(Y^L\right)^k\right]}{S(d)}$$

$$\Updownarrow$$

$$\mathrm{E}\left[\left(Y^L\right)^k\right] = \mathrm{E}\left[\left(Y^P\right)^k\right] \cdot S(d) \tag{1.2.2.5}$$

Coach's Remarks

Note that Y^L is the general notation for the payment per loss variable. However, Y^L is not always $(X-d)_+$. It is only the case if the policy has an ordinary deductible and no other coverage modifications.

For example, for a policy that has a policy limit u,

$$Y^L = X \wedge u$$

And for a policy that has a deductible d and a policy limit u,

$$Y^L = (X \wedge m) - (X \wedge d)$$

where m, which will be introduced in Section 1.2.3, is the maximum covered loss.

The same can be said about Y^P. Its expression depends on the coverage modifications.

Also, note that when there is no deductible, $Y^P = Y^L$. This is because all losses will result in payments.

Example 1.2.2.3

For an insurance policy:

- Losses follow an exponential distribution with mean 500.
- The policy has an ordinary deductible of 100 per loss.
- Y^P is the payment per payment random variable.

Calculate $\mathrm{Var}\left[Y^P\right]$.

Solution

Let X be the loss. Define the variables:

$$X \sim \text{Exponential}\ (500)$$

$$Y^L = (X - 100)_+$$

$$Y^P = X - 100 \mid X > 100$$

Our goal is to calculate

$$\mathrm{Var}\left[Y^P\right] = \mathrm{E}\left[\left(Y^P\right)^2\right] - \mathrm{E}\left[Y^P\right]^2$$

Use Equation 1.2.2.5 to calculate the payment per payment moments.

$$\begin{aligned}\mathrm{E}\left[Y^P\right] &= \frac{\mathrm{E}\left[Y^L\right]}{S(d)} \\ &= \frac{409.37}{0.8187} \\ &= 500\end{aligned}$$

Recall that the expected payment per loss was calculated in Example 1.2.2.1 as 409.37, and that $S(100) = e^{-100\,/\,500} = 0.8187$.

Next,

$$\begin{aligned}
\mathrm{E}\left[\left(Y^L\right)^2\right] &= \int_d^\infty 2(x-d)S(x)\,\mathrm{d}x \\
&= \int_{100}^\infty (2x-200)e^{-x\,/\,500}\,\mathrm{d}x \\
&= 2\left[x\left(-500e^{-x\,/\,500}\right) - 500^2 e^{-x\,/\,500}\right]_{100}^{\infty} - 200\left[-500e^{-x\,/\,500}\right]_{100}^{\infty} \\
&= 409{,}365.3765
\end{aligned}$$

$$\begin{aligned}
\mathrm{E}\left[\left(Y^P\right)^2\right] &= \frac{\mathrm{E}\left[\left(Y^L\right)^2\right]}{S(d)} \\
&= 500{,}000
\end{aligned}$$

Thus, the variance is

$$\begin{aligned}
\mathrm{Var}\left[Y^P\right] &= 500{,}000 - 500^2 \\
&= \mathbf{250{,}000}
\end{aligned}$$

■

Alternative Solution

Alternative Solution

A simpler way of solving this problem is by recalling the memoryless property of the exponential distribution.

For $X \sim \text{Exponential}\ (\theta)$ and $d > 0$,

$$X - d \mid X > d \sim \text{Exponential}\ (\theta)$$

Thus, in this case,

$$Y^P \sim \text{Exponential}\ (500)$$

$$\begin{aligned} \text{Var}\left[Y^P\right] &= 500^2 \\ &= \mathbf{250{,}000} \end{aligned}$$

$\blacksquare$

$(X \wedge d)$ is called the limited loss variable, while $(X - d \mid X > d)$ is called the *excess loss variable*. Its mean, denoted by $e(d)$, is the *mean excess loss function*.

$$e(d) = \text{E}[X - d \mid X > d]$$

Sometimes, we will use $e_X(d)$ to indicate the loss variable in the subscript.

The table below includes shortcuts for calculating the mean excess loss for certain loss distributions.

Loss	Mean	Excess Loss	Mean Excess Loss
Exponential (θ)	θ	Exponential (θ)	θ
Uniform (a, b)	$\dfrac{a+b}{2}$	Uniform $(0, b-d)$	$\dfrac{b-d}{2}$
Pareto (α, θ)	$\dfrac{\theta}{\alpha-1}$	Pareto $(\alpha, \theta+d)$	$\dfrac{\theta+d}{\alpha-1}$
S-P Pareto (α, θ)	$\dfrac{\alpha\theta}{\alpha-1}$	Pareto (α, d)	$\dfrac{d}{\alpha-1}$
Beta $(1, b, \theta)$	$\dfrac{\theta}{1+b}$	Beta $(1, b, \theta-d)$	$\dfrac{\theta-d}{1+b}$

The proofs are included in the appendix at the end of this section.

Franchise Deductibles

The franchise deductible is slightly different from the ordinary deductible. When a loss is greater than the deductible, a policy with a franchise deductible will pay the **full** amount of the loss.

$$Y^L = \begin{cases} 0, & X \le d \\ X, & X > d \end{cases}$$

In contrast, a policy with an ordinary deductible will only pay the amount in excess of the deductible.

The expected payment of a policy with a franchise deductible can be calculated by taking the expected payment of a policy with an ordinary deductible and adding back the deductible when a payment is made.

$$\begin{aligned} \mathrm{E}\left[Y^L\right] &= \int_d^\infty x \cdot f(x)\,\mathrm{d}x \\ &= \int_d^\infty x \cdot f(x)\,\mathrm{d}x - \int_d^\infty d \cdot f(x)\,\mathrm{d}x + \int_d^\infty d \cdot f(x)\,\mathrm{d}x \\ &= \int_d^\infty (x-d) \cdot f(x)\,\mathrm{d}x + d \cdot \int_d^\infty f(x)\,\mathrm{d}x \end{aligned}$$

$$= \mathrm{E}\left[(X-d)_+\right] + d \cdot S(d)$$

The expected payment per payment is obtained by dividing the expected payment per loss by the survival function evaluated at the deductible.

$$\begin{aligned}\mathrm{E}\left[Y^P\right] &= \frac{\mathrm{E}\left[(X-d)_+\right] + d \cdot S(d)}{S(d)} \\ &= e(d) + d\end{aligned}$$

Example 1.2.2.4

Loss amounts have a distribution with survival function

$$S(x) = \sqrt{\left(\frac{100}{x+100}\right)^5}, \qquad x > 0$$

An insurance coverage for these losses has a franchise deductible of 50.

Calculate the expected payment per payment.

Solution

$$S(x) = \sqrt{\left(\frac{100}{x+100}\right)^5} = \left(\frac{100}{x+100}\right)^{2.5} = \left(\frac{\theta}{x+\theta}\right)^{\alpha}$$
$$\Downarrow$$
$$X \sim \text{Pareto}\ (2.5,\ 100)$$

For a policy with a franchise deductible, the expected payment per payment is

$$\mathrm{E}\left[Y^P\right] = e(d) + d$$

Recall that for $X \sim \text{Pareto}\ (\alpha,\, \theta)$ and $d > 0$,

$$X - d \mid X > d \sim \text{Pareto}\ (\alpha,\, \theta + d)$$

Thus,

$$\begin{aligned} e(d) &= \mathrm{E}[X - d \mid X > d] \\ &= \frac{\theta + d}{\alpha - 1} \end{aligned}$$

Calculate the expected payment per payment:

$$\begin{aligned} \mathrm{E}\left[Y^P\right] &= \frac{100 + 50}{2.5 - 1} + 50 \\ &= \mathbf{150} \end{aligned}$$

■

Impact of Deductibles on Claim Frequency

Recall that the number of payments is not necessarily equal to the number of losses.

When there is a deductible of d, a payment will only be made when the loss is greater than the deductible.

than the deductible.

If the number of losses N is modeled by an $(a,\ b,\ 0)$ class distribution, the number of payments N' will follow the same distribution but with modified parameters.

The parameters are modified as follows:

	$\boldsymbol{N}$	$\boldsymbol{N'}$
Poisson	λ	$v\lambda$
Binomial	$m,\ q$	$m,\ vq$
Negative binomial	$r,\ \beta$	$r,\ v\beta$

where v is the probability of a loss being greater than the deductible, $\Pr(X > d)$. In other words, v is the probability that a loss results in a positive payment.

The proofs of these modifications are included in the appendix at the end of this section.

Example 1.2.2.5

A policyholder's number of losses per year follows a binomial distribution with parameters $m = 2$ and $q = 0.3$.

Each loss is subject to a deductible d where the probability of a loss exceeding the deductible is 0.8.

The number of losses and the loss amounts are mutually independent.

Determine the probability of a policyholder receiving no payment from the insurance company in a year.

Solution

Modify the frequency from number of losses to number of payments.

$$N \sim \text{Binomial}\ (2,\ 0.3)$$

Let N' be the number of payments. The modified parameter q is

$$0.3 \cdot \Pr(X > d) = 0.3(0.8) = 0.24$$

Thus,

$$N' \sim \text{Binomial}\ (2,\ 0.24)$$

The probability of no payments is

$$p_{N'}(0) = \binom{2}{0} 0.24^0 (1 - 0.24)^2 = \mathbf{0.5776}$$

■

1.2.3 Coinsurance

30m

Coinsurance is the portion of loss the **insurer** is responsible for. Suppose a policy has coinsurance of α, where $0 < \alpha < 1$. For a loss of amount X, the insurer will pay αX while the policyholder will pay the rest, $(1-\alpha)X$.

Based on the expected value property covered in Section 1.0.4, the expected payment is simply the expected loss multiplied by the coinsurance.

$$\mathrm{E}[\alpha X] = \alpha \cdot \mathrm{E}[X]$$

In the case where a policy has both a **deductible** and **coinsurance**, the coinsurance typically only comes into play **after** the deductible is met. The payment per loss variable is as shown below. For a policy with deductible d and coinsurance α,

$$Y^L = \begin{cases} 0, & X \le d \\ \alpha\,(X-d), & X > d \end{cases}$$

The policyholder pays the first d of the loss and then shares the excess amount with the insurer. Then, the average payment is

$$\mathrm{E}\left[Y^L\right] = \alpha\,(\mathrm{E}[X] - \mathrm{E}[X \wedge d])$$

Coach's Remarks

If a question specifies that the coinsurance is applied **before** the deductible, then instead of applying the entire loss X against the deductible, the insurer will only apply the coinsured portion of the loss, αX. The policyholder will still be responsible for the remaining portion, i.e. $(1-\alpha)X$, plus the amount needed to meet the deductible.

So, the payment per loss variable is as shown below. For a policy with deductible d

and coinsurance α,

$$\begin{aligned} Y^L &= \begin{cases} 0, & \alpha X \le d \\ \alpha X - d, & \alpha X > d \end{cases} \\ &= \begin{cases} 0, & X \le \dfrac{d}{\alpha} \\ \alpha\left(X - \dfrac{d}{\alpha}\right), & X > \dfrac{d}{\alpha} \end{cases} \end{aligned}$$

In conclusion, if the question specifies that the coinsurance is applied **before** the deductible, adjust the deductible by dividing it by the coinsurance. Then, solve it the same way as the case where the coinsurance is applied **after**. Thus, the average payment is

$$\mathrm{E}\left[Y^L\right] = \alpha\left(\mathrm{E}[X] - \mathrm{E}\left[X \wedge \frac{d}{\alpha}\right]\right)$$

Now that we have discussed each coverage modification individually, let's look at how they can be combined. There will be multiple examples at the end to demonstrate the most common cases.

Here is the general formula for the expected payment per loss when all of the coverage modifications we have discussed are included. We will modify and apply it to each of the examples below.

$$\mathrm{E}\left[Y^L\right] = \alpha \cdot \{\mathrm{E}[X \wedge m] - \mathrm{E}[X \wedge d]\} \tag{1.2.3.1}$$

where

- X is the loss variable,
- u is the policy limit (set $u = \infty$ if policy limit doesn't apply),

- d is the deductible (set $d=0$ if deductible doesn't apply),
- α is the coinsurance (set $\alpha=1$ if coinsurance doesn't apply), and
- m is the *maximum covered loss* and is equal to $\dfrac{u}{\alpha}+d$.

Notice the first limited expected value is evaluated at the **maximum covered loss**, rather than the policy limit. Many students make a mistake here.

The maximum covered loss is the loss amount above which the insurer pays the policy limit, i.e. when $X \ge m$, $Y^L = u$. As a result, there is a maximum covered loss only when there is a policy limit.

Altogether, for a policy with deductible d, coinsurance α, and policy limit u,

$$Y^L = \begin{cases} 0, & X \le d \\ \alpha\,(X-d), & d < X < m \\ u, & X \ge m \end{cases}$$

- If a policy only has a policy limit, then $m=u$.
- If a policy has a policy limit and a deductible, then $m=u+d$.
- If a policy has a policy limit, a deductible, and coinsurance, then $m=\dfrac{u}{\alpha}+d$.

Coach's Remarks

$(X \wedge \infty)$ means that there is no cap on the loss X. Thus, the limited expectation is simply the expected loss.

$$\mathrm{E}[X \wedge \infty] = \int_0^\infty S(x)\,\mathrm{d}x = \mathrm{E}[X]$$

$(X \wedge 0)$ means the loss X is capped at 0. Thus, the limited expectation is 0.

$$\mathrm{E}[X \wedge 0] = \int_0^0 S(x)\,\mathrm{d}x = 0$$

Therefore, the expected payment for an insurance policy with deductible 0, $\mathrm{E}\left[(X-0)_+\right]$, equals the expected loss.

$$\begin{aligned}\mathrm{E}\left[(X-0)_+\right] &= \mathrm{E}[X] - \mathrm{E}[X \wedge 0] \\ &= \mathrm{E}[X]\end{aligned}$$

Example 1.2.3.1

For an insurance policy, you are given:

- The loss severity follows a Pareto distribution with parameters $\alpha = 2$ and $\theta = 1{,}000$.
- The policy has an ordinary deductible of 200 and a policy limit of 2,000.

Calculate the expected insurance payment per loss.

Solution

Start by modifying Equation 1.2.3.1. Substitute $d = 200$, $u = 2{,}000$, and $\alpha = 1$ because there is no coinsurance. The modified formula is

$$\mathrm{E}\left[Y^L\right] = \mathrm{E}[X \wedge 2{,}200] - \mathrm{E}[X \wedge 200]$$

Look up the Pareto distribution's limited expectation formula in the exam table.

$$\begin{aligned}\mathrm{E}[X \wedge 2{,}200] &= \frac{1{,}000}{2-1}\left[1-\left(\frac{1{,}000}{2{,}200+1{,}000}\right)^{2-1}\right] \\ &= 687.50\end{aligned}$$

$$\begin{aligned}\mathrm{E}[X \wedge 200] &= \frac{1{,}000}{2-1}\left[1-\left(\frac{1{,}000}{200+1{,}000}\right)^{2-1}\right] \\ &= 166.67\end{aligned}$$

Calculate the expected payment per loss.

$$\begin{aligned}\mathrm{E}\left[Y^L\right] &= 687.50 - 166.67 \\ &= \mathbf{520.83}\end{aligned}$$

■

Coach's Remarks

The table below illustrates why the expected payment per loss expression above is correct.

	$X \wedge 2{,}200$	$X \wedge 200$	Y^L
$X \leq 200$	X	X	0
$200 < X < 2{,}200$	X	200	$X - 200$
$X \geq 2{,}200$	$2{,}200$	200	$2{,}000$

Subtracting column 3 from column 2 will produce the same outcome as the payment per loss variable, i.e.

$$Y^L = (X \wedge 2{,}200) - (X \wedge 200)$$

Therefore, the expected payment per loss is

$$\mathrm{E}\left[Y^L\right] = \mathrm{E}[X \wedge 2{,}200] - \mathrm{E}[X \wedge 200]$$

Example 1.2.3.2

Losses follow an exponential distribution with mean 1,000.

For an insurance policy:

- Each loss is subject to an ordinary deductible of 200.
- The policy will reimburse 80% of the losses in excess of 200.
- The insurance payment is capped at 2,000.

Calculate the expected payment per loss.

Solution

Modify Equation 1.2.3.1 by setting $u = 2{,}000$, $d = 200$, and $\alpha = 0.8$. The resulting formula is

$$\mathrm{E}\left[Y^L\right] = 0.8 \cdot (\mathrm{E}[X \wedge 2{,}700] - \mathrm{E}[X \wedge 200])$$

We need to compute the limited expected values.

$$\begin{aligned}\mathrm{E}[X \wedge 2{,}700] &= 1{,}000\left(1 - e^{-2{,}700\,/\,1{,}000}\right)\\ &= 932.79\end{aligned}$$

$$\begin{aligned}\mathrm{E}[X \wedge 200] &= 1{,}000\left(1 - e^{-200\,/\,1{,}000}\right)\\ &= 181.27\end{aligned}$$

Substitute the limited expected values into the formula to calculate the final answer.

$$\begin{aligned}\mathrm{E}\left[Y^L\right] &= 0.8 \cdot (932.79 - 181.27)\\ &= \mathbf{601.22}\end{aligned}$$

■

Coach's Remarks

Let's confirm that the expected payment per loss expression is correct.

- For losses below 200, $Y^L = 0$.
- For losses between 200 and 2,700, Y^L is 80% of the amount in excess of 200, i.e. $Y^L = 0.8(X - 200)$.
- For losses greater than 2,700, Y^L is capped at the policy limit, i.e.

2,000.

	$0.8\,(X \wedge 2{,}700)$	$0.8\,(X \wedge 200)$	Y^L
$X \leq 200$	$0.8X$	$0.8X$	0
$200 < X < 2{,}700$	$0.8X$	$0.8(200)$	$0.8(X - 200)$
$X \geq 2{,}700$	$0.8(2{,}700)$	$0.8(200)$	$2{,}000$

Subtracting column 3 from column 2 produces the last column, as intended.

Example 1.2.3.3

Losses follow a gamma distribution with parameters $\alpha = 2$ and $\theta = 100$.

An insurance policy covers the loss with a policy limit of 250. The insurance company is offering an alternative coverage to replace the policy limit with coinsurance c, which is the proportion of the loss paid by the insurance company, such that the expected insurance cost remains the same.

Determine c.

Solution

Let X represent the losses.

$$X \sim \text{Gamma}\ (2,\ 100)$$

For the coverage with a policy limit of 250, modify Equation 1.2.3.1 by setting $u = 250, d = 0$, and $\alpha = 1$:

$$\mathrm{E}[X \wedge 250]$$

For the alternative coverage with coinsurance c, modify Equation 1.2.3.1 by setting $u = \infty, d = 0$, and $\alpha = c$:

$$c \cdot \mathrm{E}[X]$$

Equate the two expected payments. The goal is to solve for c.

$$c \cdot \mathrm{E}[X] = \mathrm{E}[X \wedge 250]$$

First, calculate the mean.

$$\begin{aligned} \mathrm{E}[X] &= 2(100) \\ &= 200 \end{aligned}$$

Next, calculate the limited expected value. The formula is provided in the exam table. Substitute $k = 1$ into the formula to get:

$$\mathrm{E}[X \wedge u] = \alpha\theta\Gamma(\alpha + 1;\, u/\theta) + u\left[1 - \Gamma(\alpha;\, u/\theta)\right]$$

Substitute the parameters and policy limit into the formula. We will need to evaluate incomplete gamma functions.

$$\mathrm{E}[X \wedge 250] = 2(100)\Gamma(3;\, 2.5) + 250\left[1 - \Gamma(2;\, 2.5)\right]$$

Recall the shortcut introduced in Section 1.1.5. Start by defining a Poisson variable.

$$N \sim \text{Poisson}\left(\lambda = \frac{250}{100} = 2.5\right)$$

Evaluate the incomplete gamma functions.

$$\begin{aligned}\Gamma(3;\ 2.5) &= 1 - \Pr(N < 3) \\ &= 1 - [p_N(0) + p_N(1) + p_N(2)] \\ &= 1 - \left[e^{-2.5} + e^{-2.5}(2.5) + \frac{e^{-2.5}(2.5)^2}{2}\right] \\ &= 0.4562\end{aligned}$$

$$\begin{aligned}\Gamma(2;\ 2.5) &= 1 - \Pr(N < 2) \\ &= 1 - [p_N(0) + p_N(1)] \\ &= 1 - \left[e^{-2.5} + e^{-2.5}(2.5)\right] \\ &= 0.7127\end{aligned}$$

Substitute the calculated values into the limited expected value formula.

$$\begin{aligned}\mathrm{E}[X \wedge 250] &= 2(100)(0.4562) + 250(1 - 0.7127) \\ &\approx 163\end{aligned}$$

Lastly, solve for c.

$$\begin{aligned}c(200) &\approx 163 \\ c &\approx \mathbf{81.5\%}\end{aligned}$$

■

Coach's Remarks

The incomplete gamma functions can also be evaluated by integrating the gamma PDF. Integration by parts is required. Knowing the Poisson shortcut is an advantage but not a must.

Example 1.2.3.4

An insurance company will provide a financial incentive to its insurance agents if the total incurred losses for the year are less than 500.

The bonus will be 15% of the amount by which the total incurred losses are under 500.

The total losses in a year follow a Pareto distribution with parameters $\alpha = 3$ and $\theta = 400$.

Determine the expected bonus.

Solution

Start by defining the bonus random variable.

Let X be the incurred loss amount.

If the incurred losses are less than 500, the bonus is 15% of the difference between the incurred losses and 500.

If the incurred losses are greater than 500, no bonus is paid.

If the incurred losses are greater than 500, no bonus is paid.

Thus, the bonus variable is

$$B = \begin{cases} 0.15(500 - X), & X < 500 \\ 0, & X \geq 500 \end{cases} = 0.15(500 - X)_+$$

The expected bonus is

$$\begin{aligned} \mathrm{E}[B] &= 0.15\,\mathrm{E}\left[(500 - X)_+\right] \\ &= 0.15\,(\mathrm{E}[500] - \mathrm{E}[500 \wedge X]) \end{aligned}$$

$(500 \wedge X)$ can be expressed as $\min(500,\ X)$. The order of the items in the parentheses does not matter. Thus, it can also be expressed as $\min(X,\ 500)$, or $(X \wedge 500)$, which means

$$\begin{aligned} \mathrm{E}[B] &= 0.15\,(\mathrm{E}[500] - \mathrm{E}[X \wedge 500]) \\ &= 0.15\left\{500 - \frac{400}{3-1}\left[1 - \left(\frac{400}{500+400}\right)^{3-1}\right]\right\} \\ &= \mathbf{50.93} \end{aligned}$$

$\blacksquare$

1.2.4 Inflation

25m

Inflation increases costs, but policy modifications generally aren't changed in a given policy period to adjust for inflation. So, actuaries are interested in modeling the effects of inflation. Inflation is an application of scaling (multiplying a random variable by a constant), and exam questions may pair inflation with coverage modifications.

Note that the following identity holds for constants d and r.

$$\mathrm{E}[(1+r)X \wedge d] = (1+r)\mathrm{E}\left[X \wedge \frac{d}{1+r}\right]$$

Let's practice with a few examples.

Example 1.2.4.1

This year, losses follow a single-parameter Pareto distribution with the PDF:

$$f(x) = \frac{20{,}000}{x^3}, \qquad x > 100$$

Individual losses increase 10% each year due to inflation.

An insurance policy has a policy limit of 220.

Determine the expected insurance payment next year.

Solution

Let X represent this year's losses.

$$X \sim \text{S-P Pareto}\ (2,\ 100)$$

$1.1X$ represents **next** year's losses. The goal is to calculate

$$\mathrm{E}[1.1X \wedge 220] = 1.1\mathrm{E}\left[X \wedge \frac{220}{1.1}\right]$$

Look up the limited expected value formula from the exam table.

$$\begin{aligned} 1.1\mathrm{E}\left[X \wedge \frac{220}{1.1}\right] &= 1.1\mathrm{E}[X \wedge 200] \\ &= 1.1\left[\frac{2 \cdot 100}{2-1} - \frac{(1)100^2}{(2-1)200^{2-1}}\right] \\ &= \mathbf{165} \end{aligned}$$

■

Alternative Solution

We can apply the concept of scaling discussed in Section 1.1.8. A scaled random variable follows the same distribution with the θ parameter scaled. Thus, next year's losses follow a single-parameter Pareto distribution, where α remains as 2 and $\theta = 1.1(100) = 110$.

$$1.1X \sim \text{S-P Pareto}\ (2, 110)$$

The limited expected value is

$$\begin{aligned}\mathrm{E}[1.1X \wedge 220] &= \frac{2 \cdot 110}{2-1} - \frac{(1)110^2}{(2-1)220^{2-1}} \\ &= \mathbf{165}\end{aligned}$$

■

Coach's Remarks

This question can also be solved using first principles.

$$\begin{aligned}\mathrm{E}[1.1X \wedge 220] &= \int_0^\infty \min(1.1x,\ 220)\, f(x)\,\mathrm{d}x \\ &= \int_0^\infty 1.1 \min\left(x,\ \frac{220}{1.1}\right) f(x)\,\mathrm{d}x \\ &= 1.1 \int_0^\infty \min(x,\ 200)\, f(x)\,\mathrm{d}x \\ &= 1.1\mathrm{E}[X \wedge 200]\end{aligned}$$

Example 1.2.4.2

In 2016, losses were uniformly distributed on the interval $[0,\ 1{,}000]$.

Losses in 2017 are 20% higher than in 2016.

An insurance policy covers each loss subject to an ordinary deductible of 300.

Calculate the difference between the loss elimination ratios in 2016 and 2017, $\text{LER}_{2016} - \text{LER}_{2017}$.

Solution

Recall the loss elimination ratio formula for an ordinary deductible.

$$\text{LER} = \frac{\text{E}[X \wedge d]}{\text{E}[X]}$$

Let X represent the loss in 2016.

$$X \sim \text{Uniform}\ (0,\ 1{,}000)$$

Calculate the mean and limited expected value of losses in 2016.

$$\begin{aligned}\text{E}[X] &= \frac{0 + 1{,}000}{2} \\ &= 500\end{aligned}$$

$$\begin{aligned}\text{E}[X \wedge 300] &= \int_0^{300} S(x)\,\text{d}x \\ &= \int_0^{300} \left(1 - \frac{x}{1{,}000}\right) \text{d}x \\ &= \left[\quad x^2 \quad \right]^{300}\end{aligned}$$

$$= \left[x - \frac{x^2}{2{,}000} \right]_0^{250}$$
$$= 255$$

Thus, the loss elimination ratio in 2016 is

$$\begin{aligned}\text{LER}_{2016} &= \frac{255}{500} \\ &= 0.51\end{aligned}$$

Losses in 2017 are 20% higher. Thus, losses in 2017 are represented by $1.2X$.

$$\begin{aligned}\mathrm{E}[1.2X] &= 1.2\,\mathrm{E}[X] \\ &= 1.2\,(500) \\ &= 600\end{aligned}$$

$$\begin{aligned}\mathrm{E}[1.2X \wedge 300] &= 1.2\,\mathrm{E}[X \wedge 250] \\ &= 1.2 \int_0^{250} S(x)\,\mathrm{d}x \\ &= 1.2 \int_0^{250} \left(1 - \frac{x}{1{,}000}\right) \mathrm{d}x \\ &= 1.2 \left[x - \frac{x^2}{2{,}000} \right]_0^{250} \\ &= 262.5\end{aligned}$$

Thus, the loss elimination ratio in 2017 is

$$\begin{aligned}\text{LER}_{2017} &= \frac{262.5}{600} \\ &= 0.4375\end{aligned}$$

Calculate the final answer.

$$\begin{aligned}\text{LER}_{2016} - \text{LER}_{2017} &= 0.51 - 0.4375 \\ &= \mathbf{0.0725}\end{aligned}$$

■

Coach's Remarks

Alternatively, use the fact that

$$1.2X \sim \text{Uniform}\,(0,\ 1{,}200)$$

to calculate the mean and limited expected value of losses in 2017.

We can add the effects of inflation to Equation 1.2.3.1 as follows:

$$\text{E}\left[Y^L\right] = \alpha\,(1+r)\left(\text{E}\left[X \wedge \frac{m}{1+r}\right] - \text{E}\left[X \wedge \frac{d}{1+r}\right]\right) \qquad (1.2.4.1)$$

where

- X is the loss variable,
- u is the policy limit (set $u = \infty$ if policy limit doesn't apply),

- d is the deductible (set $d = 0$ if deductible doesn't apply),
- α is the coinsurance (set $\alpha = 1$ if coinsurance doesn't apply),
- r is the inflation rate (set $r = 0$ if inflation doesn't apply), and
- m is the maximum covered loss and equals $\dfrac{u}{\alpha} + d$.

Likewise, Equation 1.2.2.5 becomes

$$\mathrm{E}\left[\left(Y^P\right)^k\right] = \frac{\mathrm{E}\left[\left(Y^L\right)^k\right]}{S_X\left(\dfrac{d}{1+r}\right)}$$

$$\Updownarrow$$

$$\mathrm{E}\left[\left(Y^L\right)^k\right] = \mathrm{E}\left[\left(Y^P\right)^k\right] \cdot S_X\left(\frac{d}{1+r}\right) \tag{1.2.4.2}$$

Alternatively, you can choose to stick with Equation 1.2.3.1 and replace X in the equation with $X^* = (1+r)X$. For example,

$$X \sim \text{Pareto}\,(\alpha,\, \theta) \quad \Rightarrow \quad X^* \sim \text{Pareto}\,(\alpha,\, \theta^*)$$

where $X^* = (1+r)X$ and $\theta^* = (1+r)\theta$.

The expected payment per loss would then be

$$\mathrm{E}\left[Y^L\right] = \alpha\left(\mathrm{E}[X^* \wedge m] - \mathrm{E}[X^* \wedge d]\right)$$

The alternative method is most useful when the distribution of X^* is easy to derive, such as when the distribution of X has a scale parameter.

Coach's Remarks

Equation 1.2.4.1 is applicable to almost all situations we have discussed, except the following two cases that are not commonly seen.

For a **franchise deductible**, since the full amount is paid when the loss is greater than the deductible, two changes need to be made to the ultimate formula. First, the maximum covered loss will be d less than it would be for an ordinary deductible.

$$m = \frac{u}{\alpha}$$

Second, add back the effect of the deductible for the portion of loss above the deductible, i.e.

$$+\,\alpha \cdot d \cdot S_X\left(\frac{d}{1+r}\right)$$

For **coinsurance that is applied before the deductible**, replace d with the deductible divided by the coinsurance factor.

$$d = \frac{\text{Deductible}}{\alpha}$$

Example 1.2.4.3

This year's losses follow a distribution that has the following CDF values and limited expected values:

x	$F(x)$	$\mathrm{E}[X \wedge x]$
500	0.44	375
600	0.49	429
1,800	0.79	818
2,415	0.85	925
3,000	0.89	1,000
4,350	0.93	1,115
∞	1.00	1,500

A policy covering the losses has an ordinary deductible of 600.

Assume losses are 20% higher next year.

The insurer is considering adding one of the following to the policy starting next year:

- A policy limit of 3,000.
- A coinsurance of α.

Determine α such that the expected payment per loss will be equal regardless of the insurer's decision.

Solution

Calculate the expected payment per loss under each option by modifying Equation 1.2.4.1.

If the insurer chooses to add the policy limit, the expected payment will be

$$u = 3{,}000 \quad d = 600 \quad \alpha = 1 \quad r = 0.2 \quad m = 3{,}600$$

$$\mathrm{E}\left[Y^L\right] = 1.2\left(\mathrm{E}\left[X \wedge \frac{3{,}600}{1.2}\right] - \mathrm{E}\left[X \wedge \frac{600}{1.2}\right]\right)$$

$$
\begin{aligned}
&= 1.2\left(\mathrm{E}[X \wedge 3{,}000] - \mathrm{E}[X \wedge 500]\right) \\
&= 1.2\,(1{,}000 - 375) \\
&= 750
\end{aligned}
$$

If the insurer chooses to add the coinsurance, the expected payment will be

$$u = \infty \quad d = 600 \quad \alpha = \alpha \quad r = 0.2 \quad m = \infty$$

$$
\begin{aligned}
\mathrm{E}\left[Y^L\right] &= \alpha\,(1.2)\left(\mathrm{E}\left[X \wedge \frac{\infty}{1.2}\right] - \mathrm{E}\left[X \wedge \frac{600}{1.2}\right]\right) \\
&= \alpha\,(1.2)\left(\mathrm{E}[X] - \mathrm{E}[X \wedge 500]\right) \\
&= \alpha\,(1.2)\,(1{,}500 - 375) \\
&= 1{,}350\alpha
\end{aligned}
$$

Equate the two expected payments to solve for α:

$$
\begin{aligned}
750 &= 1{,}350\alpha \\
\alpha &= \mathbf{0.5556}
\end{aligned}
$$

$\blacksquare$

If a question asks for the variance of the payment, it will likely

- involve a distribution that has a special property, such as memoryless, so that the variance can be calculated easily (as illustrated in Example 1.2.2.3), or
- have a simple payment structure so that you can solve it using first principles (see

Example 1.2.4.4 below).

Example 1.2.4.4

Losses follow a probability distribution that has the following survival function:

$$S(x) = 1 - \frac{x}{200}, \qquad 0 \le x \le 200$$

An insurance policy has a deductible of 20 and a maximum covered loss of 100.

Calculate the variance of payment per payment, $\mathrm{Var}\left[Y^P\right]$.

Solution

Based on the survival function, losses follow a uniform distribution on the interval 0 to 200.

Also, a maximum covered loss of 100 indicates a policy limit of 80.

$$u = m - d = 100 - 20 = 80$$

The variance is the second moment minus the first moment squared.

$$\mathrm{Var}\left[Y^P\right] = \mathrm{E}\left[\left(Y^P\right)^2\right] - \mathrm{E}\left[Y^P\right]^2$$

A simpler way to calculate Y^P's moments is by calculating Y^L's moments, and then converting them.

Start by determining the payment per loss variable. The insurer pays

- nothing for losses below the deductible;
- the amount in excess of the deductible for losses between the deductible and the maximum covered loss;
- the policy limit for losses greater than the maximum covered loss.

The payment per loss variable is

$$Y^L = \begin{cases} 0, & X \le 20 \\ X - 20, & 20 < X < 100 \\ 80, & X \ge 100 \end{cases}$$

Calculate the first and second moments using first principles.

$$\begin{aligned} \mathrm{E}\left[Y^L\right] &= \int_{20}^{100} (x-20) \cdot \frac{1}{200}\,\mathrm{d}x + 80\,S(100) \\ &= \left[\frac{(x-20)^2}{400}\right]_{20}^{100} + 80\left(1 - \frac{100}{200}\right) \\ &= 56 \end{aligned}$$

$$\begin{aligned} \mathrm{E}\left[\left(Y^L\right)^2\right] &= \int_{20}^{100} (x-20)^2 \cdot \frac{1}{200}\,\mathrm{d}x + 80^2 S(100) \\ &= \left[\frac{(x-20)^3}{600}\right]_{20}^{100} + 80^2(0.5) \\ &= 4{,}053.3333 \end{aligned}$$

Using Equation 1.2.2.5, convert Y^L's moments to Y^P's moments:

$$\mathrm{E}\left[Y^P\right] = \frac{\mathrm{E}\left[Y^L\right]}{S(d)} = \frac{56}{0.9} = 62.2222$$

$$\mathrm{E}\left[\left(Y^P\right)^2\right] = \frac{\mathrm{E}\left[\left(Y^L\right)^2\right]}{S(d)} = \frac{4{,}053.3333}{0.9} = 4{,}503.7037$$

where $S(d) = S(20) = 1 - \dfrac{20}{200} = 0.9$.

Finally, calculate the variance.

$$\mathrm{Var}\left[Y^P\right] = 4{,}503.7037 - (62.2222)^2 = \mathbf{632.0988}$$

$\blacksquare$

1.2 Summary

10m

Payment Per Loss

Y^L is the payment per loss variable.

Policy Limits

The policy limit is the maximum amount the **insurer** will pay for a single loss. For an insurance policy with a policy limit u,

$$Y^L = X \wedge u = \begin{cases} X, & X < u \\ u, & X \geq u \end{cases}$$

$$\begin{aligned} \mathrm{E}\left[\left(Y^L\right)^k\right] &= \mathrm{E}\left[(X \wedge u)^k\right] \\ &= \int_0^u x^k \cdot f(x)\,\mathrm{d}x + u^k \cdot S(u) \\ &= \int_0^u k\,x^{k-1} S(x)\,\mathrm{d}x \end{aligned}$$

$(X \wedge u)$ is the limited loss variable, and its mean, $\mathrm{E}[X \wedge u]$, is the limited expected value.

Deductibles

The deductible is the amount the **policyholder** has to pay before receiving any reimbursement. There are two types of deductible: ordinary and franchise.

ORDINARY DEDUCTIBLE

$$Y^L = (X-d)_+ = \begin{cases} 0, & X \le d \\ X-d, & X > d \end{cases} = X - (X \wedge d)$$

$$\mathrm{E}\left[Y^L\right] = \mathrm{E}\left[(X-d)_+\right] = \mathrm{E}[X] - \mathrm{E}[X \wedge d]$$

$$\begin{aligned} \mathrm{E}\left[\left(Y^L\right)^k\right] &= \mathrm{E}\left[(X-d)_+^k\right] \\ &= \int_d^\infty (x-d)^k \cdot f(x)\,\mathrm{d}x \\ &= \int_d^\infty k\,(x-d)^{k-1} S(x)\,\mathrm{d}x \end{aligned}$$

LOSS ELIMINATION RATIO

For a policy with an ordinary deductible d,

$$\mathrm{LER} = \frac{\mathrm{E}[X \wedge d]}{\mathrm{E}[X]}$$

FRANCHISE DEDUCTIBLE

$$Y^L = \begin{cases} 0, & X \le d \\ X, & X > d \end{cases}$$

$$\mathrm{E}\left[Y^L\right] = \mathrm{E}\left[(X-d)_+\right] + d \cdot S(d)$$

Payment Per Payment

Y^P is the payment per payment variable, i.e. $Y^P = Y^L \mid Y^L > 0$.

$$\mathrm{E}\left[Y^P\right] = \frac{\mathrm{E}\left[Y^L\right]}{S(d)} \qquad \Leftrightarrow \qquad \mathrm{E}\left[Y^L\right] = \mathrm{E}\left[Y^P\right] \cdot S(d)$$

$(X - d \mid X > d)$ is the excess loss variable, and its mean, $e(d)$, is the mean excess loss.

$$e(d) = \mathrm{E}[X - d \mid X > d] = \frac{\mathrm{E}\left[(X-d)_+\right]}{S(d)}$$

SPECIAL CASES

Loss	Mean	Excess Loss	Mean Excess Loss
Exponential (θ)	θ	Exponential (θ)	θ
Uniform $(a,\, b)$	$\dfrac{a+b}{2}$	Uniform $(0,\, b-d)$	$\dfrac{b-d}{2}$
Pareto $(\alpha,\, \theta)$	$\dfrac{\theta}{\alpha-1}$	Pareto $(\alpha,\, \theta+d)$	$\dfrac{\theta+d}{\alpha-1}$
S-P Pareto $(\alpha,\, \theta)$	$\dfrac{\alpha\theta}{\alpha-1}$	Pareto $(\alpha,\, d)$	$\dfrac{d}{\alpha-1}$
Beta $(1,\, b,\, \theta)$	$\dfrac{\theta}{1+b}$	Beta $(1,\, b,\, \theta-d)$	$\dfrac{\theta-d}{1+b}$

IMPACT ON CLAIM FREQUENCY

If the number of losses N is modeled by an $(a,\, b,\, 0)$ class distribution, then the number of payments N' will follow the same distribution with modified parameters.

	N	N'
Poisson	λ	$v\lambda$
Binomial	$m,\, q$	$m,\, vq$
Negative binomial	$r,\, \beta$	$r,\, v\beta$

where $v = \Pr(X > d)$.

The Ultimate Formula

$$\mathrm{E}\left[Y^L\right] = \alpha\,(1+r)\left(\mathrm{E}\left[X \wedge \frac{m}{1+r}\right] - \mathrm{E}\left[X \wedge \frac{d}{1+r}\right]\right)$$

where

- X is the loss variable,
- d is the deductible (set to 0 if not applicable),
- u is the policy limit (set to ∞ if not applicable),
- α is the coinsurance (set to 1 if not applicable),
- r is the inflation rate (set to 0 if not applicable), and
- m is the maximum covered loss, which equals $\dfrac{u}{\alpha} + d$.

Appendix

15m

Impact of Deductibles on Claim Frequency

Let

- N be the number of losses
- N' be the number of payments (i.e. losses that exceed the deductible)
- $v = \Pr(X > d)$ be the probability that a loss exceeds the deductible

$$\begin{aligned}
\Pr\left(N' = k\right) &= \sum_{n=k}^{\infty} \Pr(N = n) \binom{n}{k} v^k (1 - v)^{n-k} \\
&= \sum_{n=k}^{\infty} \Pr(N = n) \frac{n!}{(n-k)!\,k!} v^k (1 - v)^{n-k} \\
&= \sum_{n=k}^{\infty} \Pr(N = n) \frac{n\,(n-1)\ldots(n-k+1)}{k!} v^k (1 - v)^{n-k} \\
&= \mathrm{E}\left[\frac{N\,(N-1)\ldots(N-k+1)}{k!} v^k (1 - v)^{N-k}\right] \\
&= \frac{v^k}{k!} \cdot \mathrm{E}\left[N\,(N-1)\ldots(N-k+1)(1 - v)^{N-k}\right] \\
&= \frac{v^k}{k!} \cdot P_N^{(k)}(1 - v)
\end{aligned}$$

Poisson

For $N \sim \text{Poisson}\ (\lambda)$,

$$P(z) = e^{\lambda(z-1)}$$

$$P'(z) = \lambda \cdot e^{\lambda(z-1)}$$

$$P''(z) = \lambda^2 \cdot e^{\lambda(z-1)}$$

$$\vdots$$

$$P^{(k)}(z) = \lambda^k \cdot e^{\lambda(z-1)}$$

Thus,

$$\begin{aligned}\Pr\left(N' = k\right) &= \frac{v^k}{k!} \cdot \lambda^k \cdot e^{\lambda(1-v-1)} \\ &= \frac{(v\lambda)^k e^{-v\lambda}}{k!}\end{aligned}$$

Conclude that $N' \sim \text{Poisson}\ (v\lambda)$.

■

Binomial

For $N \sim \text{Binomial}\ (m,\ q)$,

$$P(z) = [1 + q\,(z-1)]^m$$

$$P'(z) = m[1 + q\,(z-1)]^{m-1}\,(q)$$

$$P''(z) = m\,(m-1)[1 + q\,(z-1)]^{m-2}\left(q^2\right)$$

$$\vdots$$

$$P^{(k)}(z) = m\,(m-1)\ldots(m-k+1)[1+q\,(z-1)]^{m-k}\left(q^k\right)$$

Thus,

$$\begin{aligned}\Pr\left(N' = k\right) &= \frac{v^k}{k!}\cdot m\,(m-1)\ldots(m-k+1)(1-vq)^{m-k}\left(q^k\right)\\ &= \frac{m\,(m-1)\ldots(m-k+1)}{k!}(vq)^k(1-vq)^{m-k}\\ &= \binom{m}{k}(vq)^k(1-vq)^{m-k}\end{aligned}$$

Conclude that $N' \sim \text{Binomial}\,(m,\,vq)$.

$\blacksquare$

Negative Binomial

For $N \sim \text{Negative Binomial}\,(r,\,\beta)$,

$$P(z) = [1-\beta\,(z-1)]^{-r}$$

$$P'(z) = r[1-\beta\,(z-1)]^{-(r+1)}\,(\beta)$$

$$P''(z) = r\,(r+1)[1-\beta\,(z-1)]^{-(r+2)}\left(\beta^2\right)$$

$$\vdots$$

$$P^{(k)}(z) = r\,(r+1)\ldots(r+k-1)[1-\beta\,(z-1)]^{-(r+k)}\left(\beta^k\right)$$

$$P^{(k)}(z) = r\,(r+1)\ldots(r+k-1)[1-\beta\,(z-1)]^{-(r+k)}\,(\beta^k)$$

Thus,

$$\begin{aligned}\Pr\big(N' = k\big) &= \frac{v^k}{k!}\cdot r\,(r+1)\ldots(r+k-1)(1+v\beta)^{-(r+k)}\,\big(\beta^k\big)\\ &= \frac{r\,(r+1)\ldots(r+k-1)}{k!}\cdot\frac{(v\beta)^k}{(1+v\beta)^{(r+k)}}\end{aligned}$$

Conclude that $N' \sim \text{Negative Binomial}\,(r,\ v\beta)$.

$\blacksquare$

Mean Excess Loss - Special Cases

Exponential

For $X \sim \text{Exponential}\,(\theta)$ and $d > 0$,

$$\begin{aligned}S_{X-d|X>d}(x) &= \Pr(X-d > x \mid X > d)\\ &= \Pr(X > x+d \mid X > d)\\ &= \frac{\Pr(X > x+d)}{\Pr(X > d)}\\ &= \exp\left(-\frac{x+d}{\theta}\right) \div \exp\left(-\frac{d}{\theta}\right)\end{aligned}$$

$$\begin{aligned} &= \exp\left(-\frac{x+d}{\theta} + \frac{d}{\theta}\right) \\ &= \exp\left(-\frac{x}{\theta}\right) \end{aligned}$$

Thus, conclude that the excess loss:

$$(X - d \mid X > d) \sim \text{Exponential }(\theta)$$

and the mean excess loss:

$$\mathrm{E}[X - d \mid X > d] = \boldsymbol{\theta}$$

■

Uniform

For $X \sim \text{Uniform }(a,\, b)$ and $a < d < b$,

$$\begin{aligned} S_{X-d|X>d}(x) &= \Pr(X - d > x \mid X > d) \\ &= \Pr(X > x + d \mid X > d) \\ &= \frac{\Pr(X > x + d)}{\Pr(X > d)} \\ &= \frac{b - (x + d)}{b - a} \div \frac{b - d}{b - a} \\ &= \frac{b - x - d)}{b - d} \\ &= \frac{(b - d) - x}{(b - d) - 0} \end{aligned}$$

Thus, conclude that the excess loss:

Thus, conclude that the excess loss:

$$(X-d \mid X>d) \sim \text{Uniform}\,(0,\, b-d)$$

and the mean excess loss:

$$\mathrm{E}[X-d \mid X>d] = \boldsymbol{\frac{b-d}{2}}$$

$\blacksquare$

Pareto

For $X \sim \text{Pareto}\,(\alpha,\, \theta)$ and $d > 0$,

$$\begin{aligned} S_{X-d|X>d}(x) &= \Pr(X-d>x \mid X>d) \\ &= \Pr(X>x+d \mid X>d) \\ &= \frac{\Pr(X>x+d)}{\Pr(X>d)} \\ &= \left(\frac{\theta}{x+d+\theta}\right)^{\alpha} \div \left(\frac{\theta}{d+\theta}\right)^{\alpha} \\ &= \left(\frac{d+\theta}{x+d+\theta}\right)^{\alpha} \end{aligned}$$

Thus, conclude that the excess loss:

$$(X-d \mid X>d) \sim \text{Pareto}\,(\alpha,\, \theta+d)$$

and the mean excess loss:

$$\mathrm{E}[X-d \mid X>d]=\boldsymbol{\frac{\theta+d}{\alpha-1}}$$

■

Single-Parameter Pareto

For $X \sim \text{S-P Pareto}\ (\alpha,\ \theta)$ and $d>\theta$,

$$\begin{aligned} S_{X-d \mid X>d}(x) &= \Pr(X-d>x \mid X>d) \\ &= \Pr(X>x+d \mid X>d) \\ &= \frac{\Pr(X>x+d)}{\Pr(X>d)} \\ &= \left(\frac{\theta}{x+d}\right)^{\alpha} \div \left(\frac{\theta}{d}\right)^{\alpha} \\ &= \left(\frac{d}{x+d}\right)^{\alpha} \end{aligned}$$

Thus, conclude that the excess loss:

$$(X-d \mid X>d) \sim \text{Pareto}\ (\alpha,\ d)$$

and the mean excess loss:

$$\mathrm{E}[X-d \mid X>d]=\boldsymbol{\frac{d}{\alpha-1}}$$

■

Beta

For $X \sim \text{Beta}\,(a = 1,\, b,\, \theta)$

$$\begin{aligned} f_X(x) &= \frac{\Gamma(1+b)}{\Gamma(1)\,\Gamma(b)}\left(\frac{x}{\theta}\right)\left(1-\frac{x}{\theta}\right)^{b-1}\frac{1}{x} \\ &= \frac{b}{\theta}\left(1-\frac{x}{\theta}\right)^{b-1} \end{aligned}$$

$$\begin{aligned} F_X(x) &= \int_0^x \frac{b}{\theta}\left(1-\frac{t}{\theta}\right)^{b-1}\mathrm{d}t \\ &= \left[-\left(1-\frac{t}{\theta}\right)^{b}\right]_0^x \\ &= 1-\left(1-\frac{x}{\theta}\right)^{b} \\ &= 1-\left(\frac{\theta-x}{\theta}\right)^{b} \end{aligned}$$

For $0 < d < \theta$,

$$\begin{aligned} S_{X-d|X>d}(x) &= \Pr(X-d > x \mid X > d) \\ &= \Pr(X > x+d \mid X > d) \\ &= \frac{\Pr(X > x+d)}{\Pr(X > d)} \\ &= \left(\frac{\theta-x-d}{\theta}\right)^{b} \div \left(\frac{\theta-d}{\theta}\right)^{b} \\ &= \left(\frac{\theta-d-x}{\theta-d}\right)^{b} \end{aligned}$$

Thus, conclude that the excess loss:

$$(X - d \mid X > d) \sim \text{Beta}\left(a = 1,\ b,\ \theta - d\right)$$

and the mean excess loss:

$$\text{E}[X - d \mid X > d] = \boldsymbol{\frac{\theta - d}{1 - b}}$$

$\blacksquare$