Chapter LM 3-7: Loss Models

Functions and moments:
- The cumulative distribution function of a random variable X is the probability \( F(X) = Pr(X \leq x) \)
- \( S(X) \) is the survival function, the complement of \( F(X) \), the probability of surviving longer than x, \( Pr(X > x) \)
- For continuous random variable, \( f(x) \) is the probability density function, \( f(x) = \frac{d}{dx}F(X) \)
- For discrete random variable, \( p(x) \) is the probability mass function, \( p(x) = Pr(X = x) \)
- \( h(x) \) is the hazard rate function, \( h(x) = \frac{f(x)}{S(x)} \)
- The expected value of X is defined by, \( E[X] = \int_{-\infty}^{\infty} x f(x) \, dx \), and more generally, \( E[g(x)] = \int_{-\infty}^{\infty} g(x) f(x) \, dx \)

\[ E(aX + bY) = aE(X) + bE(Y), \] regardless of whether \( X \) and \( Y \) are independent or not. \[ E[(X + Y)^2] = E(X^2) + 2E(XY) + E(Y^2). \]

- The \( n^{th} \) raw moment of X is defined as \( \mu'_n = E[X^n] \). \( \mu = \mu'_1 \) is the mean. The \( n^{th} \) central moment of X (\( n \neq 1 \)) is defined as \( \mu_n = E[(X - \mu_1)^n] \).

- Central moments can be calculated from the raw moments:
  \[ \begin{align*}
  \mu_2 &= \mu'_2 - \mu^2 \\
  \mu_3 &= \mu'_3 - 3\mu'_2\mu + 2\mu^3 \\
  \mu_4 &= \mu'_4 - 4\mu'_3\mu + 6\mu'_2\mu^2 - 3\mu^4
  \end{align*} \]

Measures
- The variance is \( \text{Var}(X) = \mu_2 = E(X^2) - E(X)^2 \), and is denoted by \( \sigma^2 \)

\[ \text{Var}(aX + bY) = a^2\text{Var}(X) + b^2\text{Var}(Y) + 2ab \text{Cov}(X,Y) \]

If all the \( X_i \)'s are independent and have identical distributions, and we set \( X = X_i \) for all i, then \( \text{Var}(\sum_{i=1}^{n} X_i) = n \text{Var}(X) \)

However, \( \text{Var}(nX) = n^2\text{Var}(X) \)

- The standard deviation \( \sigma \) is the positive square root of the variance
- The coefficient of variation is \( \frac{\sigma}{\mu} \), relative dispersion
- The skewness is \( \gamma_1 = \frac{\mu_3}{\sigma^3} \)
- The kurtosis is \( \gamma_2 = \frac{\mu_4}{\sigma^4} \), thickness of tails
- The covariance is defined by:
  \[ \text{Cov}(X,Y) = E[(X - \mu_x)(Y - \mu_y)] = E[XY] - E[X]E[Y] \]

\[ \text{for independent random variables, Cov}(X,Y) = 0 \]

- The correlation coefficient is defined by:
  \[ \rho_{XY} = \frac{\text{Cov}(X,Y)}{\sigma_x \sigma_y} \]

Sample Mean and Variance
A sample is a set of observations from \( n \) identically distributed random variables.

The sample mean, \( \bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} = E[X] \)

The variance of the sample mean of \( X_1,...,X_n \),
\[ \text{Var}(\bar{X}) = \frac{\sum_{i=1}^{n} \text{Var}(X)}{n^2} = \frac{\text{Var}(X)}{n} = \frac{\sigma^2}{n} \]

Percentile
A 100p\text{th} percentile is a number \( \pi_p \) such that \( F(\pi_p) = Pr(X \leq \pi_p \geq p \) and \( F(\pi_p^-) = Pr(X < \pi_p) \leq p \). If \( F \) is strictly increasing, it is the unique point at which \( Pr(X \leq \pi_p) = p \). A median is a 50\text{th} percentile.

Normal approximation
Le \( X \) be a normal random variable with mean \( \mu \) and variance \( \sigma^2 \), \( X \sim n(\mu, \sigma^2) \).

Suppose we want a number \( x \) such that \( Pr(X \leq x) = p \), where \( p \) is the percentile:

\[ Pr(X \leq x) = p \]

\[ Pr\left(\frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma}\right) = p \]

\[ \Phi\left(\frac{x - \mu}{\sigma}\right) = p \]

\[ \frac{x - \mu}{\sigma} = \Phi^{-1}(p) \]

\[ x = \mu + \sigma \Phi^{-1}(p) \]

Note: \( \Phi^{-1}(p) \) can be looked up on the normal distribution table (i.e., \( \Phi^{-1}(0.95) = 1.645 \))

Mixtures
A finite mixture distribution is a random variable \( X \) whose distribution function \( F(X) \) is given by a discrete sum of probability measures:

\[ F(X) = \sum_{i=1}^{n} w_i F_i(X), \quad \text{weights } w_i \text{ adds up to 1} \]

A mixture distribution is an appropriate model for a situation where a single event occurs. However, the single event may be of many different types, and the type is random.

The mean of a mixture is the weighted average of the means of the components. The variance of a mixture is not the weighted average of the variances of its components. The variance is computed by the second moment and then subtracting the square of the mean. It is also possible to have continuous mixtures.
Bernoulli shortcut
A Bernoulli distribution is one where the random variable is either 0 or 1. It is 1 with probability q and 0 with probability 1 − q. Its mean is q, and its variance is q(1 − q).

If X is Bernoulli and Y can only assume the values a and b, with a having probability q, then Y = (a − b)X + b. This means that the variance of Y is \((a − b)^2\) Var(X) = \((a − b)^2q(1 − q)\).

**For any random variable which assumes only 2 values, the variance is the squared difference between the two values times the probabilities of the two values (as highlighted above).**

Payment per loss with deductible
An ordinary deductible d means that the first d of each claim is not paid. The random variable for the payment per claim is denoted (X − d),

\[
E[(X − d)_+] = \int_0^\infty (x − d)f(x)dx = \int_d^\infty S(x)dx
\]

The random variable for the amount not paid due to the deductible is the minimum of X and d, and is denoted by \(X \wedge d\).

\[
E[X \wedge d] = \int_0^d xf(x)dx + dS(x) = \int_0^d S(x)dx
\]

The expected value of the loss is \(E[X]\),

\[
E[X] = E[X \wedge d] + E[(X − d)_+]
\]

Payment per payment with deductible
The random variable for payment per payment on an insurance with a deductible is the deductible is the payment per loss random variable conditioned on \(x > d\), or \(X − d \mid X > d\). Its expected value is,

\[
e(d) = \frac{E[(X − d)_+]}{S(d)}
\]

This variable is shifted by d and truncated. \(e(d)\) is also known as mean residual life and mean excess life.

Special cases:
1. For an exponential distribution, \(e(d) = 0\)
2. For a Pareto distribution (2 parameters),
\[
e(d) = \frac{\theta + d}{\alpha − 1}
\]
3. For a single parameter Pareto,
\[
e(d) = \frac{d}{\alpha − 1}
\]

Therefore, the expected value of the loss is also,

\[
E[X] = E[X \wedge d] + e(d)S(d)
\]

Payment per loss with claims limit
For an insurance coverage with a claims limit u, the expected payment per loss is \(E[X \wedge u]\). If the insurance coverage has a claim limit u and a deductible d, then the expected payment per loss is \(E[X \wedge u] − E[X \wedge d]\). For the expected payment per payment, divide this \(S(d)\).

Parametric distribution
All of the continuous distributions are scale families. This means that if the random variable is multiplied by a constant, the new random variable is in the same family. \(\theta\) is the *scale parameter*. This means that if \(X\) has the Whatever distribution with parameters \(\theta\) and OtherParameters, then \(cX\) has the Whatever distribution with parameters \(c\theta\) and the same OtherParameters. This makes handling inflation easy.

\[
F_y(y) = Pr(Y \leq y) = Pr(cX \leq y) = Pr(X \leq \frac{y}{c}) = F_x\left(\frac{y}{c}\right)
\]

Tail Weight
1. The more positive raw or central moments exist, the less the tail weight.
2. To compare two distributions, the limits of the ratios of the survival functions, or equivalently the ratios of the density functions, can be examined as \(x \to \infty\). A ratio going to infinity implies the function in the numerator has heavier tail weight.
3. An increasing hazard rate function means a lighter tail and a decreasing one means a heavier tail.
4. An increasing mean residual life function means a heavier tail and vice versa. A decreasing hazard rate function implies an increasing mean residual life and an increasing hazard rate function implies a decreasing mean residual life, but the converse isn’t necessarily true.

Splicing
*For a spliced distribution, the sum of the probabilities of being in each splice must add up to 1.*

Deductibles
An *ordinary deductible d* on a policy means that d is subtracted from each loss to obtain the payment, and nothing is paid if the resulting number is not positive.

A *franchise deductible d* means that if the loss is below d, nothing is paid, but if the loss is d or higher, the full amount is paid.

- **Payment per loss for a franchise deductible:**
  \[
  E[(X − d)_+] + d S(d)
  \]
- **Payment per payment for a franchise deductible:**
  \[
  e(d) + d
  \]

**Loss Elimination Ratio (LER)**
The LER is defined as the proportion of the loss which the insurer doesn’t pay as a result of the deductible.

\[
LER(d) = \frac{E[X \wedge d]}{E[X]}
\]

Inflation decreases the loss elimination ratio.
Inflation

Uniform inflation multiplies all claims by a constant. In the presence of a deductible, it raises the average cost per claim. For scale-parameterized distributions, the inflated variable can be derived from the original variable by multiplying the scale parameter \( \theta \) by 1 plus the inflation rate. For the lognormal distribution, the inflated variable is obtained by adding \( \ln(1 + r) \) to the original parameter \( \mu \), where \( r \) is the inflation rate.

If \( Y = (1 + r)X \), where \( X \) is the original variable, \( Y \) the inflated variable, and \( r \) the inflation rate, then the expected value \( E[Y \wedge d] \) can be calculated as follows:

\[
E[Y \wedge d] = (1 + r)E[X \wedge \frac{d}{1+r}]
\]

Coinsurance

Coinsurance of \( a \) means that a portion, \( a \), of each loss is reimbursed by insurance. The expected payment per loss if there is \( a \) coinsurance, \( d \) deductible, and \( u \) maximum covered loss is

\[
E[X \wedge u] = E[X \wedge d]
\]

If there is inflation of \( r \),

\[
E[X \wedge d] = a \alpha [E(X \wedge u) - E(X \wedge d)]
\]

Bonuses

Express the bonus in terms of the earned premium; it’ll always be something like \( \max[0, c(rp - X)] \), where \( r \) is the loss ratio, \( P \) is earned premium, \( X \) is losses, and \( c \) is some constant. Then you can pull out \( crP \) and write it as \( crP - c \min(rp, X) = crP - c(X \wedge rP) \). We can then use the tables to calculate the expected value. Pareto with \( \alpha = 2 \) is used so often for this type of problem, and the formula for this distribution is:

\[
E[X \wedge d] = \frac{\theta}{d + \theta}
\]

Discrete distribution

The \((a, b, 0)\) class:

- A sum of \( n \) Poisson random variables \( N_1, \ldots, N_n \) with parameters \( \lambda_1, \ldots, \lambda_n \) has a Poisson distribution whose parameter is \( \sum_{i=1}^{n} \lambda_i \).
- For the negative binomial distribution, the variance is always greater than the mean.
- A sum of \( n \) negative binomial random variables \( N_1, \ldots, N_n \) with parameters \( r_1, \ldots, r_n \) has a negative binomial distribution with parameters \( \beta \) and \( \sum_{i=1}^{n} r_i \).
- A special case of the negative binomial distribution is the geometric distribution, which is a negative binomial distribution with \( r = 1 \).
- The geometric distribution is the discrete counterpart of the exponential distribution, and has a similar memoryless property.

\[
\Pr(N \geq n + k | N \geq n) = \Pr(N \geq k)
\]

- For the binomial distribution, the variance is always less than the mean.
- A sum of \( n \) binomial random variables \( N_1, \ldots, N_n \) having the same \( q \) and parameters \( m_1, \ldots, m_n \) has a binomial distribution with parameters \( q \) and \( \sum_{i=1}^{n} m_i \).

The distributions in the \((a, b, 0)\) class is defined by the following property. If we let \( p_n = Pr(X = k) \), then

\[
k \left( \frac{p_k}{p_{k-1}} \right) = ak + b \quad \text{where} \quad a \text{ is the slope}
\]

If the variance is greater than the mean or the slope is increasing \( \rightarrow \) negative binomial

If the variance is equal to the mean or the slope is zero \( \rightarrow \) Poisson

If the variance is less than the mean or the slope is decreasing \( \rightarrow \) binomial

The \((a, b, 1)\) class:

One way to obtain a distribution in this class is to take a distribution from the \((a, b, 0)\) class and truncate it at 0. Let \( p_n \) be the probabilities from the \((a, b, 0)\) class distribution we start with and \( p_n^T \) be the probabilities of the new distribution. Then the probabilities of the new distribution are defined by the following equation.

\[
p_n^T = \frac{p_n}{1 - p_o}
\]

These distributions are called zero-truncated distributions.

Poison/Gamma Mixture

The negative binomial can be derived as a gamma mixture of Poissons.

If loss frequency follows a Poisson distribution with parameter \( \lambda \), but \( \lambda \) is not fixed and varies according to a gamma distribution \( (\alpha, \theta) \) the conditional loss frequency is Poisson with parameter \( \lambda \); and, the unconditional loss frequency for is negative binomial \((r, \beta)\), where \( r = \alpha, \beta = 0 \).

For a gamma distribution \( (\alpha, \theta) \), the mean is \( \alpha\theta \) and the variance is \( \alpha\theta^2 \); for a negative binomial distribution, the mean is \( r\beta \) and the variance is \( r\beta(1 + \beta) \).

If a gamma mixture of Poissons, i.e., \( \lambda \) varies according to gamma distribution \( (\alpha, \theta) \), is adjusted to a fractional interval \( 1/x \rightarrow \) the mixture has mean \( \frac{a\theta}{x} \) and variance \( \frac{a\theta^2}{x^2} \).
Frequency Distributions, Exposure and Coverage Modifications

1. Exposure Modification *(Not included on May 2007 syllabus)*
   - Suppose a model is based on \( n_1 \) exposures, and now we want a model for \( n_2 \) exposures
     - i. If the model is *Poisson* with parameter \( \lambda \) → the new model is Poisson with parameter \( \lambda \left( \frac{n_2}{n_1} \right) \)
     - ii. If the model is *negative binomial* \((r, \beta)\) → the new model is negative binomial \((r \left( \frac{n_2}{n_1} \right), \beta)\)
     - iii. If the model is *binomial* \((m, q)\) → the new model is binomial \((m \left( \frac{n_2}{n_1} \right) = m^*, q)\), where the only acceptable revised \( m^* \) is still an integer

2. Coverage Modification
   - Change in deductible, so that the number of claims for amounts > 0 changes
   - Uniform inflation, which affects frequency if there is a deductible or a claim limit
   - Modified frequency, frequency of positive paid claims, has the same distribution as the original frequency but with different parameters. Suppose the probability of paying a claim, i.e., severity being > deductible, is \( v \)
     - i. If the model for loss frequency is *Poisson* with parameter \( \lambda \) → the new parameter for frequency of paid claims is \( v \lambda \)
     - ii. If the model is *negative binomial* \((r, \beta)\) → the modified frequency is negative binomial \((r, v \beta)\)
     - iii. If the model is *binomial* \((m, q)\) → the modified frequency is binomial \((m, v q)\)

3. Summary

<table>
<thead>
<tr>
<th>Model</th>
<th>Original Parameters Exposure ( n_{11}, \Pr(X&gt;0)=1 )</th>
<th>Exposure Modification Exposure ( n_{22}, \Pr(X&gt;0)=1 )</th>
<th>Coverage Modification Exposure ( n_{11}, \Pr(X&gt;0)=v )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poisson</td>
<td>( \lambda )</td>
<td>( \left( \frac{n_2}{n_1} \right) \lambda )</td>
<td>( v \lambda )</td>
</tr>
<tr>
<td>Binomial</td>
<td>( m, q )</td>
<td>( \left( \frac{n_2}{n_1} \right) m, q )</td>
<td>( m, v q )</td>
</tr>
<tr>
<td>Negative Binomial</td>
<td>( r, \beta )</td>
<td>( \left( \frac{n_2}{n_1} \right) r, \beta )</td>
<td>( r, v \beta )</td>
</tr>
</tbody>
</table>

- The same parameter that gets multiplied by \( v \) in the \((a, b, 0)\) class gets multiplied by \( v \) in the \((a, b, 1)\) class. \( p_0 = 1 - \sum p_k \) is the balancing item, \( 1 \leq k \leq \infty \)
Aggregate Loss Models: Approximating Distribution

1. Aggregate losses are the total losses paid. Let $N$ be the frequency random variable, i.e., the number of losses. Let $X_i, i = 1, \ldots, n$ be the severity random variable, i.e., the sizes of the $N$ losses. If $S$ is the aggregate loss variable $\rightarrow S = \sum_{i=1}^{N} X_i$.
   - If the $X_i$ are independent and have the same distribution, $S$ is called a compound distribution. $N$ is called the primary distribution, and the common distribution for the $X_i$ is called the secondary distribution.

2. Collective Risk Model
   - $S = \sum_{i=1}^{N} X_i$, $N$ is the number of claims and $X_i$ is the size of each claim
   - $X_i$'s are independent identically distributed random variables, i.e., every claim size has the same probability distribution and is independent of any other claim size
   - $X_i$'s are independent of $N$, i.e., the claim counts are independent of the claim sizes

3. Individual Risk Model
   - $S = \sum_{i=1}^{n} X_i$, $n$ is the number of insured in the group and $X_i$ is the aggregate claims of each individual member
   - $X_i$'s are independent but not necessarily identically distributed random variables, different insureds could have different distributions of aggregate losses
   - There is no random variable $N$, where $n$ is fixed number and equals the size of the group

4. Compound Expectation and Variance
   - Assume a collective risk model, compound distribution, then
     i. $E[S] = E[N] E[X]$, and
     ii. $\text{Var}(S) = E[N] \text{Var}(X) + \text{Var}(N) E[X]^2$. This is the special case of the conditional variance formula
     $\text{Var}(Y) = \text{Var}_C[\text{E}(Y|C)] + \text{E}_C[\text{Var}(Y|C)]$, where the condition $C$ is that $N = n$.
     iii. For a compound Poisson distribution, i.e., when the primary distribution is Poisson with parameter $\lambda$, the compound variance formula reduces to $\text{Var}(S) = \lambda E[X^2]$
     iv. The compound variance formula can only be used when $N$ and $X_i$ are independent and when $E[X]$ is constant. It cannot be used to calculate the unconditional variance of $S$ from the unconditional moment $s$ of $N$ and $X$ when $E[X]$ varies, even when $N$ and $X$ are unconditionally independent (i.e., when $E[X]$ varies, the conditional variance formula must be used).

5. Approximating Distributions
   - The aggregate distribution may be approximated with a normal distribution. If severity is discrete, then the aggregate loss distribution is discrete, and a continuity correction is required (i.e., adding 0.5 to $S$ when calculating the probability of being greater than or subtracting 0.5 from $S$ when calculating the probability of being less than).
     i. $\text{Pr}(S > s) = 1 - \Phi \left( \frac{s - \mu}{\sigma} \right)$
     ii. $\text{Pr}(S \leq s) = \Phi \left( \frac{s - \mu}{\sigma} \right)$
   - Approximation is frequently used for the collective risk model, where $E[S] = E[N] E[X]$, and $\text{Var}(S) = E[N] \text{Var}(X) + \text{Var}(N) E[X]^2$
   - When the sample is not large enough and has a heavy tail, the symmetric normal distribution is not appropriate, and sometimes lognormal distribution is used instead
   - Approximation can also be used for the individual risk model. In this case, the mean is the sum of the means (i.e., $E[S] = \sum \mu_i$, where $0 \leq i \leq n$) and the variance is the sum of the variances (i.e., $\text{Var}(S) = \sum \sigma_i^2$, where $0 \leq i \leq n$)

**Note:** In general with gamma distribution, $\Gamma(x) = (x - 1) \Gamma(x - 1) = (x - 1) (x - 2) \Gamma(x - 2) \ldots$ etc.
Aggregate Loss Models: The Recursive Formula

1. By assuming that the distribution of the \( X_i \)'s is discrete, then, in principle, it is possible to calculate the probabilities of all the possible value of \( S \), and allows the computation of the distribution function.

2. **DEFINITION: Probability Function Notation:** For,
   - Frequency distribution, \( p_n = \Pr(N = n) \)
   - Severity distribution, \( f_n = \Pr(X = n) \)
   - Aggregate loss distribution, \( g_n = \Pr(S = n) = f_i(n) \)

3. By the definitions above, \( g_n = \sum_{k=0}^{\infty} p_k \sum_{i=1}^{\infty} \ldots \sum_{i_k} f_{i_1} f_{i_2} \ldots f_{i_k} \) is the convolution formula, the product of the \( f_i \)'s is called the \( k \)-fold convolution of the \( f_i \)'s, or \( f^k \). If \( f_0 \neq 0 \), this is an infinite sum, since any number of 0s can be included in the second sum
   - \( F_s(n) = g_1 + g_2 + \ldots + g_n \).

4. Recursive formula is sometimes more efficient than the convolution formula and it automatically takes care of the probability that \( X = 0 \)
   - For the \((a, b, 0)\) class, \( g_k = 1/(1 - \alpha f_0) \sum_{j=1}^{\infty} \left( a + \frac{b_j}{k} \right) f_j g_{k-j} \), \( k = 1, 2, 3, \ldots \)
     i. For a Poisson distribution, where \( a = 0 \) and \( b = \lambda \), the formula simplifies to \( g_k = \frac{\lambda}{k} \sum_{j=1}^{\infty} f_j g_{k-j} \), \( k = 1, 2, 3, \ldots \)
   - For the \((a, b, 1)\) class, \( g_k = \{[p_1 - (a + b)p_0] f_k + \sum_{j=1}^{\infty} \left( a + \frac{b_j}{k} \right) f_j g_{k-j} / (1 - \alpha f_0) \}, k = 1, 2, 3, \ldots \)
   - To start the recursion, \( g_0 \) is needed, and if \( \Pr(X = 0) = 0 \), this is \( p_0 \). In general, \( g_0 = P_1(f_0) \), where \( P_1(z) \) is the probability generating function of the primary distribution. Formulas for \( P_1(z) \) for the \((a, b, 0)\) and \((a, b, 1)\) classes are included in the exam table.

**Note:** When computing the probability of aggregate claims, probability of 0 would need to be eliminated; i.e., if the distribution is binomial \((m, q) \rightarrow \) the revised parameters are \((m, [1 - \Pr(X=0)]q)\) and the revised severity distribution \( f_1 = \Pr(X=1) / [1 - \Pr(X=0)], f_2 = \Pr(X=2) / [1 - \Pr(X=0)] \) ... etc.

Aggregate Losses: Aggregate Deductible

1. An aggregate deductible is a deductible that is applied to aggregate losses rather than individual losses
2. Let \( N \sim \) loss frequency distribution, \( X \sim \) loss severity distribution, and \( S \sim \) aggregate loss distribution
   - In the absence of an aggregate deductible, if \( (N \cap X) = 0 \) (i.e., independent), then \( \E[S] = \E[N] \E[X] \)
   - The expected value of aggregate losses above the deductible is called the net stop-loss premium. If \( S \) is aggregate losses and \( d \) is the deductible, then the net stop-loss premium = \( \E[(S - d)_+] \).
     i. If severity is discrete, then aggregate loss will be discrete since frequency is discrete and \( \E[(S - d)_+] = \E[S] - \E[S \wedge d] \), where \( \E[S] = \E[N] \E[X] \) and \( \E[S \wedge d] \) is a finite integral

Continuous case: \( \E[(S - d)_+] = \E[S] - d + \int_0^d (d - s) f(s)_d s \)
Discrete case: \( \E[(S - d)_+] = \E[S] - d + \sum_{s=0}^{d-1} (d - s) p(s) \)

If \( d \) is not at a breakpoint in the data, and \( a < d < b \), then \( \E[(S - d)_+] = \frac{b-d}{b-a} \E[(S - a)_+] + \frac{d-a}{b-a} \E[(S - b)_+] \), by linear interpolation
   - Same as earlier, let \( p_n = \Pr(N = n) \), \( f_n = \Pr(X = n) \) and \( g_n = \Pr(S = n) \)
     i. For discrete distribution in which the only possible values are multiples of \( h \):
        \[ \E[S \wedge d] = \sum_{j=0}^{[d/h]-1} h_j g_{hj} + d[1 - \Pr(d)] \], where the sum is over all multiple of \( h \) less than \( d \)
     ii. Example: \( \E[S \wedge d] = g_i(X_i) + \ldots + g_{ih}(X_{ih}) + d \Pr(S \leq d) \), where \( S(d) = \Pr(S \geq d) = 1 - (g_i + \ldots + g_{ih}) = 1 - \Pr(S \leq d) = 1 - \Pr(d) \), \( g_i = \Pr(S = i) = p_i f_i \) and \( i \leq i + j \leq d \)
Aggregate Losses: Miscellaneous Topics (Unlikely to show up on exam)

1. If frequency has a geometric distribution mean $\beta$ and severities are exponential mean $\theta$, the aggregate distribution is a 2-point mixture of degenerate distribution at 0 with weight $\frac{1}{1+\beta}$ and an exponential distribution with mean $\theta(1 + \beta)$, weight $\frac{\beta}{1+\beta}$.
   - $Pr(S = 0) = 1 - \frac{1}{1+\beta}$, since $S$ is 0 only when $N = 0$ and the probability that a geometric is 0 is $\frac{1}{1+\beta}$.

2. Method for discretizing the distribution – First, a span will be decided, span is the distance between the points that will have a positive probability in the discretized distribution.
   - Method of Rounding – The severity values within a span are rounded to the endpoints. If $h$ is the span, $p_{kh} = F[(k + 0.5 - 0)h] - F[(k - 0.5 - 0)h]$, where $0$ indicates that the lower bound is included but the upper bound is not; as usual in rounding 0.5 rounds up. This rounding convention makes no difference if $F$ is continuous everywhere.

Ruin Theory

1. Definitions
   - The probability of ruin starting with initial surplus $u$, $\psi(u)$, as the probability surplus ever goes below 0. $\tilde{\psi}(u)$ is defined as the probability of surplus goes to 0 when surplus is only checked at integer times $t = 0, 1, 2, \ldots$; such a surplus process is called a discrete surplus process, whereas one where surplus is checked at all $\mathbb{R}$ values of $t$ is a continuous surplus process. The corresponding variable $\psi(u, t)$ and $\tilde{\psi}(u, t)$ are defined as the probability surplus goes below 0 no later than time $t$.
   - $\phi(u)$ is defined as the probability of never getting ruined if starting with surplus $u$; it is $1 - \psi(u)$. Corresponding definition apply for $\tilde{\phi}(u)$, $\phi(u, t)$, and $\tilde{\phi}(u, t)$.

2. The infinite-horizon probability of ruin is higher than the finite-horizon probability of ruin, since the more time is allowed, the more time a company has to get ruined. The probability of ruin is a continuous surplus process is higher than that in a discrete surplus process, since surplus is checked more often (continuously).

3. Summary of the Inequalities:

\[
\begin{align*}
\psi(u) &\geq \psi(u, t) \geq \tilde{\psi}(u, t) \\
\psi(u) &\geq \tilde{\psi}(u) \geq \tilde{\psi}(u, t) \\
\psi(u) &\geq \psi(u + k) \geq \psi(u, t) \\
\phi(u) &\leq \phi(u, t) \leq \tilde{\phi}(u, t) \\
\phi(u) &\leq \phi(u, t) \leq \phi(u, t)
\end{align*}
\]

4. Definitions of Ruin Functions:

<table>
<thead>
<tr>
<th>Horizon</th>
<th>Frequency of checking</th>
<th>Ruin probability</th>
<th>Survival probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Infinite</td>
<td>Continuous</td>
<td>$\psi(u)$</td>
<td>$\phi(u)$</td>
</tr>
<tr>
<td>Infinite</td>
<td>Annual</td>
<td>$\tilde{\psi}(u)$</td>
<td>$\tilde{\phi}(u)$</td>
</tr>
<tr>
<td>Through time t</td>
<td>Continuous</td>
<td>$\psi(u, t)$</td>
<td>$\phi(u, t)$</td>
</tr>
<tr>
<td>Through time t</td>
<td>Annual</td>
<td>$\tilde{\psi}(u, t)$</td>
<td>$\tilde{\phi}(u, t)$</td>
</tr>
</tbody>
</table>

5. Calculating Ruin Probabilities – Use the straightforward convolution method. Exhaustively tabulate all possible values of surplus at each discrete time $t$, with the probability of each value. When a value goes below 0, you no longer consider any further development for that surplus.