Normal and Lognormal Approximations

25M

There are times when we need to approximate the distribution of the aggregate loss. Usually, this is because calculating exact probabilities from the aggregate loss distribution is time-consuming. When dealing with the sum of i.i.d. random variables, the **normal** distribution is a good approximation. Based on the ***Central Limit Theorem***, this approximation improves as the number of random variables increases.

The aggregate loss is the sum of N i.i.d. random variables, X_i. Thus, we approximate S with a normal distribution.

$$S \sim \text{Normal}\left(\text{E}[S],\ \text{Var}[S]\right)$$

Let's see this in action.

Example S2.5.3.1

The number of losses reported per year has mean 50 and variance 300. The individual losses have mean 30 and standard deviation 60. The number of losses and individual loss amounts are independent.

Using the normal approximation, determine the probability that the annual aggregate loss is greater than 1,700.

Solution

Apply the compound mean and variance formulas.

$$\begin{aligned} \mu_S &= \mu_N \mu_X \\ &= 50(30) \\ &= 1{,}500 \end{aligned}$$

$$\begin{aligned}\sigma_S^2 &= \mu_N \sigma_X^2 + \sigma_N^2 \mu_X^2 \\ &= 50\left(60^2\right) + 300\left(30^2\right) \\ &= 450{,}000\end{aligned}$$

Using the normal approximation, the aggregate loss follows a normal distribution with mean 1,500 and variance 450,000.

$$S \sim \text{Normal}\ (1{,}500,\ 450{,}000)$$

Calculate the desired probability.

$$\begin{aligned}\Pr(S > 1{,}700) &= 1 - \Pr(S \le 1{,}700) \\ &= 1 - \Pr\left(Z \le \frac{1{,}700 - 1{,}500}{\sqrt{450{,}000}}\right) \\ &= 1 - \Phi(0.29814) \\ &= 1 - 0.6172 \\ &= \mathbf{0.3828}\end{aligned}$$

Continuity Correction

The normal distribution is continuous. When using a continuous distribution to approximate a **discrete** distribution, we need to apply *continuity correction*.

Coach's Remarks

We should always use continuity correction when approximating a discrete quantity with a continuous distribution.

While it is true that the frequency model is usually discrete and the severity model is usually continuous, a question can certainly model the frequency using a continuous distribution and/or model the severity using a discrete distribution.

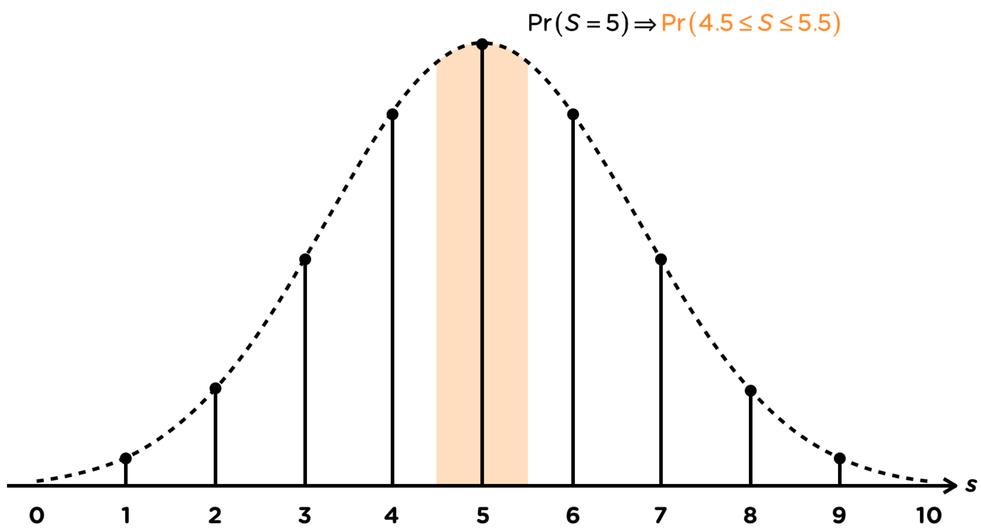
The aggregate loss model is discrete when **both** the frequency and severity models are discrete.

Consider the following example:

Assume S can be any integer from 0 to 10. Determine the probability that S equals 5 using the normal approximation.

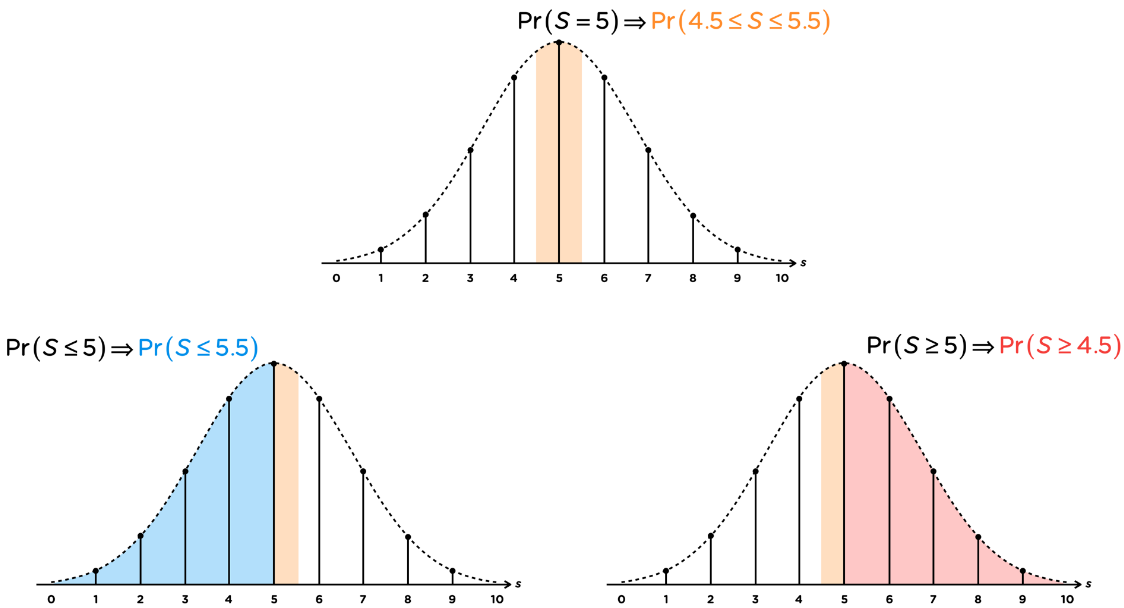
Since the normal distribution is continuous, approximating S using the normal distribution would imply $\Pr(S=5)=0$ when its actual value is non-zero.

To address this issue, we approximate $\Pr(S=5)$ as $\mathbf{Pr(4.5 \leq S \leq 5.5)}$.

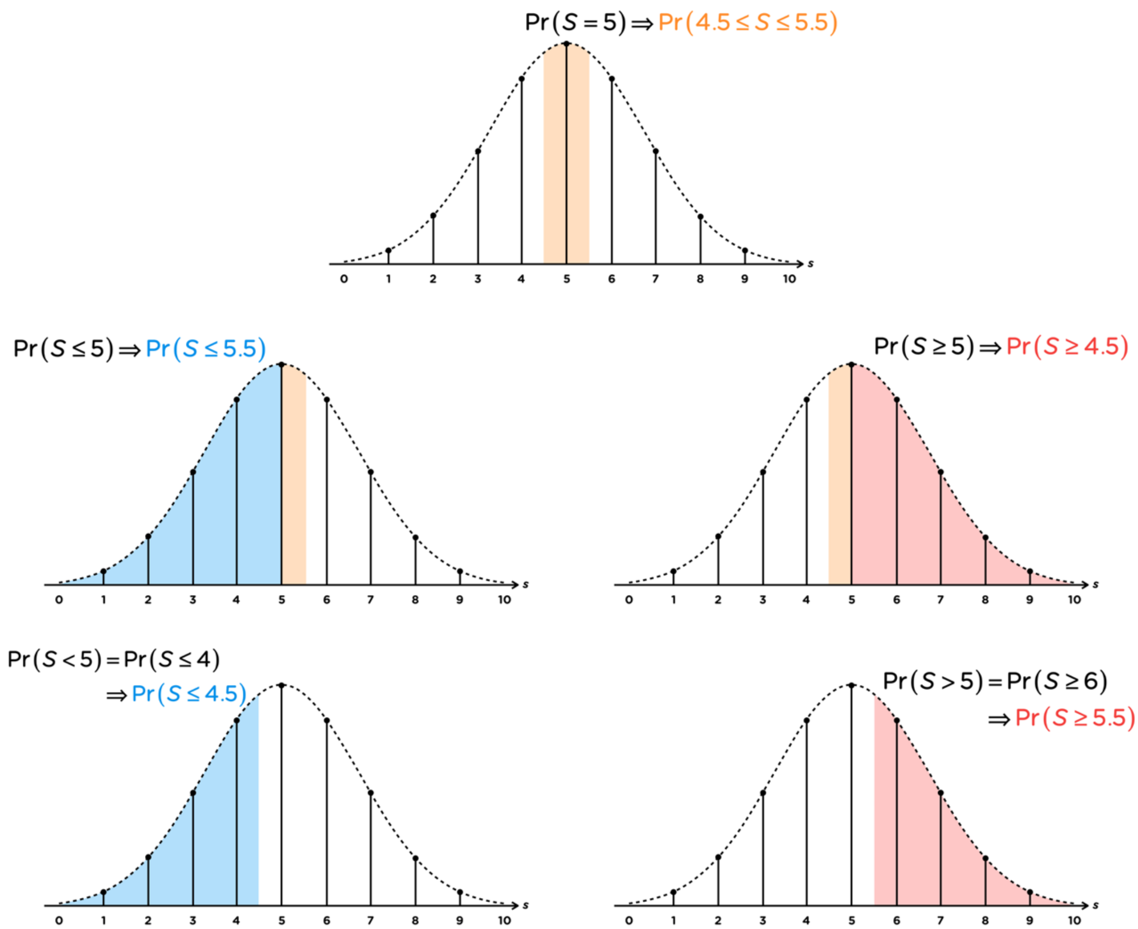


We can use the approximation above to derive other normal-approximated probabilities. For example, if asked for $\Pr(S \leq 5)$, we need to include the range 4.5 to 5.5, since we want to include the probability of $S=5$. As a result, $\Pr(S \leq 5)$ is approximated by $\Pr(S \leq 5.5)$.

Likewise, $\Pr(S \geq 5)$ is approximated by $\Pr(S \geq 4.5)$.



$\Pr\left(S < 5\right)$, the complement of $\Pr\left(S \geq 5\right)$, is approximated by $1 - \Pr\left(S \geq 4.5\right)$, or $\Pr\left(S \leq 4.5\right)$.



Example S2.5.3.2

The annual number of road crashes in Town ABC follows a Poisson distribution with mean 30. For each road crash, the number of deaths has the following distribution:

Number of Deaths	Probability
0	80%
1	10%
2	5%
3	2.5%
4	2.5%

The number of road crashes and the number of deaths for each road crash are mutually independent.

Using the normal approximation with continuity correction, determine the probability of having more than 15 road deaths in Town ABC in a year.

Solution

Let N be the number of road crashes in a year.

$$N \sim \text{Poisson}\ (30)$$

$$\mu_N = \sigma_N^2 = 30$$

Let X be the number of deaths in a road crash. Calculate the severity mean and variance.

$$\begin{aligned}\mu_X &= 0.8(0) + 0.1(1) + 0.05(2) + 0.025(3) + 0.025(4) \\ &= 0.375\end{aligned}$$

$$\begin{aligned}\sigma_X^2 &= \left[0.8\left(0^2\right) + 0.1\left(1^2\right) + 0.05\left(2^2\right) + 0.025\left(3^2\right) + 0.025\left(4^2\right)\right] - 0.37 \\ &= 0.7844\end{aligned}$$

Let S be the total road deaths in a year. Calculate the aggregate mean and variance.

$$\begin{aligned}\mu_S &= \mu_N \mu_X \\ &= 30(0.375) \\ &= 11.25\end{aligned}$$

$$\begin{aligned}\sigma_S^2 &= \mu_N \sigma_X^2 + \sigma_N^2\,\mu_X^2 \\ &= 30(0.7844) + 30\left(0.375^2\right) \\ &= 27.75\end{aligned}$$

S is approximated by a normal distribution.

$$S \sim \text{Normal}\ (11.25,\ 27.75)$$

Because S is discrete, we need to use the continuity correction.

$$\Pr(S > 15) \Rightarrow \Pr(S > 15.5)$$

Calculate the final answer.

$$\begin{aligned}\Pr(S > 15.5) &= 1 - \Pr(S \le 15.5) \\ &= 1 - \Pr\left(Z \le \frac{15.5 - 11.25}{\sqrt{27.75}}\right) \\ &= 1 - \Phi(0.80678) \\ &= 1 - 0.7901 \\ &= \mathbf{0.2099}\end{aligned}$$

■

Sometimes, S is discrete but does not assume the value of every integer. For example, the possible values of S might be 50, 200, 500, ... and nothing in between. In that case, we apply continuity correction using the **midpoints** between possible values. This is consistent with the way we apply continuity correction when all integer values are possible.

An example will better illustrate this idea.

Example S2.5.3.3

For aggregate loss, S:

- The number of losses, N, has the following distribution.

n	$\Pr(N = n)$
1	0.8
2	0.2

- The individual loss amount, X, has the following distribution.

x	$\Pr(X = x)$
0	0.1
200	0.6
$1{,}000$	0.3

- The number of losses and the individual loss amount are mutually independent.

Using the normal approximation with continuity correction, determine the probability that

1. S is at most 400.
2. S is greater than 800.

Solution to (1)

Calculate the frequency mean and variance.

$$\begin{aligned}\mu_N &= 1(0.8) + 2(0.2) \\ &= 1.2\end{aligned}$$

$$\begin{aligned}\sigma_N^2 &= \left[1^2(0.8) + 2^2(0.2)\right] - 1.2^2 \\ &= 0.16\end{aligned}$$

Calculate the severity mean and variance.

$$\begin{aligned}\mu_X &= 0(0.1) + 200(0.6) + 1{,}000(0.3) \\ &= 420\end{aligned}$$

$$\begin{aligned}\sigma_X^2 &= \left[0^2(0.1) + 200^2(0.6) + 1{,}000^2(0.3)\right] - 420^2 \\ &= 147{,}600\end{aligned}$$

Calculate the aggregate mean and variance.

$$\begin{aligned}\mu_S &= 1.2(420) \\ &= 504\end{aligned}$$

$$\begin{aligned}\sigma_S^2 &= 1.2(147{,}600) + 0.16\left(420^2\right) \\ &= 205{,}344\end{aligned}$$

The question asks for $\Pr(S \le 400)$, which includes $\Pr(S = 400)$.

To apply continuity correction, the probability of $S = 400$ should include the probability of S from

- the midpoint between 400 and its previous possible value to
- the midpoint between 400 and its next possible value.

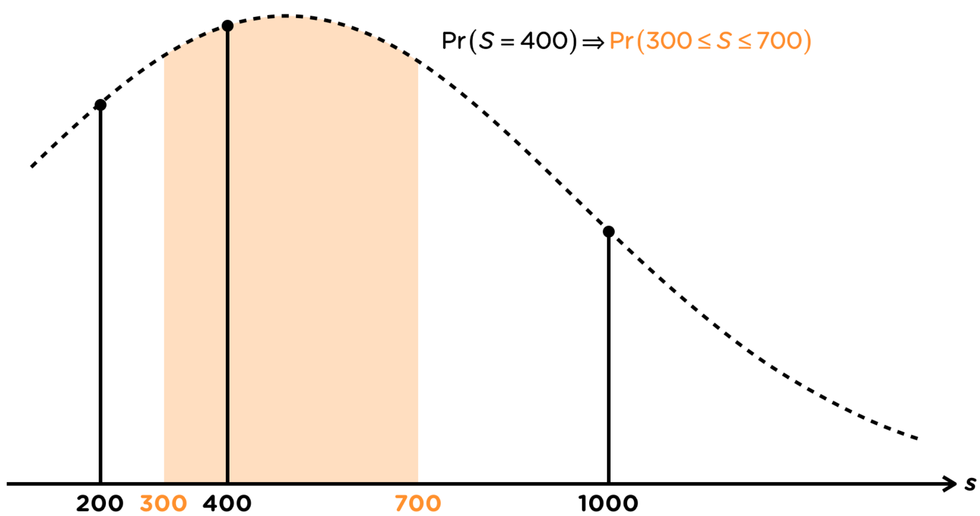
The possible values of S include:

s	**Description**
⋮	⋮

s	Description
200	A claim of 200; two claims of 0 and 200
400	Two claims of 200
1,000	A claim of 1,000; two claims of 0 and 1,000
⋮	⋮

Thus, $\Pr(S=400)$ is approximated by

$$\Pr\left(\frac{200+400}{2} \le S \le \frac{400+1{,}000}{2}\right) = \Pr(300 \le S \le 700)$$



Consequently, $\Pr(S \le 400)$ is approximated by $\Pr(S \le 700)$.

Solution to (1) (cont.)

The final answer is

$$\Pr(S \le 400) \Rightarrow \Pr(S \le 700)$$

$$\begin{aligned}\Pr(S \le 700) &= \Pr\left(Z \le \frac{700 - 504}{\sqrt{205{,}344}}\right)\\ &= \Phi(0.43253)\\ &= \mathbf{0.66732}\end{aligned}$$

■

Solution to (2)

Since the aggregate loss is discrete, $\Pr(S > 800)$ is equivalent to $\Pr(S > 400)$ or $\Pr(S \ge 1{,}000)$. Both are approximated by $\Pr(S > 700)$. Thus,

$$\Pr(S > 800) \Rightarrow \Pr(S > 700)$$

$\Pr(S \le 700)$ has been calculated to be 0.66732. Therefore,

$$\Pr(S > 700) = 1 - 0.66732 = \mathbf{0.33268}$$

■

Coach's Remarks

Here is another way to calculate the aggregate mean and variance.

s	$\Pr(S = s)$
0	$0.8(0.1) + 0.2(0.1^2) = 0.082$
200	$0.8(0.6) + 0.2[2(0.1)(0.6)] = 0.504$
400	$0.2(0.6^2) = 0.072$
1,000	$0.8(0.3) + 0.2[2(0.1)(0.3)] = 0.252$
1,200	$0.2[2(0.6)(0.3)] = 0.072$
2,000	$0.2(0.3^2) = 0.018$

$$\begin{aligned}\mu_S &= 0(0.082) + 200(0.504) + 400(0.072) + 1{,}000(0.252) + 1{,}200(0.0\\ &= 504\end{aligned}$$

$$\begin{aligned}\sigma_S^2 &= \left[0^2(0.082) + 200^2(0.504) + 400^2(0.072) + 1{,}000^2(0.252) + 1{,}20\right.\\ &= 205{,}344\end{aligned}$$

Lognormal Approximation

Besides normal, **lognormal** may be good candidates for approximating the aggregate loss distribution. The process is similar to normal approximation, except we will need to determine the distribution parameters by equating the means and variances. The steps are:

1. Calculate the aggregate mean and variance.
2. Equate to lognormal mean and variance to solve for the parameters of the approximating distribution.
3. Calculate the desired quantity based on the approximating distribution.

Coach's Remarks

We follow the exact same steps when applying the normal approximation. However, since the normal parameters are its mean and variance, there is no additional calculation needed to determine the parameters.

Example S2.5.3.4

The aggregate loss, S, has mean 5 and variance 1.

Determine the mode of the aggregate loss distribution

1. using the lognormal approximation.
2. using the normal approximation.

Solution to (1)

Lognormal approximation

Equate the lognormal's mean and variance to the aggregate mean and variance. The moment formulas are provided on the exam table.

$$e^{\mu+0.5\sigma^2} = 5$$

$$\left(e^{\mu+0.5\sigma^2}\right)^2\left(e^{\sigma^2}-1\right) = 1$$

Solve for the lognormal's parameters μ and σ^2.

$$\begin{aligned}\frac{\left(e^{\mu+0.5\sigma^2}\right)^2\left(e^{\sigma^2}-1\right)}{\left(e^{\mu+0.5\sigma^2}\right)^2} &= \frac{1}{5^2}\\ e^{\sigma^2}-1 &= 0.04\\ \sigma^2 &= \ln\left(0.04+1\right)\\ &= 0.0392\end{aligned}$$

$$\begin{aligned}e^{\mu+0.5(0.0392)} &= 5\\ \mu &= \ln\left(5\right)-0.5(0.0392)\\ &= 1.5898\end{aligned}$$

Therefore,

$$S \sim \text{Lognormal}\left(1.5898,\ 0.0392\right)$$

The mode of S is