Aggregate Loss Models
Chapter 9

Note. Here is a proof for \( E[X_1 + \cdots + X_N | N = n] = E[X_1 + \cdots + X_n] = nE(X_1) \): We prove this for \( E[S|N = 2] = 2E(X_1) \). First of all let \( \Omega \) be the common support of the random variables \( \{X_1, \ldots, X_n\} \) (these variables are identically distributed and so have identical density function and so have identical support).

**Next Step.** Note that on the event \( \{N = 2\} \) the random variable \( S \) becomes \( X_1 + X_2 \) so it takes on all values \( s = x_1 + x_2 \) where \( x_1 \in \Omega \) and \( x_2 \in \Omega \). So the support of the conditional random variable \( S|\{N = 2\} \) is the set \( \Delta = \{s = x_1 + x_2 : x_1 \in \Omega, x_2 \in \Omega\} \).

**Next Step.** For \( s \in \Delta \) we have

\[
P(S = s | N = 2) = P(S = x_1 + x_2 | N = 2) = P(X_1 + X_2 = x_1 + x_2 | N = 2)
\]

\[
= \frac{P(X_1 + X_2 = x_1 + x_2 , N = 2)}{P(N = 2)} = \frac{P(X_1 + X_2 = x_1 + x_2)P(N = 2)}{P(N = 2)} = P(X_1 + X_2 = x_1 + x_2)
\]

**Next Step.**

\[
E(S|N = 2) = \sum_{s \in \Delta} s P(S = s | N = 2) = \sum_{s} \sum_{x_1 \in \Omega, x_2 \in \Omega} (x_1 + x_2) P(X_1 + X_2 = x_1 + x_2)
\]

\[
= E(X_1 + X_2) = E(X_1) + E(X_2) = 2E(X_1) \quad \checkmark
\]

**Theorem.** Let \( \{X_1, X_2, \ldots \} \) be an independent and identically distributed sample from a claim. If the number of claims \( N \) is independent of the claim amount, then

\[
E(S) = E(N) E(X_1)
\]

**Proof.**

\[
E(S) = E\left(\sum_{j=1}^{N} X_j \right)
\]

\[
= \sum_{k=0}^{\infty} P(N = k) E(\sum_{j=1}^{N} X_j | N = k)
\]

\[
= \sum_{k=0}^{\infty} P(N = k) E(\sum_{j=1}^{k} X_j) \quad \text{independence is used here - check it out}
\]

\[
= \sum_{k=0}^{\infty} P(N = k) k E(X_1)
\]

\[
= \left( \sum_{k=0}^{\infty} k P(N = k) \right) E(X_1) = E(N) E(X_1)
\]
**Theorem.** Let \( \{X_1, \ldots, X_n\} \) be an independent and identically distributed sample from a claim. If the number of claims \( N \) is independent of the claim amount, then

(i) \[ E(\sum_{i=1}^{N} X_i|N) = N E(X_1) \]

(ii) \[ \text{Var}(\sum_{i=1}^{N} X_i|N) = N \text{Var}(X_1) \]

(iii) \[ \text{Var}(\sum_{i=1}^{N} X_i) = E(N) \text{Var}(X_1) + \text{Var}(N) E(X_1)^2 \quad \text{compound variance formula} \]

**Proof of (i).** For each fixed value \( n \) of \( N \) we have

\[
E(\sum_{i=1}^{N} X_i|N)(n) = E(\sum_{i=1}^{n} X_i|N = n) = E(\sum_{i=1}^{n} X_i) = n E(X_1)
\]

Since this is true for all \( n \) we have proved the claim of part (i).

**Proof of (ii).** For each fixed value \( n \) of \( N \) we have

\[
\text{Var}(\sum_{i=1}^{N} X_i|N)(n) = \text{Var}(\sum_{i=1}^{n} X_i|N = n) = \text{Var}(\sum_{i=1}^{n} X_i) = n \text{Var}(X_1)
\]

Since this is true for all \( n \) we have proved the claim of part (ii).

**Proof of (iii).**

\[
\text{Var}(S) = E[\text{Var}(S|N)] + \text{Var}[E(S|N)]
\]

\[
= E[\text{Var}(\sum_{i=1}^{N} X_i) + \text{Var}(E(\sum_{i=1}^{N} X_i))]
\]

\[
= E[N \text{Var}(X_1)] + \text{Var}(N E(X_1))
\]

\[
= E(N) \text{Var}(X_1) + E(X_1)^2 \text{Var}(N)
\]

**Note.** When the primary distribution is Poisson(\( \lambda \)), then the compound variance formula reduces to

\[ \text{Var}(S) = \lambda E(X^2) \quad \text{compound variance for Poisson primary} \]
**Example.** Let $N \sim \text{Poisson}(2)$ and let $\{X_1, X_2, \ldots\}$ be an independent identically distributed sequence with common distribution $N(15, \sigma^2 = 5)$. Calculate $E(\sum_{i=1}^{N} X_i)$ and $\text{Var}(\sum_{i=1}^{N} X_i)$.

**Solution.** We have

$$E(N) = 2, \quad \text{Var}(N) = 2$$

Then from the formulas above, we will have:

$$E(\sum_{i=1}^{N} X_i) = E(N) E(X_1) = (2)(15) = 30$$

$$\text{Var}(\sum_{i=1}^{N} X_i) = E(N) \text{Var}(X_1) + \text{Var}(N) E(X_1)^2 = (2)(5) + (2)(15)^2 = 460$$
Aggregate loss: total amount of loss in one period.

Individual risk model: is a model where we study the aggregate loss $S = X_1 + \cdots + X_n$ where the losses $\{X_1, \ldots, X_n\}$ are (stochastically) independent.

Collective risk model: is a model where we study the aggregate loss $S = X_1 + \cdots + X_N$ where the losses $\{X_1, X_2, \ldots\}$ are (stochastically) independent and identically distributed (i.i.d.) and where $N$ is a discrete random variable (called frequency) indicating the number of losses, and that the $X_i$’s are independent of $N$. The distribution of $N$ is called the primary distribution and the common distribution of $X_i$’s is called the secondary distribution. The collective model is a special case of compound distribution in which we add up a random number of identically distributed random variables.

In an individual risk model, $n$ is the number of insureds and $X_i$ is the claim size for the individual $i$. So under this model we can have $P(X_i = 0) > 0$, i.e. there is positive chance that the insured $i$ will not have any claim in the period of study. But in the collective model, $N$ represents the variable number of claims and $X_i$ denotes the $i$-th amount of the claim that has occurred. So, under the collective model we have $P(X_i = 0) = 0$.

Convention: For $N = 0$ we set $S = 0$.

Example *, Let $S$ have a Poisson frequency distribution with parameter $\lambda = 5$. The individual claim amount has the following distribution:

<table>
<thead>
<tr>
<th>$x$</th>
<th>$f_X(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>0.8</td>
</tr>
<tr>
<td>500</td>
<td>0.16</td>
</tr>
<tr>
<td>1000</td>
<td>0.04</td>
</tr>
</tbody>
</table>

Calculate the probability that aggregate claims will be exactly 600.

Solution.

Case 1. $N = 2$ in which we have
\begin{align*}
\left\{\begin{array}{ll}
X_1 = 100, & X_2 = 500 \\
X_1 = 500, & X_2 = 100
\end{array}\right.
\end{align*}

with probability

\[
P(S = 600, N = 2) = P(N = 2, X_1 = 100, X_2 = 500) + P(N = 2, X_1 = 500, X_2 = 100) = 2 \left( \frac{e^{-5.5^2}}{2!} \right) (0.8)(0.16) = 0.0216
\]

**Case 2.** \(N = 6\) in which we have \(X_1 = \cdots = X_6 = 100\). Then:

\[
P(S = 600, N = 6) = P(N = 6, X_1 = 100, \ldots, X_6 = 100) = \left( \frac{e^{-5.5^6}}{6!} \right) (0.8)^6 = 0.0383
\]

\[
P(S = 600) = 0.0216 + 0.0383 = 0.0599
\]

**Example.** Consider the following collective risk model.

<table>
<thead>
<tr>
<th># of claims</th>
<th>probability</th>
<th>claim size</th>
<th>probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>(\frac{1}{3})</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>(\frac{3}{5})</td>
<td>25</td>
<td>(\frac{1}{3})</td>
</tr>
<tr>
<td></td>
<td></td>
<td>150</td>
<td>(\frac{2}{3})</td>
</tr>
<tr>
<td>2</td>
<td>(\frac{1}{5})</td>
<td>50</td>
<td>(\frac{2}{3})</td>
</tr>
<tr>
<td></td>
<td></td>
<td>200</td>
<td>(\frac{1}{3})</td>
</tr>
</tbody>
</table>

Find the variance of the aggregate loss.
Solution.

\[
S = \begin{cases} 
0 & \text{Prob.} = \frac{1}{5} \\
1 & \begin{cases} 
S = 25 & \text{Prob.} = \left(\frac{3}{5}\right)\left(\frac{2}{3}\right) = \frac{1}{5} \\
S = 150 & \text{Prob.} = \left(\frac{3}{5}\right)\left(\frac{1}{3}\right) = \frac{2}{5} \\
2 & \begin{cases} 
S = 100 & \text{Prob.} = \left(\frac{1}{5}\right)\left(\frac{2}{3}\right) = \frac{1}{15} \\
S = 250 & \text{Prob.} = 2\left(\frac{1}{5}\right)\left(\frac{2}{3}\right) = \frac{4}{45} \\
S = 400 & \text{Prob.} = \left(\frac{1}{5}\right)\left(\frac{1}{3}\right) = \frac{1}{15} 
\end{cases}
\end{cases}
\]

\[E(S) = (0)(\frac{1}{5}) + (25)(\frac{1}{5}) + (150)(\frac{2}{5}) + (100)(\frac{4}{15}) + (250)(\frac{4}{45}) + (400)(\frac{1}{15}) = 105\]

\[E(S^2) = (0)^2(\frac{1}{5}) + (25)^2(\frac{1}{5}) + (150)^2(\frac{2}{5}) + (100)^2(\frac{4}{15}) + (250)^2(\frac{4}{45}) + (400)^2(\frac{1}{15}) = 19125\]

\[\text{Var}(S) = E(S^2) - E(S)^2 = 19125 - 105^2 = 8100\]

**Example (Problem 9.55 of the textbook).** A population has two classes of drivers. The number of accidents per individual driver has a geometric distribution. For a driver selected at random from Class I, the geometric distribution parameter has a uniform distribution over the interval (0,1). Twenty-five percent of the drivers are in Class I. All drivers in Class II have expected number of claims 0.25. For a driver selected at random from this population, determine the probability of exactly two accidents.

**Solution.**

\[
P(N = 2 \mid \text{Class I}) = \int_0^1 P(N = 2 \mid \beta)(1) d\beta = \int_0^1 \frac{\beta^2}{(1 + \beta)^3} d\beta
\]

\[
= \int_1^2 \frac{(u - 1)^2}{u^3} du \\
= \int_1^2 \frac{u^2 - 2u + 1}{u^3} du
\]

\[
= \int_1^2 \frac{1}{u} - \frac{2}{u^2} + \frac{1}{u^3} du
\]

\[
= \int_1^{\infty} \left(\frac{1}{u} - 2u^{-2} + u^{-3}\right) du
\]

6
\[
\ln |u| + 2u^{-1} + \left. \frac{u^2}{2} \right|_{u=1}^{u=2} = 0.068
\]

\[-\frac{\beta^2}{(1 + \beta)^3} = \frac{(0.25)^2}{(1.25)^3} = 0.032
\]

\[
P(N = 2 | \text{Class II}) = P(N = 2 | \text{class II})P(\text{class II}) + P(N = 2 | \text{class I})P(\text{class I}) =
(0.068)(0.25) + (0.032)(0.75) = 0.041
\]

**Example.** There are two groups of good and bad drivers in the city. The good drivers make up 70% of the population of the drivers. The annual claim amount of each good driver is uniformly distributed on \((0, 3000)\). The annual claim for each bad driver is exponentially distributed with mean \(\theta = 5000\). As many as 100 drivers are chosen at random. Find the mean and variance of the aggregate claim.

**Solution.** We have \(S = X_1 + \cdots + X_{100}\), where each \(X_i\) is Uniform\((0, 5000)\) with probability 0.7 and is Exponential\((\theta = 5000)\) with probability 0.3.

\[
E(X) = E(X | \text{good})P(\text{good}) + E(X | \text{bad})P(\text{bad}) = (1500)(0.7) + (5000)(0.3) = 2550
\]

\[
E(X^2 | \text{good}) = \int_0^{3000} x^2 \frac{1}{3000} dx = \frac{1}{9000} \left[ x^3 \right]_0^{3000} = 3000,000
\]

Using the table, for Exponential with mean \(\theta\) the \(k\)-th moment is \(\theta^k \Gamma(k + 1)\). Therefore:

\[
E(X^2 | \text{bad}) = (5000)^2 \Gamma(3) = (25,000,000)(2) = 50,000,000
\]

\[
E(X^2) = E(X^2 | \text{good})P(\text{good}) + E(X^2 | \text{bad})P(\text{bad})
= (3000,000)(0.7) + (25,000,000)(0.3) = 9,600,000
\]

\[
\text{Var}(X) = E(X^2) - E(X)^2 = 9,600,000 - (2550)^2 = 3,097,500
\]

\[
E(S) = 100 E(X) = 255,000
\]

\[
\text{Var}(S) = 100 \text{Var}(X) = 309,750,000
\]
Some statistics of compound distribution

Note 1:
If $P_X(z) = E[z^X]$ is the PGF of a random variable $X$, then

$$P_X'(z) = E\left[\frac{d}{dz}(z^X)\right] = E[X z^{X-1}]$$

$$P_X''(z) = E\left[\frac{d}{dz}(X z^{X-1})\right] = E[X(X - 1) z^{X-2}]$$

and in general, for all $k$'s we have:

$$P_X^{(k)}(z) = E\left[X(X - 1) \cdots (X - k + 1) z^{X-k}\right]$$

By putting $z = 1$ in both sides of these equalities, we get:

$$\begin{align*}
P_X'(1) &= E(X) \\
P_X''(1) &= E[X(X - 1)] \\
P_X^{(k)}(1) &= E[X(X - 1) \cdots (X - k + 1)]
\end{align*}$$

Note 2:
For the aggregate model we have:
\[ P_S(z) = E(z^S) = E(z^0) P(N = 0) + \sum_{k=1}^{\infty} E\left[z^{X_1+\cdots+X_k} \mid N = k\right] P(N = k) \]

\[ = P(N = 0) + \sum_{k=1}^{\infty} E\left[z^{X_1+\cdots+X_k}\right] P(N = k) \]

\[ = P(N = 0) + \sum_{k=1}^{\infty} E\left[z^{X_1}\right]^k P(N = k) \quad \text{independent and identical distribution} \]

\[ = P(N = 0) + \sum_{k=1}^{\infty} w^k P(N = k) \quad w = E(z^{X_1}) \]

\[ = \sum_{k=0}^{\infty} w^k P(N = k) \]

\[ = E\left(w^N\right) = P_N(w) = P_N(P_X(z)) \]

So

\[ P_S(z) = P_N(P_X(z)) \]

Then the moment generating function of the aggregate \( S \) is:

\[ M_S(t) = P_S(e^t) = P_N[P_X(e^t)] = P_N[M_X(t)] \]

Then:

\[ \left\{ \begin{array}{l}
M'_S(t) = P'_N[M_X(t)] M'_X(t) \\
M''_S(t) = P''_N[M_X(t)] (M'_X(t))^2 + P'_N[M_X(t)] M''_X(t)
\end{array} \right. \]

\[ E(S) = M'_S(0) = P'_N[M_X(0)] M'_X(0) = P'_N(1) M'_X(0) = E(N) E(X) \]
\[
E(S^2) = M''_{S}(0) = P''_{N}[M_X(0)] \left( M'_X(0) \right)^2 + P'_{N}[M_X(0)] M''_{X}(0)
\]
\[
= P''_{N}(1) \left( M'_X(0) \right)^2 + P'_{N}(1) M''_{X}(0)
\]
\[
= E[N(N-1)]E(X)^2 + E(N)E(X^2)
\]
\[
= E(N) \left\{ E(X^2) - E(X)^2 \right\} + E(N^2)E(X)^2
\]
\[
= E(N) \text{Var}(X) + E(N^2)E(X)^2
\]

Then
\[
\text{Var}(S) = E(S^2) - E(S)^2
\]
\[
= \left\{ E(N) \text{Var}(X) + E(N^2)E(X)^2 \right\} - E(N)^2E(X)^2
\]
\[
= E(N) \text{Var}(X) + \left\{ E(N^2) - E(N)^2 \right\} E(X)^2
\]
\[
= E(N) \text{Var}(X) + \text{Var}(N)E(X)^2
\]

**Note 3:**

**Definition.** By a **Compound Poisson** distribution we mean an aggregate distribution
\[ S = X_1 + \cdots + X_N \] where \( N \) is Poisson.

**Example (problem 9.57 of the textbook)** *. Aggregate losses have a compound Poisson claim distribution with \( \lambda = 3 \) and individual claim amount distribution \( p(1) = 0.4, p(2) = 0.3, p(3) = 0.2, \) and \( p(4) = 0.1 \). Determine the probability that aggregate losses do not exceed 3.

**Solution** *
\[
P(S = 0) = P(N = 0) = e^{-3}
\]
Next:
\[ P(S = 1) = P(N = 1, X_1 = 1) = P(N = 1)P(X_1 = 1) = (e^{-3})0.4 = 1.2e^{-3} \]

Next:
\[ P(S = 2) = P(N = 1, X_1 = 2) + P(N = 2, X_1 = 1, X_2 = 1) = P(N = 1)P(X_1 = 2) + P(N = 2)P(X_1 = 1)P(X_2 = 1) = (e^{-3})0.3 + (e^{-3})0.4 = 1.62e^{-3} \]

Next:
\[ P(S = 3) = P(N = 1, X_1 = 3) + P(N = 2, X_1 = 1, X_2 = 2) + P(N = 2, X_1 = 2, X_2 = 1) + P(N = 3, X_1 = 1, X_2 = 1, X_3 = 1) = (e^{-3})0.2 + 2(e^{-3})0.4(0.3) + (e^{-3})3(0.4)^3 = 1.968e^{-3} \]

Adding up the probabilities such found, we have
\[ P(S \leq 3) = 5.788 e^{-3} = 0.288 \]

**Theorem.** Suppose that the random variables \( \{S_1, \ldots, S_n\} \) are compound Poisson with parameters \( \lambda_1, \lambda_2, \ldots, \lambda_n \) and severity distribution functions \( F_1(x), \ldots, F_n(x) \). If the \( S_i \)'s are independent, then the sum \( S = S_1 + \cdots + S_n \) has a compound Poisson distribution with parameter \( \lambda = \lambda_1 + \cdots + \lambda_n \). Further we have the following:

(i): If the severity distribution functions for \( S_1, \ldots, S_n \) are \( F_1(x), \ldots, F_n(x) \), then the severity distribution function for \( S \) is \( F(x) = \sum_{j=1}^{n} \frac{\lambda_j}{\lambda} F_j(x) \)

(ii): If the severity density functions for \( S_1, \ldots, S_n \) are \( f_1(x), \ldots, f_n(x) \), then the severity density function for \( S \) is \( f(x) = \sum_{j=1}^{n} \frac{\lambda_j}{\lambda} f_j(x) \)

(iii): If the severity PGF functions for \( S_1, \ldots, S_n \) are \( P_1(z), \ldots, P_n(z) \), then the severity PGF function for \( S \) is \( P(z) = \sum_{j=1}^{n} \frac{\lambda_j}{\lambda} P_j(z) \)

(iv): If the severity MGF functions for \( S_1, \ldots, S_n \) are \( M_1(t), \ldots, M_n(t) \), then the severity MGF function for \( S \) is \( M(t) = \sum_{j=1}^{n} \frac{\lambda_j}{\lambda} M_j(t) \)

**Proof.** Let \( P_j(z) \) be the PGF of \( F_j(x) \). Let \( N_j \) be the Poisson random variable for \( S_j \). Then
\[ P_{S_j}(z) = P_{N_j}(P_j(z)) = e^{\lambda_j (P_j(z)-1)} \]
And:

\[ P_S(z) = \prod_{j=1}^{n} P_{S_j}(z) \]

\[ = \prod_{j=1}^{n} e^{\lambda_j (P_j(z) - 1)} \]

\[ = \exp \left( \sum_{j=1}^{n} \{ \lambda_j (P_j(z) - 1) \} \right) \]

\[ = \exp \left( \sum_{j=1}^{n} \lambda_j P_j(z) - \sum_{j=1}^{n} \lambda_j \right) \]

\[ = \exp \left( \lambda \left\{ \sum_{j=1}^{n} \frac{\lambda_j}{\lambda} P_j(z) - 1 \right\} \right) \]

Because of the Uniqueness Theorem, the distribution of \( S \) is compound Poisson with parameter \( \lambda = \lambda_1 + \cdots + \lambda_n \) and severity distribution whose PGF is \( P(z) = \sum_{j=1}^{n} \frac{\lambda_j}{\lambda} P_j(z) \).

But as we know from the section on “mixing”, this is just the PGF of the distribution function

\[ F(x) = \sum_{j=1}^{n} \frac{\lambda_j}{\lambda} F_j(z) \]
Normal and Log-normal approximation for aggregate risk models

In individual risk models the $X_i$'s are assumed independent but no assumption on identical distribution is made. If in a particular problem we are sure of identical distribution, and if the number of individuals, $n$, is large, then we can use the Central Limit Theorem. But in collective risk models the number $N$ of terms in $S = X_1 + \cdots + X_N$ is a random number, so we may now have doubt about using the CLT, but if the expectation $E(N)$ is large, then we can use the Central Limit Theorem.

**Note.** When the severity distribution of $X$ is discrete, use continuity correction. The continuity correction works this way: Suppose that $a$ and $b$ are two values of $S$ such that $S$ does not take any value in the interval $(a, b)$. Then the probabilities $P(S > a)$ and $P(S \geq b)$ are the same. We use $P(S > \frac{a+b}{2})$ instead of these two. On the other hand, the probabilities $P(S \leq a)$ and $P(S < b)$ are equal, but we use $P(S < \frac{a+b}{2})$ instead of them.

Here are two examples:

**Example (from the textbook).** An insurable event has a 10% probability of occurring and when it occurs results in a loss of 5,000. Market research has indicated that consumers will pay at most 550 to purchase insurance against this event. How many policies must a company sell in order to have a 95% chance of making money (ignoring expenses)?

**Solution.** Let $n$ be the number of policies sold and let $C$ be the number of claims. Then $C \approx \text{Binomial}(n; 0.1)$. Then

\[
\begin{align*}
E(C) &= nq = (0.1)n \\
\text{Var}(C) &= nq(1 - q) = (0.1)(0.9)n
\end{align*}
\]

The cost for the company will be $5000C$ and what it receives is $550n$. We want to have
5000C < 550n with 95% probability. So, we choose n so as to satisfy:

\[
0.95 \leq P(5000C < 550n) = P(C < 0.11n) = P \left( \frac{C - E(C)}{\sqrt{\text{Var}(C)}} < \frac{0.11n - 0.1n}{\sqrt{(0.1)(0.9)n}} \right)
\]

Since \( \frac{C - E(C)}{\sqrt{\text{Var}(C)}} \approx N(0, 1) \), the minimum number n must satisfy

\[
\frac{0.11n - 0.1n}{\sqrt{(0.1)(0.9)n}} = 1.96 \Rightarrow n \approx 3457.4 \Rightarrow \text{minimum } n = 3458
\]

**Note.** In these sort of problems where we search for the number n, do not do continuity correction.

**Example.** Computer maintenance costs for a department are modeled as follows:

(i) The distribution of the number of maintenance calls each machine will need in a year is Poisson with mean 3.

(ii) The cost for a maintenance call has mean 80 and standard deviation 200.

(iii) The number of maintenance calls and the costs of the maintenance calls are all mutually independent.

The department must buy a maintenance contract to cover repairs if there is at least a 10% probability that aggregate maintenance costs in a given year will exceed 120% of the expected costs.

Using the normal approximation for the distribution of the aggregate maintenance costs, calculate the minimum number of computers needed to avoid purchasing a maintenance contract.

**Solution.** Let \( N \) be the random variable representing the number of maintenance calls for a single machine. We are given that

\[
N \sim \text{Poisson}(3) \Rightarrow E(N) = \text{Var}(N) = 3
\]
Let $Y$ be the random variable representing the maintenance cost upon each call. Let $X$ be the aggregate maintenance cost for each machine; so $N$ is the primary random variable for $X$, and $Y$ is its secondary distribution. Then

$$E(X) = E(N)E(Y) = (30)(80) = 240$$

$$\text{Var}(X) = E(N)\text{Var}(Y) + \text{Var}(N)E(Y)^2 = (3)(200)^2 + (3)(80)^2 = 139,200$$

Let $n$ be the number of machines. Let $S$ be the aggregate maintenance cost. Then

$$E(S) = nE(X) = 240n \quad \text{Var}(S) = n\text{Var}(X) = 139,200n$$

The company will avoid purchasing the contracts if $P(S > 1.2E(S)) < 0.1$. Now:

$$0.1 > P(S > 1.2E(S)) = P \left( \frac{S - E(S)}{\sqrt{\text{Var}(S)}} > \frac{0.2E(S)}{\sqrt{\text{Var}(S)}} \right)$$

$$= P \left( Z > \frac{(0.2)(240n)}{\sqrt{139,200n}} \right)$$

$$= P \left( Z > \frac{48n}{\sqrt{139,200n}} \right)$$

$$\Rightarrow \frac{48n}{\sqrt{139,200n}} > \Phi(0.9) = 1.282 \quad \Rightarrow \quad n > 99.3$$

$$\Rightarrow \text{number of computers needed} \geq 100$$

**Note.** We recall that a $X \sim \text{lognormal}(\mu, \sigma^2)$ is the one with support $(0, \infty)$ and the density function

$$f(x) = \frac{1}{x\sqrt{2\pi}\sigma^2} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}$$

In fact, we have $\ln(X) \sim N(\mu, \sigma^2)$. Some people take this as the definition of a lognormal random variable: it is a r.v. whose logarithm is normal. One can easily calculate the raw moments:

$$E(X^k) = \exp(k\mu + \frac{1}{2}k^2\sigma^2)$$

Especially

$$\begin{cases}
E(X) = \exp(\mu + \frac{1}{2}\sigma^2) \\
E(X^2) = \exp(2\mu + 2\sigma^2)
\end{cases}$$
These two equalities are usually used to find $\mu$ and $\sigma^2$; see the example below.

**Example** *. You are asked to price a Workers’ Compensation policy for a large employer. The employer wants to buy a policy from your company with an aggregate limit of 150% of total expected loss. You know the distribution for aggregate claims is Lognormal. You are also provided with the following:

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Claims</td>
<td>50</td>
<td>12</td>
</tr>
<tr>
<td>Amount of Individual Loss</td>
<td>4500</td>
<td>3000</td>
</tr>
</tbody>
</table>

Calculate the probability that the aggregate loss will exceed the aggregate limit.

**Solution.**

$E(S) = E(N)E(X) = (50)(4500) = 225,000$

$\text{Var}(S) = E(N)\text{Var}(X) + \text{Var}(N)E(X)^2 = (50)(3000)^2 + (12)^2(4500) = 3,366,000,000$

Now we find the parameters of the lognormal distribution:

$\text{exp}(\mu + \frac{1}{2}\sigma^2) = E(S) = 225,000 \implies \mu + \frac{1}{2}\sigma^2 = \ln(225000) = 12.32$

$\text{exp}(2\mu + 2\sigma^2) = E(S^2) = \text{Var}(S) + E(S)^2 = 3366000000 + (225000)^2 \implies 2\mu + 2\sigma^2 = \ln\left(3366000000 + (225000)^2\right) = 24.71$

Solving these two equations gives us:

\[
\begin{align*}
\mu &= 12.29 \\
\sigma^2 &= 0.064 \implies \sigma = 0.254
\end{align*}
\]

Then:

$P(S > 1.5E(S)) = P(S > (1.5)(225000)) = P(S > 337500) = P(\ln(S) > \ln(337500)) = P\left(\frac{\ln(S) - 12.29}{0.254} > \frac{\ln(337500) - 12.29}{0.254}\right) = P(Z > 1.72) = 1 - \Phi(1.72) = 0.043$

**Example (from the Finan’s guide - page 269)** *. You own a fancy light bulb factory. Your workforce is a bit clumsy they keep dropping boxes of light bulbs. The boxes have varying numbers of light bulbs in them, and when dropped, the entire box is destroyed. You are given:
(i) Expected number of boxes dropped per month: 50
(ii) Variance of the number of boxes dropped per month: 100
(iii) Expected value per box: 200
(iv) Variance of the value per box: 400

You pay your employees a bonus if the value of light bulbs destroyed in a month is less than 8000.
Assuming independence and using the normal approximation, calculate the probability that you will pay your employees a bonus next month.

**Solution.** Let

\[ S = X_1 + \cdots + X_N \]

where \( N \) is the number of boxes destroyed in a month, and \( X_i \) is the value of the \( i \)-th box destroyed.

\[ E(S) = E(N)E(X) = (50)(200) = 10,000 \]

\[ \text{Var}(S) = E(N)\text{Var}(X) + \text{Var}(N)E(X)^2 = (50)(400) + (100)(200)^2 = 4,020,000 \]

probability of paying a bonus \( P(S < 8000) = \) \[ P \left( \frac{S - 10,000}{\sqrt{4,020,000}} < \frac{8,000 - 10,000}{\sqrt{4,020,000}} \right) = P(N(0,1) < -0.9975) = 0.16 \]

This value was found by using the norm.dist function of Excel; in fact the command

\[ \text{= norm.dist(-0.9975 , 0 , 1 , TRUE)} \]

gives the result.

**Note.** In this example, we do not know whether the distribution of \( S \) is discrete or continuous, therefore we do not do any continuity correction.
Using convolution to find the distribution of aggregate loss from section 9.3

For the collective model $S = X_1 + \cdots + X_N$ we have

$$F_S(x) = P(S \leq x) = \sum_{n=0}^{\infty} P(S \leq x \mid N = n) P(N = n) = \sum_{n=0}^{\infty} P(N = n) F_X^{*n}(x)$$

where $F_X^{*n}(x)$ is the convolution of $F_X(x)$ with itself $n$ times.

For discrete $X$ use:

$$F_X^{*n}(x) = \sum_{k=0}^{x} F_X^{*(n-1)}(x - k) f_X(k)$$

$$f_X^{*n}(x) = \sum_{k=0}^{x} f_X^{*(n-1)}(x - k) f_X(k)$$

For continuous $X$ use:

$$F_X^{*n}(x) = \sum_{k=0}^{x} F_X^{*(n-1)}(x - k) f_X(k)$$

$$f_X^{*n}(x) = \sum_{k=0}^{x} f_X^{*(n-1)}(x - k) f_X(k)$$

Question: How do we calculate the values $F_X^{*n}(x)$ and $f_X^{*n}(x)$ to be able to calculate $F_S(x)$ and $f_S(x)$.

Answer: For discrete $X$ use

$$F_X^{*n}(x) = \sum_{k=0}^{x} F_X^{*(n-1)}(x - k) f_X(k)$$

$$f_X^{*n}(x) = \sum_{k=0}^{x} f_X^{*(n-1)}(x - k) f_X(k)$$

For continuous $X$ use:
\[
\begin{align*}
F_X^n(x) &= \int_0^x F_X^{(n-1)}(x-y) f_X(y) \, dy \\
\int_0^x F_X^{(n-1)}(x-y) f_X(y) \, dy
\end{align*}
\]

Example for the discrete case (from the Finan’s guide - page 265). An insurance portfolio produces \( N \) claims, where

<table>
<thead>
<tr>
<th>( n )</th>
<th>( P(N = n) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.5</td>
</tr>
<tr>
<td>1</td>
<td>0.2</td>
</tr>
<tr>
<td>2</td>
<td>0.2</td>
</tr>
<tr>
<td>3</td>
<td>0.1</td>
</tr>
</tbody>
</table>

Individual claim amounts have the following distribution:

<table>
<thead>
<tr>
<th>( x )</th>
<th>( f_X(x) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.9</td>
</tr>
<tr>
<td>2</td>
<td>0.1</td>
</tr>
</tbody>
</table>

Individual claim amounts and \( N \) are mutually independent. Complete the following table:

<table>
<thead>
<tr>
<th>( x )</th>
<th>( f_X^0(x) )</th>
<th>( f_X^1(x) )</th>
<th>( f_X^2(x) )</th>
<th>( f_X^3(x) )</th>
<th>( f_S(x) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\( P(N = n) \)

Solution.
Example for the discrete case (from the Finan’s guide - page 266). The number of claims in a period has a geometric distribution with mean 4. The amount of each claim $X$ follows

$$P(X = x) = 0.25 ; \ x = 1, 2, 3, 4$$

The number of claims and the claim amounts are independent. $S$ is the aggregate claim amount in the period. Calculate $F_S(3)$.

Solution.

$$f_S(2) = P(N = 0)f^0(2) + P(N = 1)f^1(2) + P(N = 2)f^2(2) + P(N = 3)f^3(2) = (0.5)(0) + (0.2)(0.1) + (0.2)(0.81) + (0.1)(0) = 0.182$$

<table>
<thead>
<tr>
<th>$x$</th>
<th>$f^0_X(x)$</th>
<th>$f^1_X(x)$</th>
<th>$f^2_X(x)$</th>
<th>$f^3_X(x)$</th>
<th>$f_S(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.5</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0.9</td>
<td>0</td>
<td>0</td>
<td>0.18</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0.1</td>
<td>0.81</td>
<td>0</td>
<td>0.182</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0.18</td>
<td>0.729</td>
<td>0.0909</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0.01</td>
<td>0.162</td>
<td>0.0182</td>
</tr>
</tbody>
</table>

$P(N = n)$ | 0.5 | 0.2 | 0.2 | 0.1

The row for the probabilities $P(N = n)$ is calculated from

$$P(N = n) = \frac{\beta^n}{(1 + \beta)^{n+1}} = \frac{4^n}{5^{n+1}}$$
Now by looking at the last column we get:

\[ F_S(3) = 0.2 + 0.04 + 0.048 + 0.0576 = 0.3456 \]

**Example for the continuous case (from the Finan’s guide - page 267).** Severities have a uniform distribution on \([0, 100]\): The frequency distribution is given by

<table>
<thead>
<tr>
<th>n</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.60</td>
</tr>
<tr>
<td>1</td>
<td>0.30</td>
</tr>
<tr>
<td>2</td>
<td>0.10</td>
</tr>
</tbody>
</table>

(i) Find the 2-fold convolution of \( F_X(x) \).

(ii) Find \( F_S(x) \).

**Solution of (i).**

(just the ordinary distribution) \( F_X^1(x) = \int_0^x f(t) \, dt = \int_0^x \frac{1}{100} \, dt = \frac{x}{100} \)

\[ F_X^2(x) = \int_0^x F_X^1(x-t) f(t) \, dt = \int_0^x \frac{x-t}{100} \, dt = \frac{1}{10000} \left[ -\frac{1}{2} (x-t)^2 \right]_{t=0}^{t=x} = \frac{1}{20000} x^2 \]

**Solution of (ii).**

For \( x \geq 0 \) we have \( F_X^0(x) = 1 \) and therefore:

\[ F_S(x) = P(N = 0) F_X^0(x) + P(N = 1) F_X^1(x) + P(N = 2) F_X^2(x) \]

\[ = 0.60 + (0.30)\left( \frac{1}{100} x \right) + (0.10)\left( \frac{1}{20000} x^2 \right) = 0.60 + \frac{3x}{1000} + \frac{x^2}{20000} \]

**Example (from the Finan’s guide - page 270)**. The number of claims, \( N \), made on an insurance portfolio follows the following distribution:
<table>
<thead>
<tr>
<th>$n$</th>
<th>$P(N = n)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.7</td>
</tr>
<tr>
<td>2</td>
<td>0.2</td>
</tr>
<tr>
<td>3</td>
<td>0.1</td>
</tr>
</tbody>
</table>

If a claim occurs, the benefit is 0 or 10 with probability 0.8 and 0.2, respectively. The number of claims and the benefit for each claim are independent. Calculate the probability that aggregate benefits will exceed expected benefits by more than 2 standard deviations.

**Solution.**

\[
E(N) = (0)(0.7) + (2)(0.2) + (3)(0.1) = 0.7
\]

\[
E(N^2) = (0^2)(0.7) + (2^2)(0.2) + (3^2)(0.1) = 1.7
\]

\[
\text{Var}(N) = E(N^2) - E(N)^2 = 1.7 - (0.7)^2 = 2
\]

\[
E(X) = (0)(0.8) + (10)(0.2) = 2
\]

\[
E(X^2) = (0^2)(0.8) + (10^2)(0.2) = 20
\]

\[
\text{Var}(X) = E(X^2) - E(X)^2 = 20 - 4 = 16
\]

\[
E(S) = E(N)E(X) = (0.7)(2) = 1.4
\]

\[
\text{Var}(S) = E(N)\text{Var}(X) + \text{Var}(N)E(X)^2 = (0.7)(16) + (2)(2^2) = 16.04
\]

\[
\text{Std}(S) = \sqrt{16.04} = 4
\]

\[
P(S > 1.4 + 2(4)) = P(S > 9.4) = 1 - P(S = 0)
\]

\[
P(S = 0) = P[(S = 0) \cap (N = 0)] + P[(S = 0) \cap (N = 2)] + P[(S = 0) \cap (N = 3)]
\]
\[ P(S = 0 \mid N = 0)P(N = 0) + P(S = 0 \mid N = 2)P(N = 2) + P(S = 0 \mid N = 3)P(N = 3) \]
\[ = (1)(0.7) + (0.8)^2(0.2) + (0.8)^3(0.1) = 0.8792. \]

Then

\[ P(S > 9.4) = 1 - 0.8792 = 0.1208 \]

**Example (from the Finan’s guide - page 271)**. For a collective risk model the number of losses, \( N \), has a Poisson distribution with \( \lambda = 20 \). The common distribution of the individual losses has the following characteristics:

(i) \( E(X) = 70 \)
(ii) \( E(X \wedge 30) = 25 \)
(iii) \( P(X > 30) = 0.75 \)
(iv) \( E(X^2 \mid X > 30) = 9000 \)

An insurance covers aggregate losses subject to an ordinary deductible of 30 per loss. Calculate the variance of the aggregate payments of the insurance.

**Solution.** We have

\[ S = (X_1 - 30)_+ + \cdots + (X_N - 30)_+ \]

For any Poisson compound distribution \( S = Y_1 + \cdots + Y_N \) we have \( \text{Var}(S) = \lambda E(Y^2) \).

Especially, for the compound Poisson \( S = (X_1 - 30)_+ + \cdots + (X_N - 30)_+ \) we have

\[ \text{Var}(S) = \lambda E[(X_1 - 30)^2_+] \]

However:

\[ E[(X - 30)^2_+] = E[(X - 30)^2_+] \mid X > 30]P(X > 30) + E[(X - 30)^2_+] \mid X \leq 30]P(X \leq 30) \]
\[ = 0.75E[(X - 30)^2_+] \mid X > 30] + (0.25)(0) = 0.75 E[(X - 30)^2] \mid X > 30] \]
\[
0.75E(X^2 - 60X + 900 | X > 30) = 0.75\left\{ E(X^2 | X > 30) - 60 E(X | X > 30) + 900 \right\}
\]
\[
= 0.75\left\{ E(X^2 | X > 30) - 60 E(X - 30 | X > 30) - 1800 + 900 \right\}
\]
\[
= 0.75\left\{ E(X^2 | X > 30) - 60 \frac{E(X) - E(X \wedge 30)}{S(30)} - 900 \right\}
\]
\[
= 0.75\left\{ 9000 - 60 \frac{70 - 25}{0.75} - 900 \right\} = 3375
\]

Example (from the Finan’s guide - page 273) *. For an insurance:

(i) The number of losses per year has a Poisson distribution with \( \lambda = 10 \).

(ii) Loss amounts are uniformly distributed on (0, 10)

(iii) Loss amounts and the number of losses are mutually independent.

(iv) There is an ordinary deductible of 4 per loss.

Calculate the variance of aggregate payments in a year.

**Solution.** The secondary distribution is \( Y = (X - 4)_+ \) where \( X \) is the ground-up loss.

\[
E(Y^2) = \int_4^{10} (x - 4)^2 f(x)dx = (0.1) \int_4^{10} (x - 4)^2dx = (0.1) \left[ \frac{1}{3}(x - 4)^3 \right]_{x=0}^{x=10} = 7.2
\]

(for the compound Poisson) \( \text{Var}(S) = \lambda E(Y^2) = (10)(7.2) = 72 \)
Stop Loss Insurance  
from section 9.3

When a deductible $d$ is applied to the aggregate loss $S$, the resulting payment

$$(S - d)_+ = \begin{cases} 
0 & S \leq d \\
S - d & S > d 
\end{cases}$$

is called stop-loss insurance. The expected value $E[(S - d)_+]$ is called net stop-loss premium. As usual, we have

$$E[(S - d)_+] = \int_d^\infty [1 - F_S(x)]dx = \int_d^\infty (x - d)f_S(x)dx$$

$$\sum_{x > d}[1 - F_S(x)] = \sum_{x > d}(x - d)f_S(x)$$

**Note.** Although this equality is true in theory, but it is not so practical to use it because it is difficult to determine the support of the random variable $S$ and so we might not know the upper limit of this integration (analyse this difficulty for the next example). A practical way of using this premium is to use the equality

$$E[(S - d)_+] = E(S) - E(S \wedge d) = E(N)E(X) - E(S \wedge d)$$

**Example.** Severities have a uniform distribution on $[0, 100]$: The frequency distribution is given by

<table>
<thead>
<tr>
<th>$n$</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.60</td>
</tr>
<tr>
<td>1</td>
<td>0.30</td>
</tr>
<tr>
<td>2</td>
<td>0.10</td>
</tr>
</tbody>
</table>

Calculate $E[(S - 5)_+]$.

**Solution.**
\[ E(X) = 50 \quad E(N) = (0)(60) + (1)(0.30) + (2)(0.10) = 0.5 \]

\[ E(S) = E(N) \cdot E(X) = 25 \]

We already calculated that \( F_S(x) = 0.60 + \frac{3x}{1000} + \frac{x^2}{200000} \). By differentiating we have \( f_S(x) = \frac{3}{1000} + \frac{x}{100000} \). Therefore

\[
E(S \wedge 5) = \int_0^5 \left( \frac{3}{1000} + \frac{x}{100000} \right) dx + 5[1 - F_S(5)] = \left[ \frac{3}{1000}x + \frac{x^2}{200000} \right]_0^5 + 5[1 - F_S(5)]
\]

\[
= \frac{3.025}{200000} + 5 \left[ 1 - \left( \frac{3}{1000} + \frac{3 \cdot 5}{1000} + \frac{5^2}{200000} \right) \right] = 1.9395
\]

\[ E[(S - 5)_] = E(S) - E(S \wedge 5) = 25 - 1.9395 = 23.0605 \quad \checkmark \]

**Important Note.** When \( X \) has continuous distribution we use the technique used in this example, but when \( X \) has discrete distribution we use the recursive formula in part (i) of the following theorem; see the example following the theorem.

**Theorem.** Suppose that \( P(a < S < b) = 0 \). Then, for \( a \leq d \leq b \), we have

(i) \[ E[(S - d)_] = E[(S - a)_] - (d - a)[1 - F_S(a)] \]

(ii) \[ E[(S - d)_] = \frac{b - d}{b - a} E[(S - a)_] + \frac{d - a}{b - a} E[(S - b)_] \quad \text{linear interpolation} \]

**Proof.** \[ E[(S - d)_] = \int_d^\infty [1 - F_S(x)]dx = \int_a^\infty [1 - F_S(x)]dx - \int_a^d [1 - F_S(x)]dx \]

\[ = \int_a^\infty [1 - F_S(x)]dx - \int_a^d [1 - F_S(a)]dx \]

\[ = E[(S - a)_] - (d - a)[1 - F_S(a)] \]
So:

\[ E[(S - d)_+] = E[(S - a)_+] - (d - a)[1 - F_S(a)] \quad (\ast) \]

Writing this equality for \( d = b \) we get:

\[ E[(S - b)_+] = E[(S - a)_+] - (b - a)[1 - F_S(a)] \]

\[ \Rightarrow 1 - F_S(a) = \frac{E[(S - a)_+] - E[(S - b)_+]}{b - a} \]

Putting this into the equality (\ast)

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

\[ = \frac{b-d}{b-a} E[(S - a)_+] + \frac{d-a}{b-a} E[(S - b)_+] \quad \checkmark \]

**Theorem (not mentioned explicitly in the textbook).** If the aggregate loss \( S \) is discrete, then the function \( f(t) = E[(S - t)_+] \) as a function of \( t \) is decreasing.

**Proof.** Take the two points \( a \) and \( b \) as in the previous theorem. The equality

\[ E[(S - b)_+] = E[(S - a)_+] - (b - a)[1 - F_S(a)] \]

shows that \( E[(S - b)_+] < E[(S - a)_+] \) for \( a < b \). Furthermore, for every \( a < d < b \), the value of \( E[(S - d)_+] \) is found by taking the linear interpolation of \( E[(S - a)_+] \) and \( E[(S - b)_+] \), therefore the function \( f(t) = E[(S - t)_+] \) is decreasing over the interval \( t \in [a, b] \).

**Note.** If \( X \) is any loss random variable, then by taking \( S = X \), we notice the above theorems hold for a single loss variable too. Specially, the function \( f(t) = E[(X - t)_+] \) is decreasing.

**Example (from the Finan’s guide - page 288)**. For a collective risk model:

(i) The number of losses has a Poisson distribution with \( \lambda = 2 \).

(ii) The common distribution of the individual losses is:
<table>
<thead>
<tr>
<th>$x$</th>
<th>$f(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.6</td>
</tr>
<tr>
<td>2</td>
<td>0.4</td>
</tr>
</tbody>
</table>

An insurance covers aggregate losses subject to a deductible of 3. Calculate the expected aggregate payments of the insurance.

**Solution.**

\[ E(X) = (1)(0.6) + (2)(0.4) = 1.4 \]

\[ E(S) = E(N) E(X) = (2)(1.4) = 1.8 \]

\[
E[(S - 3)+] = E[(S - 2)+] + F_S(2) - 1 \\
= E[(S - 1)+] + F_S(2) + F_S(1) - 2 \\
= E(S) + F_S(0) + F_S(1) + F_S(2) - 3 \\
= E(S) + 3f_S(0) + 2f_S(1) + f_S(2) - 3 \\
= 1.8 + 3f_S(0) + 2f_S(1) + f_S(2) - 3 \quad (*)
\]

**Next Step:**

\[ F_S(0) = P(S = 0) = P(N = 0) = e^{-2} \]

\[ f_S(1) = P(S = 1) = P(N = 1, X = 1) = P(N = 1) P(X = 1) = \left(\frac{e^{-2}}{1!}\right)(0.6) = (1.2)e^{-2} \]

\[ f_S(2) = P(S = 2) = P(N = 1, X_2 = 2) + P(N = 2, X_1 = 1, X_2 = 1) \]

\[ = \frac{e^{-2}}{1!}(0.4) + \frac{e^{-2}}{2!}(0.6)^2 = e^{-2}(0.8) + 2e^{-2}(0.6)^2 \]

By putting these values into \((*)\), we get:

\[ = 1.8 + 3e^{-2} + \left\{2e^{-2}(1.2)\right\} + \left\{e^{-2}(0.8) + 2e^{-2}(0.6)^2\right\} - 3 \]
Example. The number of claims in a period has a geometric distribution with mean 4. The amount of each claim \( X \) follows

\[
P(X = x) = 0.25 ; \ x = 1, 2, 3, 4
\]

The number of claims and the claim amounts are independent. \( S \) is the aggregate claim amount in the period. Calculate \( E[(S - 2)_] \) and \( E[(S - 1.6)_] \).

Solution. We had:

\[
\begin{array}{cccccc}
 x & f_X^0(x) & f_X^1(x) & f_X^2(x) & f_X^3(x) & f_S(x) \\
 0 & 1 & 0 & 0 & 0 & 0.2 \\
 1 & 0 & 0.25 & 0 & 0 & 0.04 \\
 2 & 0 & 0.25 & 0.0625 & 0 & 0.048 \\
 3 & 0 & 0.25 & 0.125 & 0.0156 & 0.0576 \\
\end{array}
\]

\[
P(N = n) = 0.2 \quad 0.16 \quad 0.128 \quad 0.1024
\]

\[
E(X) = (0.25)(1) + (0.25)(2) + (0.25)(3) + (0.25)(4) = 2.5
\]

\[
E(S) = E(N) E(X) = (4)(2.5) = 10
\]

\[
F_S(0) = f_S(0) = 0.2
\]

\[
F_S(1) = f_S(0) + f_S(1) = 0.24
\]

\[
E[(S - 1)_+] = E[(S - 0)_+] - [1 - F_S(0)] = E(S) - [1 - F_S(0)] = 10 - [1 - 0.2] = 9.2
\]

\[
E[(S - 2)_+] = E[(S - 1)_+] - [1 - F_S(1)] = E(S) - [1 - F_S(1)] = 9.2 - [1 - 0.24] = 8.44 \quad \checkmark
\]

\[
E[(S - 1.6)_+] = (0.4)E[(S - 1)_+] + (0.6)E[(S - 2)_+] = (0.4)(9.2) + (0.6)(8.44) = 8.744 \quad \checkmark
\]
Note. When $S$ is discrete and its support consists of jumps of step $h$, the following theorem gives a formula for $E[(S - a)_+]$ for all $a$ being in the support, and then using the above interpolation formula one can find all other values of $E[(S - a)_+]$ for $a$ not being in the support.

**Theorem.** If the support of $S$ consists of the values $\{0, h, 2h, 3h, \ldots\}$ then for every $jh$ in the support set we have

$$E[(S - a)_+] = h \sum_{n=0}^{\infty} \left\{ 1 - F_S((n + j)h) \right\} = h \times \text{sum of all survival values starting from } jh$$

Note. This theorem is not of practical use and it might be used for theoretical discussions.
Panjer Recursive Formulas

section 9.6

Calculating the convolutions to find $f_S(x)$ or $F_S(x)$ is a difficult matter. The Panjer’s recursive formula is an efficient way of calculating the density of $S$.

**Theorem.** Consider a collective risk model in which the severity random variable has support in the set of non-negative integer values

$$X \in \{0, 1, 2, \ldots\}$$

and the frequency distribution is in $(a, b, 1)$

$$p_n = \left(a + \frac{b}{n}\right)p_{n-1} \quad n = 2, 3, \ldots$$

Then the density of the aggregate loss $S \in \{0, 1, 2, \ldots\}$ satisfies the recursive formula:

$$f_S(n) = \frac{\{p_1 - (a + b)p_0\} f_X(n) + \sum_{k=1}^{n} \left(a + \frac{b}{n} k\right) f_X(k) f_S(n - k)}{1 - a f_X(0)}$$

**Theorem.** If in the above theorem the frequency distribution is in $(a, b, 0)$, then we have $p_1 - (a + b)p_0 = 0$, and therefore the recursive formula reduces to

$$f_S(n) = \frac{\sum_{k=1}^{n} \left(a + \frac{b}{n} k\right) f_X(k) f_S(n - k)}{1 - a f_X(0)}$$

**Note.** If the frequency distribution is Poisson, then we have $a = 0$ and $b = \lambda$ and it is in $(a, b, 0)$, therefore for this frequency distribution we have

$$f_S(n) = \frac{\lambda}{n} \sum_{k=1}^{n} k f_X(k) f_S(n - k)$$

**Note.** If in the above two theorems we have $X \in \{0, 1, \ldots, m\}$, then the corresponding formulas reduce to

$$f_S(n) = \frac{\{p_1 - (a + b)p_0\} f_X(n) + \sum_{k=1}^{\min\{n, m\}} \left(a + \frac{b}{n} k\right) f_X(k) f_S(n - k)}{1 - a f_X(0)}$$
where in the case of \((a, b, 0)\) we have \(p_1 - (a + b)p_0 = 0\).

**Example** *. You are given:

(i) Aggregate claims has a compound Poisson distribution with \(\lambda = 0.8\).

(ii) Individual claim amount distribution is

<table>
<thead>
<tr>
<th>(x)</th>
<th>(P(X = x))</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.5</td>
</tr>
<tr>
<td>2</td>
<td>0.3</td>
</tr>
<tr>
<td>3</td>
<td>0.2</td>
</tr>
</tbody>
</table>

(iii) The probabilities for certain values of the aggregate claims, \(S\), are:

<table>
<thead>
<tr>
<th>(x)</th>
<th>(P(S = x))</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.1438</td>
</tr>
<tr>
<td>3</td>
<td>0.1198</td>
</tr>
<tr>
<td>5</td>
<td>0.0294</td>
</tr>
</tbody>
</table>

Determine \(P(S = 4)\).

**Solution.** We use the recursive formula although it can be done directly too.

\[
f_S(n) = \frac{\lambda}{n} \sum_{k=1}^{n} k f_X(k) f_S(n - k) = \frac{0.8}{n} \sum_{k=1}^{n} k f_X(k) f_S(n - k)
\]

\[P(S = 0) = P(N = 0) = e^{-0.8} = 0.4493\]

\[P(S = 1) = (0.8)f_X(1)f_S(0) = (0.8)(0.5)e^{-0.8} = 0.1797\]

\[P(S = 4) = \left(\frac{0.8}{4}\right) \left\{ f_X(1)f_S(3) + 2f_X(2)f_S(2) + 3f_X(3)f_S(1) + 4f_X(4)f_S(0) \right\} =
\]

\[\left(0.2\right) \left\{ (0.5)(0.1438) + (2)(0.3)(0.1198) + (3)(0.2)(0.0294) \right\} = 0.0508\]

**Example (Problem 9.47 of the textbook)** *. Aggregate claims are compound Poisson

with \(\lambda = 2\), \(f_X(1) = \frac{1}{4}\), and \(f_X(2) = \frac{2}{4}\). For a premium of 6, an insurer covers aggregate

claims and agrees to pay a dividend (a refund of premium) equal to the excess, if any, of 75%
of the premium over 100% of the claims. Determine the excess of premium over expected claims and dividends. Use the recursive formulas.

**Solution.** 75% of the premium is \((0.75)(6) = 4.5\), so the dividend is

\[
D = \begin{cases} 
4.5 - S & \text{if } S \leq 4.5 \\
0 & \text{if } S > 4.5
\end{cases}
\]

(this is for the “if any” part)

So

\[D \in \{4.5, 3.5, 2.5, 1.5, 0.5\}\]

We are asked to find \(6 - E(S) - E(D)\).

\[
f_S(n) = \frac{\lambda}{n} \sum_{k=1}^{n} k f_X(k) f_S(n-k) = \frac{2}{n} \sum_{k=1}^{n} k f_X(k) f_S(n-k)
\]

\[f_S(0) = P(N = 0) = e^{-2}
\]

\[f_S(1) = (2)(1)(\frac{1}{4})e^{-2} = \frac{1}{2} e^{-2}
\]

\[f_S(2) = \frac{2}{3} \left\{ (\frac{1}{4})(\frac{1}{2} e^{-2}) + 2(\frac{3}{4})(e^{-2}) \right\} = \frac{13}{8} e^{-2}
\]

\[f_S(3) = \frac{2}{3} \left\{ (\frac{1}{4})(\frac{13}{8} e^{-2}) + 2(\frac{3}{4})(\frac{1}{2} e^{-2}) \right\} = \frac{37}{32} e^{-2}
\]

\[f_S(4) = \frac{2}{4} \left\{ (\frac{1}{4})(\frac{505}{512} e^{-2}) + 2(\frac{3}{4})(\frac{13}{8} e^{-2}) \right\} = \frac{505}{384} e^{-2}
\]

\[E(D) = (4.5)(e^{-2}) + (3.5)(\frac{1}{2} e^{-2}) + (2.5)(\frac{13}{8} e^{-2}) + (1.5)(\frac{37}{32} e^{-2}) + (0.5)(\frac{505}{384} e^{-2}) = 1.6411
\]

\[E(X) = (1)(\frac{1}{4}) + (2)(\frac{3}{4}) = \frac{7}{4} \quad \Rightarrow \quad E(S) = E(N) E(X) = (2)(\frac{7}{4}) = \frac{7}{2}
\]

Answer = \(6 - E(S) - E(D) = 6 - 3.5 - 1.6411 = 0.8589\)