STT 459: Actuarial Models

Albert Cohen

Actuarial Sciences Program
Department of Mathematics
Department of Statistics and Probability
C336 Wells Hall
Michigan State University
East Lansing MI
48823
albert@math.msu.edu
acohen@stt.msu.edu
Some examples and theorem proofs in these slides, and on in class exam preparation slides, are taken from our textbook ”Loss Models: From Data to Decisions” 4th edition, by Klugman, Panjer, and Wilmott.

Please note that Wiley owns the copyright for that material. No portion of the Wiley textbook material may be reproduced in any part or by any means without the permission of the publisher. We are very thankful to the publisher for allowing posting of these notes on our class website.

Also, we will from time-to-time look at problems from released previous Exams C by the SOA. All such questions belong in copyright to the Society of Actuaries, and we make no claim on them. It is of course an honor to be able to present analysis of such examples here.
Recall our notion of a sample space $S = \Omega$, and attach to it a probability measure $P$ and the set of possible events $\mathcal{F}$.

Consider now the experiment of tossing 3 fair coins. Define the random variable $Y = \text{Number of heads that appear.}$ Then

- $P\{Y = 0\} = P\{(T, T, T)\}$
- $P\{Y = 1\} = P\{(T, T, H), (T, H, T), (H, T, T)\}$
- $P\{Y = 2\} = P\{(T, H, H), (H, T, H), (H, H, T)\}$
- $P\{Y = 3\} = P\{(H, H, H)\}$
Random Variables

Recall our notion of a sample space $S = \Omega$, and attach to it a probability measure $\mathbb{P}$ and the set of possible events $\mathcal{F}$. In our previous experience, $\mathcal{F} = 2^\Omega$. 

Consider now the experiment of tossing 3 fair coins. Define the random variable $Y = \text{Number of heads that appear}$. Then:

- $P\{Y = 0\} = P\{(T, T, T)\}$
- $P\{Y = 1\} = P\{(T, T, H), (T, H, T), (H, T, T)\}$
- $P\{Y = 2\} = P\{(T, H, H), (H, T, H), (H, H, T)\}$
- $P\{Y = 3\} = P\{(H, H, H)\}$
Random Variables

Recall our notion of a sample space \( S = \Omega \), and attach to it a probability measure \( P \) and the set of possible events \( \mathcal{F} \).

In our previous experience, \( \mathcal{F} = 2^\Omega \).

For now, we take this to be the case. We can define, on this Probability Space \((\Omega, \mathcal{F}, P)\) a Random Variable \( X \) as any function

\[
X : \Omega \rightarrow \mathbb{R} \quad (1)
\]

Consider now the experiment of tossing 3 fair coins. Define the random variable \( Y = \) Number of heads that appear. Then

\[
P\{Y = 0\} = P\{(T, T, T)\} \\
P\{Y = 1\} = P\{(T, T, H), (T, H, T), (H, T, T)\} \\
P\{Y = 2\} = P\{(T, H, H), (H, T, H), (H, H, T)\} \\
P\{Y = 3\} = P\{(H, H, H)\}
\]
Random Variables

Recall our notion of a sample space $S = \Omega$, and attach to it a probability measure $\mathbb{P}$ and the set of possible events $\mathcal{F}$.

In our previous experience, $\mathcal{F} = 2^\Omega$.

For now, we take this to be the case. We can define, on this Probability Space $(\Omega, \mathcal{F}, \mathbb{P})$ a Random Variable $X$ as any function

$$X : \Omega \rightarrow \mathbb{R}$$

Consider now the experiment of tossing 3 fair coins. Define the random variable $Y = \text{Number of heads that appear}$. Then
Random Variables

Recall our notion of a sample space $S = \Omega$, and attach to it a probability measure $\mathbb{P}$ and the set of possible events $\mathcal{F}$. In our previous experience, $\mathcal{F} = 2^\Omega$. For now, we take this to be the case. We can define, on this *Probability Space* $(\Omega, \mathcal{F}, \mathbb{P})$ a *Random Variable* $X$ as any function

$$X : \Omega \rightarrow \mathbb{R}$$ (1)

Consider now the experiment of tossing 3 fair coins. Define the random variable $Y = \text{Number of heads that appear}$. Then

$$\mathbb{P}[\{Y = 0\}] = \mathbb{P}[\{(T, T, T)\}]$$
$$\mathbb{P}[\{Y = 1\}] = \mathbb{P}[\{(T, T, H), (T, H, T), (H, T, T)\}]$$
$$\mathbb{P}[\{Y = 2\}] = \mathbb{P}[\{(T, H, H), (H, T, H), (H, H, T)\}]$$
$$\mathbb{P}[\{Y = 3\}] = \mathbb{P}[\{(H, H, H)\}]$$
Consider a random variable $X$ that takes on a countable number of values, denoted by the set $R \equiv \{x_1, x_2, \ldots \}$. For any $a \in R$, we define

$$p(a) = \mathbb{P}[X = a]$$

$$\sum_{i=1}^{\infty} p(x_i) = 1 \quad (2)$$
Expected Value

For a random variable $X$ that takes on a countable number of values, define

$$\mathbb{E}[X] = \sum_{x_i \in R} x_i p(x_i)$$  \hfill (3)

As an important example, consider an event $A$ in our $\sigma$-field and define the random variable $I_A = \begin{cases} 1, & \text{if } A \text{ occurs} \\ 0, & \text{if } A \text{ does not occur} \end{cases}$\hfill (4)

Then $\mathbb{E}[I_A] = P[A]$. 
Expected Value

For a random variable $X$ that takes on a countable number of values, define

$$\mathbb{E}[X] = \sum_{x_i \in R} x_i p(x_i)$$

As an important example, consider an event $A$ in our $\sigma-$field and define the random variable

$$I_A = \begin{cases} 1, & \text{if } A \text{ occurs} \\ 0, & \text{if } A \text{ does not occur} \end{cases}$$
For a random variable $X$ that takes on a countable number of values, define

$$\mathbb{E}[X] = \sum_{x_i \in R} x_i p(x_i)$$  \hfill (3)

As an important example, consider an event $A$ in our $\sigma-$field and define the random variable

$$I_A = \begin{cases} 
1, & \text{if } A \text{ occurs} \\
0, & \text{if } A \text{ does not occur}
\end{cases}$$  \hfill (4)

Then $\mathbb{E}[I_A] = \mathbb{P}[A]$. 
If $X$ is a random variable with mean $\mu = \mathbb{E}[X]$, then the variance $\text{Var}(X)$ and standard deviation $\text{SD}(X)$ are defined as

$$\text{Var}(X) = \mathbb{E} [(X - \mu)^2]$$
If $X$ is a random variable with mean $\mu = \mathbb{E}[X]$, then the variance $\text{Var}(X)$ and standard deviation $\text{SD}(X)$ are defined as

$$
\text{Var}(X) = \mathbb{E} [(X - \mu)^2] \\
= \mathbb{E} [X^2] - \mathbb{E}[X]^2
$$
If $X$ is a random variable with mean $\mu = \mathbb{E}[X]$, then the variance $\text{Var}(X)$ and standard deviation $\text{SD}(X)$ are defined as

$$\text{Var}(X) = \mathbb{E} \left[ (X - \mu)^2 \right]$$

$$= \mathbb{E} [X^2] - \mathbb{E}[X]^2$$

$$\text{SD}(X) = \sqrt{\text{Var}(X)}$$

A useful identity is

$$\text{Var}(a \cdot X + b) = a^2 \cdot \text{Var}(X)$$

Also notice that $\mathbb{E}[X^2] \geq (\mathbb{E}[X])^2$. This is true regardless of the distribution of $X$. 
Bernoulli and Binomial Random Variables

Define a *Bernoulli* random variable $X$ with $R = \{0, 1\}$ and

$$p = p(1) = \mathbb{P}[X = 1]$$

$$1 - p = p(0) = \mathbb{P}[X = 0]$$

This represents the outcome of 1 trial. Now, imagine this experiment is carried out $n$ times. Let $Y_{n,p}$ denote the outcome of this repeated experiment. We now have $R = \{0, 1, 2, \ldots, n\}$ and

$$p(i) = \mathbb{P}[Y_{n,p} = i] = \binom{n}{i} \cdot p^i \cdot (1 - p)^{n-i}$$

$$\mathbb{E}[Y_{n,p}] = n \cdot p$$

$$\text{Var}(Y_{n,p}) = n \cdot p \cdot (1 - p)$$
The Poisson Random Variable

Recall our r.v. $X$ with $p(i) = \frac{e^{-\lambda} \lambda^i}{i!}$. Straightforward computation shows that $\mathbb{E}[X] = \lambda = \text{Var}(X)$.

$X$ can in fact approximate $Y_{n,p}$ if $\lambda = n \cdot p$, for large $n$.

As an example, consider an item that is produced by a certain machine. The probability that it will be defective is $p = 0.1$. The probability that, in a sample of 10, at most one item will be defective, is

$$
\mathbb{P}[Y_{10,0.1} \leq 1] = \binom{10}{0} 0.1^0 \cdot 0.9^{10} + \binom{10}{1} 0.1^1 \cdot 0.9^9
$$

$$
= 0.7361 \approx 0.7358 = e^{-1} + e^{-1}
$$

$$
= e^{-(10 \cdot 0.1)} \cdot \frac{(10 \cdot 0.1)^0}{1} + e^{-(10 \cdot 0.1)} \cdot \frac{(10 \cdot 0.1)^1}{1}
$$

$$
= \mathbb{P}[X = 0] + \mathbb{P}[X = 1]
$$

(9)
### Geometric Random Variable

- **Number of trials required until success**: 
  \[ P[X = n] = (1 - p)^{n-1} \cdot p \]
- **Expected Value**: 
  \[ E[X] = p \cdot 1 + (1 - p) \cdot E[1 + X] \Rightarrow E[X] = \frac{1}{p} \]
• Geometric Random Variable: *Number of trials required until success*:
\[ P[X = n] = (1 - p)^{n-1} \cdot p \]
- \( \mathbb{E}[X] = p \cdot 1 + (1 - p) \cdot \mathbb{E}[1 + X] \Rightarrow \mathbb{E}[X] = \frac{1}{p} \)
- \( \mathbb{E}[X^2] = p \cdot 1^2 + (1 - p) \cdot \mathbb{E}[(X + 1)^2] \Rightarrow \mathbb{E}[X^2] = \frac{2-p}{p^2} \)
Other Discrete Probability Distributions

- Geometric Random Variable:  
  Number of trials required until success:  
  \[ P[X = n] = (1 - p)^{n-1} \cdot p \]
  - \( E[X] = p \cdot 1 + (1 - p) \cdot E[1 + X] \Rightarrow E[X] = \frac{1}{p} \)
  - \( E[X^2] = p \cdot 1^2 + (1 - p) \cdot E[(X + 1)^2] \Rightarrow E[X^2] = \frac{2-p}{p^2} \)
  - \( Var(X) = E[X^2] - (E[X])^2 = \frac{1-p}{p^2} \)
Other Discrete Probability Distributions

- Geometric Random Variable: *Number of trials required until success*: 
  \[ \mathbb{P}[X = n] = (1 - p)^{n-1} \cdot p \]
  - \[ \mathbb{E}[X] = p \cdot 1 + (1 - p) \cdot \mathbb{E}[1 + X] \Rightarrow \mathbb{E}[X] = \frac{1}{p} \]
  - \[ \mathbb{E}[X^2] = p \cdot 1^2 + (1 - p) \cdot \mathbb{E}[(X + 1)^2] \Rightarrow \mathbb{E}[X^2] = \frac{2-p}{p^2} \]
  - \[ \text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{1-p}{p^2} \]

- Negative Binomial Random Variable: *n trials needed for r successes*: 
  \[ \mathbb{P}[X_{r,p} = n] = \binom{n-1}{r-1} p^r \cdot (1 - p)^{n-r} \]
  - \[ \mathbb{E}[X_{r,p}] = \frac{r}{p} \]
• **Geometric Random Variable:** *Number of trials required until success*: 
\[ \mathbb{P}[X = n] = (1 - p)^{n-1} \cdot p \]
- \[ \mathbb{E}[X] = p \cdot 1 + (1 - p) \cdot \mathbb{E}[1 + X] \Rightarrow \mathbb{E}[X] = \frac{1}{p} \]
- \[ \mathbb{E}[X^2] = p \cdot 1^2 + (1 - p) \cdot \mathbb{E}[(X + 1)^2] \Rightarrow \mathbb{E}[X^2] = \frac{2-p}{p^2} \]
- \[ \text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{1-p}{p^2} \]

• **Negative Binomial Random Variable:** *n trials needed for r successes*: 
\[ \mathbb{P}[X_{r,p} = n] = \binom{n-1}{r-1} p^r \cdot (1 - p)^{n-r} \]
- \[ \mathbb{E}[X_{r,p}] = \frac{r}{p} \]
- \[ \mathbb{E}[X_{r,p}^2] = \frac{r}{p} \cdot \left( \frac{r+1}{p} - 1 \right) \]
Other Discrete Probability Distributions

- Geometric Random Variable: *Number of trials required until success*:
  \[ P(X = n) = (1 - p)^{n-1} \cdot p \]
  - \[ E[X] = p \cdot 1 + (1 - p) \cdot E[1 + X] \Rightarrow E[X] = \frac{1}{p} \]
  - \[ E[X^2] = p \cdot 1^2 + (1 - p) \cdot E[(X + 1)^2] \Rightarrow E[X^2] = \frac{2-p}{p^2} \]
  - \[ Var(X) = E[X^2] - (E[X])^2 = \frac{1-p}{p^2} \]

- Negative Binomial Random Variable: *n trials needed for r successes*:
  \[ P(X_{r,p} = n) = \binom{n-1}{r-1} p^r \cdot (1 - p)^{n-r} \]
  - \[ E[X_{r,p}] = \frac{r}{p} \]
  - \[ E[X_{r,p}^2] = \frac{r}{p} \cdot \left( \frac{r+1}{p} - 1 \right) \]
  - \[ Var(X_{r,p}) = E[X_{r,p}^2] - (E[X_{r,p}])^2 = \frac{r \cdot (1-p)}{p^2} \]
Hypergeometric Random Variable: sample of size $n$ chosen randomly from population $m$ white balls and $N - m$ black balls. Let $Z_{n,N,m}$ denote the number of white balls selected:

- $\mathbb{P}(Z_{n,N,m} = i) = \frac{\binom{m}{i}\binom{N-m}{n-i}}{\binom{N}{n}}$ for $i = 0, 1, \ldots, n$
Hypergeometric Random Variable: *sample of size n chosen randomly from population m white balls and N − m black balls.* Let \( Z_{n,N,m} \) denote the number of white balls selected:

- \( \mathbb{P}[Z_{n,N,m} = i] = \binom{m}{i} \binom{N-m}{n-i} \frac{1}{\binom{N}{n}} \) for \( i = 0, 1, \ldots, n \)

- \( \mathbb{E}[Z_{n,N,m}] = \frac{n \cdot m}{N} \)

- \( \mathbb{E}[Z^2_{n,N,m}] = \frac{n \cdot m}{N} \left( \frac{(n-1)(m-1)}{N-1} + 1 \right) \)

- \( \text{Var}(Z_{n,N,m}) = \mathbb{E}[Z^2_{n,N,m}] - (\mathbb{E}[Z_{n,N,m}])^2 = n \cdot \frac{m}{N} \cdot (1 - \frac{m}{N}) \cdot \left(1 - \frac{n-1}{N-1}\right) \)
Hypergeometric Random Variable: sample of size \( n \) chosen randomly from population \( m \) white balls and \( N - m \) black balls. Let \( Z_{n,N,m} \) denote the number of white balls selected:

- \( \mathbb{P}[Z_{n,N,m} = i] = \frac{\binom{m}{i}\binom{N-m}{n-i}}{\binom{N}{n}} \) for \( i = 0, 1, \ldots, n \)
- \( \mathbb{E}[Z_{n,N,m}] = \frac{n \cdot m}{N} \)
- \( \mathbb{E}[Z_{n,N,m}^2] = \frac{n \cdot m}{N} \cdot \left( \frac{(n-1)(m-1)}{N-1} + 1 \right) \)
- \( \text{Var}(Z_{n,N,m}) = \mathbb{E}[Z_{n,N,m}^2] - (\mathbb{E}[Z_{n,N,m}])^2 = n \cdot \frac{m}{N} \cdot \left( 1 - \frac{m}{N} \right) \cdot \left( 1 - \frac{n-1}{N-1} \right) \)

Zipf (Pareto) Distribution: Assume \( \alpha > 0 \) and that

\[
\mathbb{P}[X = k] = \left( \sum_{l=1}^{\infty} \frac{1}{l^{1+\alpha}} \right)^{-1} \cdot \frac{1}{k^{1+\alpha}}.
\]

Then \( \mathbb{E}[X] = ? \)
Expectation is Linear

For a collection of random variables \( \{X_i\}_{i=1}^n \), we have

\[
\mathbb{E} \left[ \sum_{i=1}^n X_i \right] = \sum_{i=1}^n \mathbb{E}[X_i].
\]

As an example, if \( n \) dice are rolled, what is the expected value for the total number of dots?
Expectation is Linear

For a collection of random variables \( \{X_i\}_i=1^n \), we have
\[
\mathbb{E} \left[ \sum_{i=1}^n X_i \right] = \sum_{i=1}^n \mathbb{E}[X_i].
\]

As an example, if \( n \) die are rolled, what is the expected value for the total number of dots?
Recall that for an r.v. $X$, we define $F(b) := \mathbb{P}[X \leq b]$. It follows from the continuity property of probability measures that

- If $a < b$ then $F(a) < F(b)$
- $\lim_{b \to \infty} F(b) = 1$
- $\lim_{b \to -\infty} F(b) = 0$
- $\lim_{b_n \downarrow b} F(b_n) = F(b)$
Recall that for an r.v. $X$, we define $F(b) := \mathbb{P}[X \leq b]$. It follows from the continuity property of probability measures that

- If $a < b$ then $F(a) < F(b)$
- $\lim_{b \to \infty} F(b) = 1$
- $\lim_{b \to -\infty} F(b) = 0$
- $\lim_{b_n \downarrow b} F(b_n) = F(b)$

As an example,

$$\mathbb{P}[X < b] = \mathbb{P}\left[ \lim_{n \to \infty} \left\{ X \leq b - \frac{1}{n} \right\} \right]$$

$$= \lim_{n \to \infty} \mathbb{P}\left[ X \leq b - \frac{1}{n} \right]$$

$$= \lim_{n \to \infty} F\left( b - \frac{1}{n} \right)$$

(10)
Continuous Random Variables

Recall our notion of a sample space $S = \Omega$, and attach to it a probability measure $\mathbb{P}$ and the set of possible events $\mathcal{F}$.
Continuous Random Variables

Recall our notion of a sample space $S = \Omega$, and attach to it a probability measure $\mathbb{P}$ and the set of possible events $\mathcal{F}$.

If we take a random variable that can take on any real value, then we call it a *continuous random variable* if attached to it there is a non-negative function $f : \mathbb{R} \to \mathbb{R}_+$ where

$$
\mathbb{P}[X \in B] = \int_B f(x) \, dx \tag{11}
$$

for any measurable set $B$. We can consider all practical sets to be measurable. Of course, it follows from our definition that we must have

$$
\mathbb{P}[X \in \mathbb{R}] = \int_{-\infty}^\infty f(x) \, dx = 1.
$$

For intervals $[a, b]$ and $(-\infty, b]$, we have
Continuous Random Variables

Recall our notion of a sample space $S = \Omega$, and attach to it a probability measure $\mathbb{P}$ and the set of possible events $\mathcal{F}$.

If we take a random variable that can take on any real value, then we call it a **continuous random variable** if attached to it there is a non-negative function $f : \mathbb{R} \to \mathbb{R}^+$ where

$$\mathbb{P}[X \in B] = \int_B f(x) \, dx \quad (11)$$

for any measurable set $B$. We can consider all practical sets to be measurable. Of course, it follows from our definition that we must have

$$\mathbb{P}[X \in \mathbb{R}] = \int_{-\infty}^{\infty} f(x) \, dx = 1.$$ 

For intervals $[a, b]$ and $(-\infty, b]$, we have

$$\mathbb{P}[a \leq X \leq b] = \int_a^b f(x) \, dx \quad (12)$$

$$\mathbb{P}[X \leq b] = \int_{-\infty}^b f(x) \, dx$$
Example

Suppose that $X$ is c.r.v. with probability density function

$$f(x) = \begin{cases} C \cdot (2 - x), & \text{if } 0 < x < 2 \\ 0, & \text{otherwise} \end{cases}$$

(13)

What is $C$? Compute $P[X > 1]$
Example

Suppose that X is c.r.v. with probability density function

\[ f(x) = \begin{cases} 
C \cdot (2 - x), & \text{if } 0 < x < 2 \\
0, & \text{otherwise} 
\end{cases} \]  \tag{13}

What is $C$? Compute $P[X > 1]$

Answer: Since $f$ is a density, we require

\[ 1 = \int_{-\infty}^{\infty} f(x) \, dx = C \cdot \int_{0}^{2} (2 - x) \, dx = C \cdot 2 \]
Example

Suppose that \( X \) is c.r.v. with probability density function

\[
f(x) = \begin{cases} 
C \cdot (2 - x), & \text{if } 0 < x < 2 \\
0, & \text{otherwise.}
\end{cases}
\] (13)

What is \( C \)? Compute \( \mathbb{P}[X > 1] \)

Answer: Since \( f \) is a density, we require

\[
1 = \int_{-\infty}^{\infty} f(x) \, dx = C \cdot \int_{0}^{2} (2 - x) \, dx = C \cdot 2 \Rightarrow C = \frac{1}{2}
\]
Example

Suppose that $X$ is c.r.v. with probability density function

$$f(x) = \begin{cases} 
  C \cdot (2 - x), & \text{if } 0 < x < 2 \\
  0, & \text{otherwise}.
\end{cases} \quad (13)$$

What is $C$? Compute $\mathbb{P}[X > 1]$

Answer: Since $f$ is a density, we require

$$1 = \int_{-\infty}^{\infty} f(x) \, dx = C \cdot \int_{0}^{2} (2 - x) \, dx = C \cdot 2 \Rightarrow C = \frac{1}{2}$$

$$\mathbb{P}[X > 1] = \int_{1}^{\infty} f(x) \, dx = \frac{1}{2} \cdot \int_{1}^{2} (2 - x) \, dx$$

$$= \frac{1}{2} \cdot \left[ (2(2) - \frac{1}{2}(2^2)) - (2(1) - \frac{1}{2}(1^2)) \right] = \frac{1}{4} \quad (14)$$
From c.d.f to p.d.f

Recall

\[ F(y) = \mathbb{P}[X \leq y] \quad (15) \]

and so

\[ \mathbb{P} \left[ x - \frac{\Delta x}{2} \leq X \leq x + \frac{\Delta x}{2} \right] = \int_{x-\frac{\Delta x}{2}}^{x+\frac{\Delta x}{2}} f(u)du \approx f(x)\Delta x. \quad (16) \]

It follows that if \( Y = 2X \), then

\[
F_Y(x) := \mathbb{P}[Y \leq x] = \mathbb{P}[2X \leq x] = \mathbb{P}[X \leq 0.5x] = F_X(0.5x) \\
\]

\[ f_Y(x) := \frac{d}{dx} F_Y(x) = \frac{d}{dx} F_X(0.5x) = 0.5f_X(0.5x). \quad (17) \]
Define $E[X] = \int_{-\infty}^{\infty} x \cdot f(x)dx$ for any c.r.v. $X$ with p.d.f. $X$. For any real valued function $g$, we can define the expectation of the transformed c.r.v $Y = g(X)$ as

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) \cdot f(x)dx.$$

(18)
Expectations

The following Lemma is useful in many computations. The proof can be seen via an interchange (Fubini) of the order of integration.

**Lemma**

For any nonnegative random variable $Y$, we have $\mathbb{E}[Y] = \int_0^\infty \mathbb{P}[Y > y]dy$.

**Proof.**

\[
\int_0^\infty \mathbb{P}[Y > y]dy = \int_0^\infty \left( \int_y^\infty f_Y(x)dx \right)dy \\
= \int_0^\infty \left( \int_0^x f_Y(x)dy \right)dx \quad (19) \\
= \int_0^\infty x \cdot f_Y(x)dx = \mathbb{E}[X].
\]
Uniform Random Variables

If we define

\[ f(x) := \begin{cases} 
\frac{1}{B - A}, & \text{if } A < x < B \\
0, & \text{otherwise}
\end{cases} \tag{20} \]

then the distribution \( F \) is computed to be

\[ F(x) = \begin{cases} 
0, & \text{if } x \leq A \\
\frac{x - A}{B - A}, & \text{if } A < x < B \\
1, & \text{if } x \geq B \tag{21} \end{cases} \]

and so

\[ \mathbb{E}[X] = \int_{-\infty}^{\infty} xf(x)dx = \int_{A}^{B} x \cdot \frac{1}{B - A} = \frac{B + A}{2} \tag{22} \]

\[ \text{Var}(X) = \mathbb{E}[X^2] - \left( \frac{B + A}{2} \right)^2 = \frac{(B - A)^2}{12}. \]
A 1-unit long city block has a mailbox somewhere at a point $U$ that is uniformly distributed over $(0, 1)$. Determine the expected length of the section of the block that contains an envelope store $E$, where $0 \leq E \leq 1$. Also, for what value of $E$ is this expected value maximized?
Take $L_E(U)$ as the length of the sub-block that contains the envelope store $E$ and note that

$$L_E(U) = \begin{cases} 
1 - U, & \text{if } U < E \\
U, & \text{if } U > E.
\end{cases}$$

(23)

It follows that

$$E[L_E(U)] := \int_0^1 L_E(u) \cdot f(u) du = \int_0^1 L_E(u) du$$

$$= \int_0^E (1 - u) du + \int_E^1 u du = \frac{1}{2} + E(1 - E)$$

(24)
We say that $X$ is a *Normal Random Variable* with parameters $\mu, \sigma^2$ if

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \text{ for } -\infty < x < \infty$$  \hspace{1cm} (25)

Furthermore, we say that $Z$ is a *Standard Normal Random Variable* if it is Normal with parameters $\mu = 0, \sigma = 1$. Consider

$$z = \frac{x - \mu}{\sigma}$$

$$\int_{-\infty}^{\infty} f_X(x) \, dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \, dx$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(z)^2}{2}} \, dz$$

$$= \int_{-\infty}^{\infty} f_Z(z) \, dz \hspace{1cm} (26)$$
We use the notation $X \sim N(\mu, \sigma^2)$ and $Z \sim N(0, 1)$. By our transformation above, it can be seen that if $Z \sim N(0, 1)$, then $X = \mu + \sigma \cdot Z \sim N(\mu, \sigma^2)$. We can see this via

\[
\Phi(z) := F_Z(z) = \Pr[Z \leq z] = \Pr\left[\frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma}\right] = \Pr[X \leq x] =: F_X(x)
\]
We use the notation $X \sim N(\mu, \sigma^2)$ and $Z \sim N(0, 1)$. By our transformation above, it can be seen that if $Z \sim N(0, 1)$, then $X = \mu + \sigma \cdot Z \sim N(\mu, \sigma^2)$. We can see this via

$$
\Phi(z) := F_Z(z) = \mathbb{P}[Z \leq z] = \mathbb{P} \left[ \frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma} \right] = \mathbb{P}[X \leq x] = F_X(x)
$$

$$
\frac{1}{\sigma} f_Z \left( \frac{x - \mu}{\sigma} \right) = \frac{d}{dx} F_Z(z) = \frac{d}{dx} F_X(x) = f_X(x)
$$
We use the notation $X \sim \mathcal{N}(\mu, \sigma^2)$ and $Z \sim \mathcal{N}(0, 1)$. By our transformation above, it can be seen that if $Z \sim \mathcal{N}(0, 1)$, then $X = \mu + \sigma \cdot Z \sim \mathcal{N}(\mu, \sigma^2)$. We can see this via

$$\Phi(z) := F_Z(z) = \mathbb{P}[Z \leq z]$$

$$= \mathbb{P}\left[ \frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma} \right]$$

$$= \mathbb{P}[X \leq x] =: F_X(x)$$

$$\frac{1}{\sigma} f_Z \left( \frac{x - \mu}{\sigma} \right) = \frac{d}{dx} F_Z(z)$$

$$= \frac{d}{dx} F_X(x) = f_X(x)$$

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(z)^2}{2}}$$

(27)
Theorem

If $S_n$ is the number of successes that occur when $n$ independent trials, each resulting in a success with probability $p$, are performed, then for any $a < b$, we have

$$\lim_{n \to \infty} \mathbb{P} \left[ a \leq \frac{S_n - n \cdot p}{\sqrt{n \cdot p \cdot (1 - p)}} \leq b \right] = \mathbb{P}[a \leq Z \leq b] = \Phi(b) - \Phi(a)$$

(28)
For any $\lambda > 0$, define the random variable $X$ with p.d.f.

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{if } x \geq 0 \\ 0, & \text{if } x < 0. \end{cases}$$

(29)

Then, if $x \geq 0$, we have $F(x) = \int_0^x \lambda e^{-\lambda y} dy = 1 - e^{-\lambda x}$. 
Exponential Random Variables

Furthermore, using integration by parts, we also compute

\[
\mathbb{E}[X^n] = \int_0^\infty x^n \lambda e^{-\lambda x} \, dx
\]

\[
= 0 + \int_0^\infty nx^{n-1} e^{-\lambda x} \, dx
\]

\[
= \frac{n}{\lambda} \int_0^\infty x^{n-1} \lambda e^{-\lambda x} \, dx
\]

\[
= \frac{n}{\lambda} \mathbb{E}[X^{n-1}]
\]
Furthermore, using integration by parts, we also compute

\[ \mathbb{E}[X^n] = \int_0^\infty x^n \lambda e^{-\lambda x} \, dx \]

\[ = 0 + \int_0^\infty nx^{n-1} e^{-\lambda x} \, dx \]

\[ = \frac{n}{\lambda} \int_0^\infty x^{n-1} \lambda e^{-\lambda x} \, dx \]

\[ = \frac{n}{\lambda} \mathbb{E}[X^{n-1}] \]

\[ \mathbb{E}[X^1] = \frac{1}{\lambda} \mathbb{E}[1] = \frac{1}{\lambda} \]

\[ \mathbb{E}[X^2] = \frac{2}{\lambda} \mathbb{E}[X] = \frac{2}{\lambda^2} \]

\[ \text{Var}(X) = \frac{1}{\lambda^2} \]
Notice that \( X \) is a *Memoryless* random variable:

\[
\Pr[X > s + t | X > t] = \frac{\Pr[X > s + t]}{\Pr[X > t]}
= \frac{1 - (1 - e^{-\lambda(t+s)})}{1 - (1 - e^{-\lambda t})} \\
= \frac{e^{-\lambda(t+s)}}{e^{-\lambda t}}
= e^{-\lambda s} = \Pr[X > s]
\]
Define the **Hazard Rate** for any crv $X$ as

$$\lambda(t) := \frac{f(t)}{1 - F(t)} = \frac{d}{dt}F(t)$$  \hspace{1cm} (32)

It follows, by solution of the above differential equation, that an equivalent definition is

$$F(t) = 1 - e^{-\int_0^t \lambda(s)ds}$$

$$F(0) = 1$$  \hspace{1cm} (33)

Crv’s may be defined via their Hazard rates. For example:
- Rayleigh distributed random variable: $\lambda(t) = a + bt$
- Exponential distributed random variable: $\lambda(t) = \lambda$. 

Albert Cohen (MSU)
STT 459: Actuarial Models
MSU Spring 2015  27 / 193
Consider a r.v $X$ with p.d.f. 

$$f(x) = \begin{cases} 
\frac{\lambda e^{-\lambda x}(\lambda x)^{\alpha-1}}{\Gamma(\alpha)}, & \text{if } x \geq 0 \\
0, & \text{if } x < 0.
\end{cases}$$ \hspace{1cm} (34)$$

where the Gamma function $\Gamma(\alpha)$ is defined as $\int_0^\infty e^{-y}y^{\alpha-1}dy$. Integration by parts shows that for $\alpha > 1$

$$\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$$
Consider a crv $X$ with p.d.f.

$$f(x) = \begin{cases} \frac{\lambda e^{-\lambda x}(\lambda x)^{\alpha-1}}{\Gamma(\alpha)}, & \text{if } x \geq 0 \\ 0, & \text{if } x < 0. \end{cases}$$

(34)

where the $Gamma$ function $\Gamma(\alpha)$ is defined as $\int_0^{\infty} e^{-y} y^{\alpha-1} dy$. Integration by parts shows that for $\alpha > 1$

$$\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$$

$$\Gamma(n) = (n - 1)! \text{ for } n \in \mathbb{N}$$

(35)
Assume that you are waiting for $n$ events to occur, and that $N(t)$ is the number of events that occur up to time $t$, which is assumed to be Poisson ($\lambda t$) distributed. Let $T_n$ denote the time at which the $n^{th}$ event occurs.
Gamma Distribution

It follows that

\[ F(t) = P[T_n \leq t] = P[N(t) \geq n] = \sum_{j=n}^{\infty} P[N(t) = j] = \sum_{j=n}^{\infty} e^{-\lambda t} (\lambda t)^j / j!\]

\[ f(t) = F'(t) = \sum_{j=n}^{\infty} e^{-\lambda t} \cdot j \cdot (\lambda t)^{j-1} \cdot \lambda / j! - \sum_{j=n}^{\infty} \lambda e^{-\lambda t} (\lambda t)^j / j! = \frac{\lambda e^{-\lambda t} (\lambda t)^{n-1}}{(n-1)!}\]
Weibull Distribution

Consider a crv $X$ with p.d.f.

\[
f(x) = \begin{cases} 
\frac{\beta}{\alpha} \left( \frac{x - \nu}{\alpha} \right)^{\beta-1} e^{-\left( \frac{x - \nu}{\alpha} \right)^{\beta}}, & \text{if } x > \nu \\
0, & \text{if } x \leq \nu 
\end{cases}
\]  

The corresponding cdf is

\[
F(x) = \begin{cases} 
0, & \text{if } x \leq \nu \\
1 - e^{-\left( \frac{x - \nu}{\alpha} \right)^{\beta}}, & \text{if } x > \nu 
\end{cases}
\]
Consider a crv $\theta \sim U \left[ -\frac{\pi}{2}, \frac{\pi}{2} \right]$, and a transformation to another random variable $X = \tan(\theta)$. Then

$$F(x) = \mathbb{P}[X \leq x]$$
$$= \mathbb{P}[\tan(\theta) \leq x]$$
$$= \mathbb{P}[\theta \leq \tan^{-1}(x)]$$
$$= \frac{1}{\pi} \left( \tan^{-1}(x) - \left( -\frac{\pi}{2} \right) \right)$$
$$= \frac{1}{2} + \frac{1}{\pi} \tan^{-1}(x)$$

$$\Rightarrow f(x) = F'(x) = \frac{1}{\pi} \frac{1}{1 + x^2}$$
Consider a crv $X$ with p.d.f.

$$f(x) = \begin{cases} \frac{x^{a-1}(1 - x)^{b-1}}{B(a, b)}, & \text{if } 0 < x < 1 \\ 0, & \text{otherwise.} \end{cases}$$  \quad (40)$$

where $B(a, b)$ is defined as $\int_0^1 x^{a-1}(1 - x)^{b-1} \, dx$. It can be shown that

$$B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a + b)}$$
Consider a crv $X$ with p.d.f.

$$f(x) = \begin{cases} 
    x^{a-1}(1-x)^{b-1} & \text{if } 0 < x < 1 \\
    0 & \text{otherwise.}
\end{cases}$$

(40)

where $B(a, b)$ is defined as $\int_0^1 x^{a-1}(1-x)^{b-1}dx$. It can be shown that

$$B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

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

$$\text{Var}(X) = \frac{ab}{(a+b)^2(a+b+1)}$$

(41)
Distribution of a Function of a Random Variable

**Theorem**

Let $X$ be a continuous random variable with p.d.f. $f_X$. Suppose that $g(x)$ is a strictly monotonic, differentiable function of $x$. Then the random variable $Y = g(X)$ has a p.d.f. defined by

$$f_Y(y) = \begin{cases} 
  f_X \left[ g^{-1}(y) \right] \left| \frac{d}{dy} g^{-1}(y) \right|, & \text{if, for some } x, \ y = g(x) \\
  0, & \text{if, for all } x, \ y \neq g(x) 
\end{cases} \quad (42)$$
Proof.

Suppose that \( y = g(x) \) for some \( x \). Then, with \( Y = g(X) \),

\[
f_Y(y) = \frac{d}{dy} F_Y(y) \\
= \frac{d}{dy} \mathbb{P}[g(X) \leq y] \\
= \frac{d}{dy} \mathbb{P}[X \leq g^{-1}(y)] \\
= \frac{d}{dy} F_X(g^{-1}(y)) \\
= f_X(g^{-1}(y)) \frac{d}{dy} g^{-1}(y)
\]

(43)

If \( y \neq g(x) \) for any \( x \), then \( F(y) \) is either 0 or 1 \( \Rightarrow f_Y(y) = 0 \).
Examples (Can we use the previous theorem?)

Let $X$ have p.d.f $f_X$, and define $Y = X^2$, $Z = |X|$. Then

$$f_Y(y) = \frac{d}{dy} F_Y(y)$$

$$= \frac{d}{dy} \mathbb{P}[X^2 \leq y]$$

$$= \frac{d}{dy} \mathbb{P}[-\sqrt{y} \leq X \leq \sqrt{y}]$$

$$= \frac{d}{dy} (F_X(\sqrt{y}) - F_X(-\sqrt{y}))$$

$$= \frac{1}{2\sqrt{y}} [f_X(\sqrt{y}) + f_X(-\sqrt{y})]$$

$$f_Z(z) = \frac{d}{dz} \mathbb{P}[-z \leq X \leq z]$$

$$= f_X(z) + f_X(-z)$$
Mixed Distributions

There are random variables $X$ such that, for a countable number of points $x_i$,

$$
\Pr[X = x_i] = p_i > 0
$$

$$
\sum_{i=0}^{\infty} p_i \leq 1.
$$

(45)

As an example, consider $X$ such that

$$
\Pr[X = 0] = \frac{1}{2}
$$

$$
f_X(x) = \frac{1}{2}e^{-x} \text{ for } x > 0.
$$

(46)
Please work through

- 2.2
- 2.3, 2.4, 2.5
Given a random variable $X$, we define the $k^{th}$ **Raw Moment** via,

$$
\mu'_k := \mathbb{E}[X^k]
$$

and the $k^{th}$ **Central Moment** via,

$$
\mu_k := \mathbb{E}[(X - \mu'_1)^k].
$$

**HW:** Compute the first 5 raw and central moments for a random variable $X$ with $f_X(x) = \lambda e^{-\lambda x}$ for all $x \geq 0$. 
Given a random variable $X$, we define

- $\mu_1' := \mu = \text{Mean}$
- $\mu_2 = \mathbb{E}[(X - \mu_1')^2] := \sigma^2 = \text{Variance}$
- $\frac{\sigma}{\mu} := \text{Coefficient of Variation}$
- $\gamma_1 := \frac{\mu_3}{\sigma^3} = \text{Skewness}$
- $\gamma_2 := \frac{\mu_4}{\sigma^4} = \text{Kurtosis}$
Given a random variable $X$, a value $u \in \mathbb{R}$, and a value $d \in \mathbb{R}$ with $\mathbb{P}[X > d] > 0$ we define

$$Y^P := X - d = \textbf{Excess Loss Variable}$$

$$e_X(d) := \mathbb{E}[Y^P \mid Y^P > 0] = \textbf{Mean Excess Loss Function}$$

$$e^k_X(d) := \mathbb{E}[(Y^P)^k \mid Y^P > 0]$$

$$Y^L := (X - d)_+ = \textbf{Left Censored and Shifted Variable}$$

$$Y := X \wedge u = \textbf{Limited Loss Variable}$$

$$\mathbb{E}[Y] := \textbf{Limited Expected Value}.$$
Please work through

- 3.2
- 3.3, 3.4, 3.5
- 3.6, 3.7, 3.11
- 3.14, 3.15, 3.16
Define (any number) \( \pi_p \) as the \( 100p^{th} \) percentile if it satisfies

\[
F(\pi_p^-) \leq p \leq F(\pi_p).
\]  

(51)

The \( 50^{th} \) percentile is also called the **median**. For all \( z \in \mathbb{R} \), we define

\[
M_X(z) := \mathbb{E}[e^{zX}] = \text{Moment Generating Function}
\]

\[
P_X(z) := \mathbb{E}[z^X] = \text{Probability Generating Function}
\]

(52)

for which these expectations exist. Note that \( M_X(z) = P_X(e^z) \).
Moment Generating Functions

**Theorem**

Let $S_k = X_1 + \ldots + X_k$, where the $X$ are independent. Then, provided the following exist:

$$M_{S_k}(z) = \prod_{j=1}^{k} M_{X_j}(z)$$

$$P_{S_k}(z) = \prod_{j=1}^{k} P_{X_j}(z)$$

(53)
Example

Consider a loss model where at time $t > 0$, the liability with a randomly selected individual is distributed as

$$
P[X_t \geq x] = e^{-\frac{x}{t}} \text{ for } x \geq 0, t > 0$$  \hspace{1cm} (54)

$$P[X_0 = 0] = 1.$$
Consider a loss model where at time $t > 0$, the liability with a randomly selected individual is distributed as

\[ \mathbb{P}[X_t \geq x] = e^{-\frac{x}{t}} \quad \text{for } x \geq 0, t > 0 \]

\[ \mathbb{P}[X_0 = 0] = 1. \]  

(54)

It follows that

\[ M_{X_t}(z) = \mathbb{E}[e^{zX_t}] = \int_0^\infty e^{zx} f_{X_t}(x) \, dx = \int_0^\infty e^{zx} \cdot -\frac{d}{dx} \left( e^{-\frac{x}{t}} \right) \, dx \]

\[ = \int_0^\infty e^{zx} \frac{1}{t} e^{-\frac{x}{t}} \, dx = \frac{1}{t} \int_0^\infty e^{-(\frac{1}{t} - z)x} \, dx = \frac{1}{t} \cdot \frac{1}{\frac{1}{t} - z} \]  

(55)

\[ = \frac{1}{1 - zt}. \]
Consider a loss model where at time $t > 0$, the liability with a randomly selected individual is distributed as

$$
P[X_t \geq x] = e^{-\frac{x}{t}} \text{ for } x \geq 0, \ t > 0$$

$$
P[X_0 = 0] = 1. \tag{54}$$

It follows that

$$
M_X(z) = \mathbb{E}[e^{zX_t}] = \int_0^\infty e^{zx} f_X(x) dx = \int_0^\infty e^{zx} \cdot - \frac{d}{dx} \left( e^{-\frac{x}{t}} \right) dx
$$

$$
= \int_0^\infty e^{zx} \frac{1}{t} e^{-\frac{x}{t}} dx = \frac{1}{t} \int_0^\infty e^{-\left(\frac{1}{t} - z\right)x} dx = \frac{1}{t} \frac{1}{\frac{1}{t} - z} \tag{55}
$$

$$
= \frac{1}{1 - zt}.
$$

**HW:** For $S_k(t) := \sum_{j=1}^k X_t^j$, compute $M_{S_k(t)}(z)$. 

Example

For a standard Normal random variable $Z \sim N(0, 1)$, we have

$$M_Z(t) = \mathbb{E}[e^{tZ}] = \int_{-\infty}^{\infty} e^{t \cdot z} f_Z(z) \, dz = \int_{-\infty}^{\infty} e^{t \cdot z} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} z^2} \, dz$$

$$= \int_{-\infty}^{\infty} e^{\frac{1}{2} t^2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} (z-t)^2} \, dz = e^{\frac{1}{2} t^2}.$$  \hspace{1cm} (56)

Therefore, for $Y = \mu + \sigma Z \sim N(\mu, \sigma^2)$, it follows that

$$M_Y(t) = \mathbb{E}[e^{t(\mu+\sigma Z)}] = e^{t\mu} \mathbb{E}[e^{t\sigma Z}] = e^{t\mu} e^{\frac{1}{2} (t\sigma)^2} = e^{t\mu + \frac{1}{2} t^2 \sigma^2}. \hspace{1cm} (57)$$
Consider an i.i.d sequence \( \{X_i\}_{i=1}^n \) such that all the \( X_i \sim X \), where for a parameter \( \theta > 0 \),

\[
f_X(x) = \frac{1}{\theta} e^{-\frac{x}{\theta}}
\]  
(58)

Define \( \bar{S}_n := \frac{X_1 + \ldots + X_n}{n} \), and so

\[
\mathbb{E}[\bar{S}_n] = \theta
\]

\[
\text{Var}[\bar{S}_n] = \frac{\theta^2}{n}.
\]  
(59)
It follows that

\[ M_{\bar{S}_n}(t) = \mathbb{E}[e^{t\bar{S}_n}] = \left[ \mathbb{E} \left[ e^{\frac{t}{n}X} \right] \right]^n \]

\[ = \left[ M_X \left( \frac{t}{n} \right) \right]^n \]

\[ = \left[ \frac{1}{1 - \frac{t\theta}{n}} \right]^n \]

\[ \therefore \lim_{n \to \infty} M_{\bar{S}_n}(t) = \lim_{n \to \infty} \left[ \frac{1}{1 - \frac{t\theta}{n}} \right]^n = e^{t\theta} \]

\[ \therefore \mathbb{P}_{\bar{S}_n} \Rightarrow \bar{\mathbb{P}}_{\bar{S}} := \mathcal{N}(\theta, 0). \]
Furthermore, define

\[ Z_n := \frac{\bar{S}_n - \mathbb{E}[\bar{S}_n]}{\sqrt{\text{Var}[\bar{S}_n]}} = \sqrt{n} \cdot \frac{\bar{S}_n - \theta}{\theta}. \quad (61) \]

It follows that

\[ M_{Z_n}(t) = \mathbb{E}[e^{tZ_n}] = \mathbb{E}\left[e^{t\sqrt{n} \cdot \frac{\bar{S}_n - \theta}{\theta}}\right] = e^{-t\sqrt{n}} \mathbb{E}\left[e^{t\sqrt{n} \cdot \frac{\bar{S}_n}{\theta}}\right] = e^{-t\sqrt{n}} M_{\bar{S}_n}\left(\frac{t\sqrt{n}}{\theta}\right) \]

\[ = e^{-t\sqrt{n}} \left[\frac{1}{1 - \frac{t}{\sqrt{n}}}\right]^n \rightarrow e^{\frac{1}{2}t^2} \]

\[ \therefore \mathbb{P}_{Z_n} \Rightarrow \mathbb{P}_{Z} := \mathcal{N}(0, 1). \]
Please work through

- 3.21
- 3.23
- 3.24
Consider a pair of random variables \((X_1, X_2)\) with \(\mathbb{E}[X_1] = \mathbb{E}[X_2]\), and recall the notation

\[ S_i(x) := \mathbb{P}[X_i > x] = 1 - F_i(x). \tag{63} \]

It can be shown via L’Hopital’s rule that

\[ \lim_{x \to \infty} \frac{S_1(x)}{S_2(x)} = \lim_{x \to \infty} \frac{f_1(x)}{f_2(x)}. \tag{64} \]

If \(\lim_{x \to \infty} \frac{S_1(x)}{S_2(x)} = \infty\), then we say that \(X_1\) has a **heavier tail** than \(X_2\), based on limiting tail behavior.
Individually, for a random variable $X$ if

$$
\frac{d}{du} (e_X(u)) > 0 \quad (65)
$$

we say that $X$ has a **heavy tail** based on mean excess loss function.

Often, the standard for comparison is the exponential distribution to determine a heavy tail: $X$ is heavier tailed than any exponential distribution if for all $\lambda > 0$,

$$
\lim_{x \to \infty} e^{\lambda x} \mathbb{P}[X > x] = \infty. \quad (66)
$$

**HW:** Please work through

- 3.25
- 3.29
- 3.30
A **Parametric Distribution** is determined by a set of parameters. A subset of set of parametric distributions, known as **Scale Distributions**, is determined by **scale parameters**. In this case, the r.v. $X$ that is linked to the parametric distribution $F_X(x)$ can be transformed by scalar multiplication into $Y = cX$, and the resulting distribution $F_Y(y) = F_X \left( \frac{y}{c} \right)$ is in the same class of distributions.

For example, if $X \sim \exp(\lambda)$ then $F_X = 1 - e^{-\lambda X}$. Then, $F_Y(y) = 1 - e^{-\frac{\lambda}{c}y}$ and so $Y \sim \exp\left(\frac{\lambda}{c}\right)$. 
A random variable $X$ is **k-point mixture** of $(X_1, \ldots, X_k)$ if

$$F_X(x) = a_1 F_{X_1}(x) + \ldots + a_k F_{X_k}(x)$$

$$1 = \sum_{n=1}^{k} a_n.$$  \hspace{1cm} (67)

For example, let $\Theta$ be another random variable, with

$$a_i = \mathbb{P}[\Theta = i]$$

$$1 = \mathbb{P}[X = X_i \mid \Theta = i]$$

$$F_{X_i}(x) = \mathbb{P}[X_i \leq x] = \mathbb{P}[X \leq x \mid \Theta = i].$$  \hspace{1cm} (68)

It follows that

$$\mathbb{P}[X \leq x] = \sum_{i=1}^{k} \mathbb{P}[X \leq x \mid \Theta = i] \mathbb{P}[\Theta = i] = \sum_{i=1}^{k} a_i F_{X_i}(x).$$  \hspace{1cm} (69)
A random variable $X$ is **variable-component mixture** of $(X_1, \ldots, X_N)$ if $N$ can be any positive integer, with

$$F_X(x) = a_1 F_{X_1}(x) + \ldots + a_N F_{X_N}(x)$$

$$1 = \sum_{n=1}^{N} a_n.$$  \hfill (70)

For example, let $\Theta, \tilde{\Theta}$ be random variables, with

$$p = \mathbb{P}[\Theta = 0] = 1 - \mathbb{P}[\Theta = 1]$$

$$p_i = \mathbb{P}[\tilde{\Theta} = i].$$

It follows that

$$\mathbb{P}[X \leq x] = p F_{X_0}(x) + (1 - p) F_{X_1}(x)$$

$$\mathbb{P}[\tilde{X} \leq x] = \sum_{i=1}^{N} a_i F_{X_i}(x).$$

\hfill (72)
A Data-Dependent Distribution is at least as complex as the data that produced it. As the number of data points increases, so does the number of parameters.

An Empirical Model is associated with a discrete model that assigns equal probability $\frac{1}{n}$ to each of the $n$ data points.

HW: 4.1, 4.2, 4.5, 4.7, 4.10.
Let $X$ be a random variable, and $Y = X^\gamma$ its transformation. Then we define

$$Y = \begin{cases} 
\text{Transformed} : \gamma > 0 \\
\text{Inverse} : \gamma = -1 \\
\text{Inverse Transformed} : \gamma < 0, \gamma \neq -1.
\end{cases}$$
Recall our definitions

\[ \Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} \, dt \]

\[ \Gamma(\alpha; x) = \frac{\int_0^x t^{\alpha-1} e^{-t} \, dt}{\Gamma(\alpha)}. \]  \hspace{1cm} (73)

If \( X \sim \text{Gamma}(\alpha) \), the distribution we use is \( \Gamma(x; \alpha) \) to transform \( X \) into \( Y = X^\gamma \).
If $S_X(x) = e^{-x}$, then assuming $\gamma > 0$

\[
Y = \begin{cases} 
\text{Transformed} & : S_Y(y) = e^{-y^\gamma} \\
\text{Inverse Transformed} & : S_Y(y) = e^{-y^{-\gamma}} \\
\text{Weibull} & : S_Y(y) = e^{-\left(\frac{\theta}{y}\right)^\gamma} 
\end{cases}
\]
Define $X | \Lambda$ as a **conditional random variable**, and $S_{X|\Lambda}(x | \lambda) = 1 - F_{X|\Lambda}(x | \lambda)$ as its **conditional survival (probability)** function.

**Theorem**

Define $f_{\Lambda}(\lambda)$ as the density of our parameter $\Lambda$, and $f_X(x)$ as the unconditional density for our random variable $X$. Then

\[
 f_X(x) = \int_{\mathbb{R}} f_{X|\Lambda}(x | \lambda)f_{\Lambda}(\lambda)d\lambda
\]

\[
 F_X(x) = \int_{\mathbb{R}} F_{X|\Lambda}(x | \lambda)f_{\Lambda}(\lambda)d\lambda
\]

\[
 \mu'_k = \mathbb{E}[X^k] = \mathbb{E}\left[\mathbb{E}[X^k | \Lambda]\right]
\]

\[
 \mu_2 = \text{Var}[X] = \mathbb{E}\left[\text{Var}[X | \Lambda]\right] + \text{Var}\left[\mathbb{E}[X | \Lambda]\right].
\]
Mixing

Theorem

If $X \mid \Lambda \sim \exp(\Lambda)$ and $\Lambda \sim \text{Gamma}$, then $X \sim \text{Pareto}$.

Proof.

By our previous theorem,

$$f_X(x) = \int_{\mathbb{R}} f_{X\mid\Lambda}(x \mid \lambda) f_{\Lambda}(\lambda) d\lambda$$

$$= \int_{\lambda=0}^{\infty} \lambda e^{-\lambda x} \frac{\lambda^{\alpha-1} e^{-\theta \lambda} \theta^\alpha}{\Gamma(\alpha)} d\lambda$$

$$= \frac{\theta^\alpha}{\Gamma(\alpha)} \int_{\lambda=0}^{\infty} \lambda^\alpha e^{-\lambda(x+\theta)} d\lambda = \frac{\theta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha + 1)}{(x + \theta)^{\alpha+1}}.$$ (75)
If $X \mid \Lambda \sim N(\Theta, \sigma)$ and $\Theta \sim N(\mu, \alpha)$, then $X \sim N(\mu, \sigma + \alpha)$.

Proof.

By our previous theorem, and by defining $\bar{\sigma} = \frac{\sigma \alpha}{\sigma + \alpha}$ and $\bar{\mu} = \frac{\sigma \mu + \alpha \mu}{\sigma + \alpha}$,

$$f_X(x) = \int_{\mathbb{R}} f_{X \mid \Theta}(x \mid \theta) f_{\Theta}(\theta) d\theta = \int_{\theta=-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\theta)^2}{2\sigma}} \frac{1}{\sqrt{2\pi\alpha}} e^{-\frac{(\theta-\mu)^2}{2\alpha}} d\theta$$

$$= \frac{1}{\sqrt{2\pi(\sigma + \alpha)}} e^{-\frac{(x-\mu)^2}{2(\sigma + \alpha)}} \int_{\theta=-\infty}^{\infty} \frac{1}{\sqrt{2\pi\bar{\sigma}}} e^{-\frac{(\theta-\bar{\mu})^2}{2\bar{\sigma}}} d\theta$$

$$= \frac{1}{\sqrt{2\pi(\sigma + \alpha)}} e^{-\frac{(x-\mu)^2}{2(\sigma + \alpha)}}.$$

(76)
A **Frailty** random variable $\Lambda > 0$ is one where the hazard rate is, for a known function $a(x)$,

$$\mu_{X|\Lambda} = \Lambda \cdot a(x).$$  \hfill (77)

It follows that for $A(x) := \int_0^x a(s)ds$,

$$S_{X|\Lambda}(x | \lambda) = e^{-\lambda \cdot A(x)}$$

$$S_X(x) = \mathbb{E}[e^{-\Lambda \cdot A(x)}] = M_\Lambda[-A(x)]$$  \hfill (78)
Frailty

For example, consider an exponential mixture (i.e. \( A(x) = x \)) and that our parameter is uniformly distributed: \( \Lambda \sim U[0, L] \).

Then

\[
S_X(x) = \int_0^\infty S_{X|\Lambda}(x | \lambda) \cdot f_\Lambda(\lambda) d\lambda
\]

\[
= \int_0^L e^{-\lambda x} \cdot \frac{1}{L} d\lambda
\]

\[
= \frac{1}{Lx} \cdot (1 - e^{-Lx}).
\]

**HW: 5.7.**
A **k-component spliced distribution** has density

\[ f_X(x) = \sum_{n=1}^{k} a_k f_k(x) \chi_{(c_{n-1}, c_n)} \]

(80)

**HW:** Can you splice together an exponential and a Pareto distribution? Give an example if you can!

**HW:**

- 5.1, 5.3, 5.6, 5.9, 5.12
- 5.13, 5.14, 5.20
- *Limiting Distributions:* 5.10
- 5.21, 5.22, 5.26
A random variable $X$ has a **linear exponential** density if

$$f(x; \theta) = \frac{p(x)e^{r(\theta)x}}{q(\theta)}.$$  \hspace{1cm} (81)

For example, if $X \sim N(\theta, \sigma)$, then

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

$$= \left[ \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{x^2}{2\sigma^2}} \right] \left[ e^{\frac{\theta}{\sigma^2}x} \right]$$

\hspace{1cm} (82)
\[ \therefore \ln[f(x; \theta)] = \ln[p(x)] + r(\theta)x - \ln q(\theta) \]

\[ \Rightarrow \frac{\partial f}{\partial \theta} = \left[ r'(\theta)x - \frac{q'(\theta)}{q(\theta)} \right] f(x; \theta) \]
\[
\ln [f(x; \theta)] = \ln [p(x)] + r(\theta)x - \ln q(\theta)
\]

\[
\Rightarrow \frac{\partial f}{\partial \theta} = \left[ r'(\theta)x - \frac{q'(\theta)}{q(\theta)} \right] f(x; \theta)
\]

\[
\Rightarrow 0 = \frac{\partial}{\partial \theta} [1] = \frac{\partial}{\partial \theta} \left[ \int_{x=-\infty}^{x=\infty} f(x; \theta) dx \right]
\]

\[
= \int_{x=-\infty}^{x=\infty} \frac{\partial f}{\partial \theta} dx = \int_{x=-\infty}^{x=\infty} \left[ r'(\theta)x - \frac{q'(\theta)}{q(\theta)} \right] f(x; \theta) dx
\]

\[
= r'(\theta)\mathbb{E}[X] - \frac{q'(\theta)}{q(\theta)}.
\]
\[
\begin{align*}
\therefore \ln [f(x; \theta)] &= \ln [p(x)] + r(\theta)x - \ln q(\theta) \\
\Rightarrow \frac{\partial f}{\partial \theta} &= \left[ r'(\theta)x - \frac{q'(\theta)}{q(\theta)} \right] f(x; \theta) \\
\Rightarrow 0 &= \frac{\partial}{\partial \theta} \left[ \int_{x=-\infty}^{\infty} f(x; \theta) \, dx \right] \\
&= \int_{x=-\infty}^{\infty} \frac{\partial f}{\partial \theta} \, dx = \int_{x=-\infty}^{\infty} \left[ r'(\theta)x - \frac{q'(\theta)}{q(\theta)} \right] f(x; \theta) \, dx \\
&= r'(\theta)\mathbb{E}[X] - \frac{q'(\theta)}{q(\theta)}.
\end{align*}
\]
\[
\Rightarrow \mathbb{E}[X] = \frac{q'(\theta)}{r'(\theta)q(\theta)} =: \mu(\theta)
\]
\[
\begin{align*}
\therefore \ln[f(x; \theta)] &= \ln[p(x)] + r(\theta)x - \ln q(\theta) \\
\Rightarrow \frac{\partial f}{\partial \theta} &= \left[r'(\theta)x - \frac{q'(\theta)}{q(\theta)}\right]f(x; \theta) \\
\Rightarrow 0 &= \frac{\partial}{\partial \theta} \left[1\right] = \frac{\partial}{\partial \theta} \left[\int_{x=-\infty}^{x=\infty} f(x; \theta)dx\right] \\
&= \int_{x=-\infty}^{x=\infty} \frac{\partial f}{\partial \theta} dx = \int_{x=-\infty}^{x=\infty} \left[r'(\theta)x - \frac{q'(\theta)}{q(\theta)}\right] f(x; \theta)dx \\
&= r'(\theta)\mathbb{E}[X] - \frac{q'(\theta)}{q(\theta)}. \\
\Rightarrow \mathbb{E}[X] &= \frac{q'(\theta)}{r'(\theta)q(\theta)} =: \mu(\theta) \\
\Rightarrow \text{Var}[X] &= \frac{\mu'(\theta)}{r'(\theta)} (\text{Proof: Left as an exercise!})
\end{align*}
\]
Let \( N \) be a discrete random variable with

\[
P[N = k] := p_k \geq 0. \tag{84}
\]

It follows that

\[
P(z) = P_N(z) = \mathbb{E}[z^N] := \sum_{k=0}^{\infty} p_k z^k
\]

\[
P^{(m)}(z) = \frac{d^m}{dz^m} \mathbb{E}[z^N] = \mathbb{E} \left[ \frac{d^m}{dz^m}(z^N) \right]
\]

\[
= \mathbb{E} \left[ N(N-1) \cdots (N-m+1) z^{N-m} \right] \tag{85}
\]

\[
= \sum_{k=m}^{\infty} k(k-1) \cdots (k-m+1) z^{k-m} p_k.
\]

\[
\therefore P^{(m)}(0) = m! p_m.
\]
For example, consider the Poisson distribution with $p_k = \frac{e^{-\lambda} \lambda^k}{k!}$ for all $k \geq 0$. We can compute

$$P(z) = P_N(z) = \mathbb{E}[z^N] := \sum_{k=0}^{\infty} \frac{e^{-\lambda} \lambda^k}{k!} z^k$$

$$= \sum_{k=0}^{\infty} \frac{e^{-\lambda} (\lambda z)^k}{k!}$$

$$= e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda z)^k}{k!} = e^{-\lambda} e^{\lambda z}$$

$$= e^{\lambda (z-1)}.$$
Theorem

If \( \{N_j\}_{j=1}^n \) is a sequence of i.i.d. Poisson random variables with respective parameters \( \{\lambda_j\}_{j=1}^n \), then

\[
N := N_1 + \ldots + N_n \sim \text{Poisson}(\lambda_1 + \ldots + \lambda_n).
\] (87)
If \( \{N_j\}_{j=1}^n \) is a sequence of i.i.d. Poisson random variables with respective parameters \( \{\lambda_j\}_{j=1}^n \), then

\[
N := N_1 + \ldots + N_n \sim \text{Poisson}(\lambda_1 + \ldots + \lambda_n).
\]  

(87)

**Proof.**

\[
P_N(z) = \prod_{j=1}^n P_{N_j}(z) = \prod_{j=1}^n e^{\lambda_j(z-1)} = e^{(\sum_{j=1}^n \lambda_j)(z-1)}. 
\]

(88)
Theorem

Assume \( \Lambda \) has a density \( f_\Lambda(\lambda) \), and \( \mathbb{P}[N = k \mid \Lambda = \lambda] = \frac{e^{-\lambda} \lambda^k}{k!} \).

Then

\[ p_k = \mathbb{P}[N = k] = \int_0^\infty \frac{e^{-\lambda} \lambda^k}{k!} f_\Lambda(\lambda) d\lambda = \frac{1}{k!} \mathbb{E}[\Lambda^k e^{-\Lambda}] \tag{89} \]

For example, if \( f_\Lambda(\lambda) = \frac{\lambda^{\alpha-1} e^{-\frac{\lambda}{\theta}}}{\theta^\alpha \Gamma(\alpha)} \), then

\[ p_k = \frac{1}{k!} \int_0^\infty e^{-\lambda} \lambda^k \frac{\lambda^{\alpha-1} e^{-\frac{\lambda}{\theta}}}{\theta^\alpha \Gamma(\alpha)} d\lambda = \frac{1}{k!} \frac{1}{\theta^\alpha \Gamma(\alpha)} \int_0^\infty e^{-\lambda(1+\frac{1}{\theta})} \lambda^{k+\alpha-1} d\lambda \]

\[ = \frac{\Gamma(k + \alpha) \theta^k}{k! \Gamma(\alpha)} \frac{\theta^k}{(1 + \theta)^{k+\alpha}} \tag{90} \]
Theorem

Assume $\Lambda$ has a density $f_\Lambda(\lambda)$, and $P[N = k \mid \Lambda = \lambda] = \frac{e^{-\lambda} \lambda^k}{k!}$.

Then

$$p_k = P[N = k] = \int_0^\infty \frac{e^{-\lambda} \lambda^k}{k!} f_\Lambda(\lambda) d\lambda = \frac{1}{k!} \mathbb{E}[\Lambda^k e^{-\Lambda}]. \quad (89)$$

For example, if $f_\Lambda(\lambda) = \frac{\lambda^{\alpha-1} e^{-\frac{\lambda}{\theta}}}{\theta^\alpha \Gamma(\alpha)}$, then

$$p_k = \frac{1}{k!} \int_0^\infty e^{-\lambda} \lambda^k \frac{\lambda^{\alpha-1} e^{-\frac{\lambda}{\theta}}}{\theta^\alpha \Gamma(\alpha)} d\lambda = \frac{1}{k!} \frac{1}{\theta^\alpha \Gamma(\alpha)} \int_0^\infty e^{-\lambda (1 + \frac{1}{\theta})} \lambda^{k + \alpha - 1} d\lambda$$

$$= \frac{\Gamma(k + \alpha)}{k! \Gamma(\alpha)} \frac{\theta^k}{(1 + \theta)^{k + \alpha}}. \quad (90)$$

HW: Compute $p_k$ if $\Lambda \sim \text{Exp}(\lambda)$. 
Define $X$ and $N$ such that for a sequence $\{X_j\}_{j=1}^n$ with each of the $X_j \sim X$ and

$$p = \mathbb{P}[X = 1] = 1 - \mathbb{P}[X = 0]$$

$$N = N_1 + \ldots + N_n.$$  \hspace{1cm} (91)

It follows that

$$P_X(z) = (1 - p)z^0 + pz^1 = 1 - p + pz$$
Binomial Distributions Revisited

Define $X$ and $N$ such that for a sequence $\{X_j\}_{j=1}^n$ with each of the $X_j \sim X$ and

$$
p = \mathbb{P}[X = 1] = 1 - \mathbb{P}[X = 0] \\
N = N_1 + \ldots + N_n.
$$

(91)

It follows that

$$
P_X(z) = (1 - p)z^0 + pz^1 = 1 - p + pz \\
\Rightarrow P_N(z) = [1 - p + pz]^n
$$

We can also show that

$$
E[N] = np \\
\text{Var}[N] = np(1 - p).
$$

(93)
Define $X$ and $N$ such that for a sequence $\{X_j\}_{j=1}^n$ with each of the $X_j \sim X$ and

$$p = \Pr[X = 1] = 1 - \Pr[X = 0]$$

$$N = N_1 + \ldots + N_n.$$  

(91)

It follows that

$$P_X(z) = (1 - p)z^0 + pz^1 = 1 - p + pz$$

$$\Rightarrow P_N(z) = [1 - p + pz]^n$$

(92)

$$\Rightarrow \Pr[N = k] = \frac{m!}{(m - k)!k!} p^k (1 - p)^{m-k}.$$  

We can also show that

$$\mathbb{E}[N] = np$$

$$\text{Var}[N] = np(1 - p).$$  

(93)
A discrete random variable \( N \) is a member of the \((a, b, 0)\) class if, for \( p_k = \Pr[N = k] \) and all \( k \in \{1, 2, 3\ldots\} \),

\[
\frac{p_k}{p_{k-1}} = a + \frac{b}{k}.
\]  \hspace{1cm} (94)

It follows that

\[
kp_k = akp_{k-1} + bp_{k-1} = a(k-1)p_{k-1} + (a+b)p_{k-1}
\]
A discrete random variable $N$ is a **member of the $(a, b, 0)$ class** if, for $p_k = P[N = k]$ and all $k \in \{1, 2, 3\ldots\}$,

$$\frac{p_k}{p_{k-1}} = a + \frac{b}{k}. \quad (94)$$

It follows that

$$kp_k = akp_{k-1} + bp_{k-1} = a(k - 1)p_{k-1} + (a + b)p_{k-1}$$

$$\Rightarrow \sum_{k=1}^{\infty} kp_k = a \sum_{k=1}^{\infty} (k - 1)p_{k-1} + (a + b) \sum_{k=1}^{\infty} p_{k-1}$$
A discrete random variable $N$ is a member of the $(a, b, 0)$ class if, for $p_k = \mathbb{P}[N = k]$ and all $k \in \{1, 2, 3\ldots\}$,

$$\frac{p_k}{p_{k-1}} = a + \frac{b}{k}.$$  \hspace{1cm} (94)

It follows that

$$kp_k = akp_{k-1} + bp_{k-1} = a(k - 1)p_{k-1} + (a + b)p_{k-1}$$

$$\Rightarrow \sum_{k=1}^{\infty} kp_k = a \sum_{k=1}^{\infty} (k - 1)p_{k-1} + (a + b) \sum_{k=1}^{\infty} p_{k-1}$$  \hspace{1cm} (95)

$$\therefore \mathbb{E}[N] = a\mathbb{E}[N] + (a + b) \cdot 1 = \frac{a + b}{1 - a}.$$  

**HW:** Ex. 6.1, 6.2.
By definition, if we take successive probabilities (frequencies) of observed data, and we know that the data is drawn from an \((a, b, 0)\)-class distribution, then it must be that for \(y_k := k \frac{p_k}{p_{k-1}}\),

\[
y_k = ak + b. \tag{96}
\]

For example, if we observe \(f\) total data points, and \(f_k\) are the total number of observed points with property \(k\), then

\[
\tilde{p}_k = \frac{f_k}{f} \Rightarrow \tilde{y}_k = k \frac{f_k}{f_{k-1}}. \tag{97}
\]

To fit, plot \(\tilde{y}_k\) against \(k\), and fit for \(a\) and \(b\).

**HW:** Read on the \((a, b, 1)\) class, read pages 90 – 94, and complete Ex. 6.4, 6.6.
According to results by **Sundt and Jewell**, and extended in **Hess, Liewald, and Schmidt**, the three distributions in this class are

<table>
<thead>
<tr>
<th>Distribution</th>
<th>a</th>
<th>b</th>
<th>( p_0 )</th>
<th>( p_k )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poisson</td>
<td>0</td>
<td>( \lambda )</td>
<td>( e^{-\lambda} )</td>
<td>( e^{-\lambda} \frac{\lambda^k}{k!} )</td>
</tr>
<tr>
<td>Negative Binomial</td>
<td>( 1 - p )</td>
<td>( (1 - p)(r - 1) )</td>
<td>( p^r )</td>
<td>( \frac{\Gamma(r+k)}{k!\Gamma(r)} p^r (1 - p)^k )</td>
</tr>
<tr>
<td>Binomial</td>
<td>( -\frac{p}{1-p} )</td>
<td>( (n + 1)\frac{p}{1-p} )</td>
<td>( (1 - p)^n )</td>
<td>( \binom{n}{k} p^k (1 - p)^{n-k} )</td>
</tr>
</tbody>
</table>

We can use this table to match empirically observed data with an appropriate distribution.
Let \( N \) be a member of \( C(a, b, 0) \) satisfying the recursive probabilities

\[
k \frac{p_k}{p_{k-1}} = \frac{3}{4} k + 3. \tag{98}
\]

Identify the distribution \( N \).
Let $N$ be a member of $C(a, b, 0)$ satisfying the recursive probabilities

$$
k \frac{p_k}{p_{k-1}} = \frac{3}{4} k + 3. \tag{98}$$

Identify the distribution $N$.

**Answer:**

As $a > 0$, $b > 0$, it follows that $N$ is a negative binomial distribution. Furthermore,

$$\frac{3}{4} = a = 1 - p$$

$$3 = b = (1 - p)(r - 1) = \frac{3}{4} (r - 1) \tag{99}$$

$$\Rightarrow (p, r) = (0.25, 5).$$
The distribution of accidents for 84 randomly selected policies is as follows:

<table>
<thead>
<tr>
<th>Number of Accidents</th>
<th>Number of Policies</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>32</td>
</tr>
<tr>
<td>1</td>
<td>26</td>
</tr>
<tr>
<td>2</td>
<td>12</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
</tr>
</tbody>
</table>

Identify the frequency model that best represents these data.
The distribution of accidents for 84 randomly selected policies, and their adjusted frequency ratios, are as follows:

<table>
<thead>
<tr>
<th>Number of Accidents</th>
<th>Number of Policies</th>
<th>$k \frac{p_k}{p_{k-1}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>26</td>
<td>0.8125</td>
</tr>
<tr>
<td>2</td>
<td>12</td>
<td>0.9231</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>1.75</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>2.2857</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>2.5</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>3</td>
</tr>
</tbody>
</table>
The distribution of accidents for 84 randomly selected policies, and their adjusted frequency ratios, are as follows:

<table>
<thead>
<tr>
<th>Number of Accidents</th>
<th>Number of Policies</th>
<th>$k \frac{p_k}{p_{k-1}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>26</td>
<td>0.8125</td>
</tr>
<tr>
<td>2</td>
<td>12</td>
<td>0.9231</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>1.75</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>2.2857</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>2.5</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>3</td>
</tr>
</tbody>
</table>

Best fit line is $0.462969k + 0.25816$, hence a negative binomial distribution. Check out Wolfram Alpha’s computation by entering "linear regression $(1, 0.8125), (2, 0.9231), (3, 1.75), (4, 2.2857), (5, 2.5), (6, 3)$" in the search bar.
Consider now the distribution of accidents for randomly selected policies is as follows:

<table>
<thead>
<tr>
<th>Number of Accidents</th>
<th>Number of Policies</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>400</td>
</tr>
<tr>
<td>1</td>
<td>200</td>
</tr>
<tr>
<td>2</td>
<td>50</td>
</tr>
<tr>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
</tr>
</tbody>
</table>

Identify the frequency model that best represents these data.
The distribution of accidents for 84 randomly selected policies, and their adjusted frequency ratios, are as follows:

<table>
<thead>
<tr>
<th>Number of Accidents</th>
<th>Number of Policies</th>
<th>$k \frac{p_k}{p_{k-1}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>200</td>
<td>1 \cdot \frac{200}{400}</td>
</tr>
<tr>
<td>2</td>
<td>50</td>
<td>2 \cdot \frac{50}{200}</td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>3 \cdot \frac{8}{50}</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>4 \cdot \frac{1}{8}</td>
</tr>
</tbody>
</table>
The distribution of accidents for 84 randomly selected policies, and their adjusted frequency ratios, are as follows:

<table>
<thead>
<tr>
<th>Number of Accidents</th>
<th>Number of Policies</th>
<th>$k \frac{p_k}{p_{k-1}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>200</td>
<td>1 \cdot \frac{200}{400}</td>
</tr>
<tr>
<td>2</td>
<td>50</td>
<td>2 \cdot \frac{50}{200}</td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>3 \cdot \frac{8}{50}</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>4 \cdot \frac{1}{8}</td>
</tr>
</tbody>
</table>

Best fit line is $-0.002k + 0.5$. Is this a Binomial or Poisson distribution? Check out Wolfram Alpha’s computation by entering ”linear regression $(1, 1 \cdot \frac{200}{400}), (2, 2 \cdot \frac{50}{200}), (3, 3 \cdot \frac{8}{50}), (4, 4 \cdot \frac{1}{8})$” in the search bar.
A bank, insurance company, hedge fund, or pension sponsor among other entities is charged with navigating the risks borne by their endeavor. In order to mitigate and price such risks, one must first set out to identify their various forms.
Multiple Exposures to Risk

A few known variants include

- Market Risk.
- Interest Rate Risk.
- FX Risk.
- Liquidity Risk.
- Credit Risk.
Market Risk

- Exposure to capital markets.
- Can affect both assets and liabilities.
- Examples include equities and other financial instruments.
Interest Rate Risk

- Exposure to changes in short rate.
- Underestimating volatility of short rate evolution (shocks to economy.)
- Exposure to changes in yield curve and term structure.
Inability to predict or affect foreign government domestic policy (prime rate) and foreign policy (devaluation.)

Can lead to inability to properly hedge financial instruments held in foreign countries.

Foreign interest rate fluctuation can lead to possible shortfalls that banks may wish to immunize their portfolio against. Currency fluctuation also leads to daily operational risk, and should factor into planning.

For example, should a band tour Europe or America this summer? Click here for an insightful article by Neil Shah in the Wall Street Journal™, with comment from the manager of a very prominent rock band.
Liquidity Risk

- Exposure to depressed prices in return for expedited sale (liquidation) of assets.
- Also affects strategies that hedge long term contracts with short term assets.
- Banks are also prone to this in the form of *bank runs*.
Credit Risk

- This also falls under **Default Risk**, where one is exposed to (counterparty) default on contracts that exchange cash flows.
- Correlated with Market Risk, although different opinions on how so.
- For life insurance companies, a life \((x)\) can be seen as a default risk if that person outlives the present value of their retirement annuity. Also known as **Longevity (Mortality) Risk**.
Credit Risk

- This also falls under **Default Risk**, where one is exposed to (counterparty) default on contracts that exchange cash flows.
- Correlated with Market Risk, although different opinions on how so.
- For life insurance companies, a life \( x \) can be seen as a default risk if that person outlives the present value of their retirement annuity. Also known as **Longevity (Mortality) Risk**.
  - This happens, for example, when pension sponsors overestimate the mortality of the insured population. Reinsurance can be bought, but such reinsurers are also susceptible to the previously defined risks, depending on their asset and liability portfolio.
  - As an example, please refer to the paper by Tsai, Tzeng, and Wang on *Hedging Longevity Risk When Interest Rates Are Uncertain*. 
Credit Risk

- Complicated to measure and perhaps even define:

\[
\text{Risk} = \mathbf{P} [\text{Default}] \quad \text{or} \quad \mathbf{E} [\text{Loss} | \text{Default}] = \mathbf{E} [\text{Loss}] \times \mathbf{P} [\text{Default}]
\]

However, does \( \mathbf{E} [\text{Loss} | \text{Default}] \) relate in some fashion to \( \mathbf{P} [\text{Default}] \)? In an adverse economy mired in recession, perhaps so. In fact, Frye in his paper “Collateral Damage” (2000) argues that the same factors that increase default rates can also decrease the value of loan collateral.

Click here for a review of Structural Models for Credit Risk published by the SOA.

Clearly, regulators, counterparties, clearinghouses, and shareholders would like a precise definition of the risk involved with an entity’s business. One possible definition is the capital a business entity should hold. So what to do?
Credit Risk

- Complicated to measure and perhaps even define:
  - Risk = \( P[\text{Default}] \)?
  - Risk = \( E[\text{Loss} \mid \text{Default}] \)?

However, does \( E[\text{Loss} \mid \text{Default}] \) relate in some fashion to \( P[\text{Default}] \)? In an adverse economy mired in recession, perhaps so. In fact, Frye in his paper “Collateral Damage” (2000) argues that the same factors that increase default rates can also decrease the value of loan collateral.

Click here for a review of Structural Models for Credit Risk published by the SOA.

Clearly, regulators, counterparties, clearinghouses, and shareholders would like a precise definition of the risk involved with an entity’s business. One possible definition is the capital a business entity should hold. So what to do?
Credit Risk

- Complicated to measure and perhaps even define:
  - Risk = \( P[\text{Default}] \)?
  - Risk = \( E[\text{Loss} \mid \text{Default}] \)?
  - or...Risk = \( E[\text{Loss}] = E[\text{Loss} \mid \text{Default}] \times P[\text{Default}] \)?

However, does \( E[\text{Loss} \mid \text{Default}] \) relate in some fashion to \( P[\text{Default}] \)? In an adverse economy mired in recession, perhaps so. In fact, Frye in his paper "Collateral Damage" (2000) argues that the same factors that increase default rates can also decrease the value of loan collateral.

Click here for a review of Structural Models for Credit Risk published by the SOA.

Clearly, regulators, counterparties, clearinghouses, and shareholders would like a precise definition of the risk involved with an entity's business. One possible definition is the capital a business entity should hold. So what to do?
Credit Risk

- Complicated to measure and perhaps even define:
  - Risk = $P[\text{Default}]$?
  - Risk = $E[\text{Loss} \mid \text{Default}]$?
  - or...Risk = $E[\text{Loss}] = E[\text{Loss} \mid \text{Default}] \times P[\text{Default}]$?
  - However, does $E[\text{Loss} \mid \text{Default}]$ relate in some fashion to $P[\text{Default}]$?
Credit Risk

- Complicated to measure and perhaps even define:
  - Risk = $P[\text{Default}]$?
  - Risk = $\mathbb{E}[	ext{Loss} \mid \text{Default}]$?
  - or...Risk = $\mathbb{E}[	ext{Loss}] = \mathbb{E}[	ext{Loss} \mid \text{Default}] \times P[\text{Default}]$?
  - However, does $\mathbb{E}[	ext{Loss} \mid \text{Default}]$ relate in some fashion to $P[\text{Default}]$?
  - In an adverse economy mired in recession, perhaps so. In fact, Frye in his paper "Collateral Damage" (2000) argues that the same factors that increase default rates can also decrease the value of loan collateral.
  - Click here for a review of Structural Models for Credit Risk published by the SOA.
Credit Risk

- Complicated to measure and perhaps even define:
  - Risk = \( P[\text{Default}] \)?
  - Risk = \( E[\text{Loss} | \text{Default}] \)?
  - or...Risk = \( E[\text{Loss}] = E[\text{Loss} | \text{Default}] \times P[\text{Default}] \)?
  - However, does \( E[\text{Loss} | \text{Default}] \) relate in some fashion to \( P[\text{Default}] \)?
  - In an adverse economy mired in recession, perhaps so. In fact, Frye in his paper “Collateral Damage” (2000) argues that the same factors that increase default rates can also decrease the value of loan collateral.
  - Click here for a review of Structural Models for Credit Risk published by the SOA.

Clearly, regulators, counterparties, clearinghouses, and shareholders would like a precise definition of the risk involved with an entity’s business. One possible definition is the capital a business entity should hold. So what to do?

Fix a probability space \((\Omega, \mathcal{F}, \mathbb{P})\) and a standard Brownian motion \(W\) that lives on this space.

Consider now a term contact with term \(T\) and let \(\alpha\) denote the management charges factor along with \(\beta\) representing the policyholder’s participation factor.

Furthermore, assume mean and standard deviation parameters \((\mu, \sigma)\) respectively and the corresponding Geometric Brownian Mutual Fund Asset

\[
S_t = S_0 e^{\mu t + \sigma W_t}. \tag{100}
\]
Using this as the model of the asset returns upon which premiums are invested, the policyholder wishes to purchase a contract that pays a **maturity benefit** credited at a rate of return which is the greater of

- the customer’s **risk discount rate** $r$, where $r < \mu$ or
- the participation rate of the stock index returns of $S$.

Symbolically, for a current premium $P$ invested in the , the contract payout value at maturity is

\[
V(T) = (1 - \alpha)P \max \left\{ e^{rT}, 1 + \beta \left( \frac{S_T}{S_0} - 1 \right) \right\}. \quad (101)
\]
Assume that the policyholder is able to fully participate in the returns from the fund (i.e. $\beta = 1$.) Then

$$V(T) = (1 - \alpha)P \frac{S_T}{S_0} + (1 - \alpha)P \max \left\{ e^{rT} - \left( \frac{S_T}{S_0} \right), 0 \right\}$$

$$:= V_1(T) + V_2(T).$$

Here, $V_1(T)$ is the *net premium*, or payoff, for investing in the index fund and $V_2(T)$ is the *guaranteed option* payoff if the index fund under-performs relative to the risk discount rate $r$.

How does one reserve to meet the obligations of $V_2(T)$. 
One can see that the probability of a payout, that $V_2(T) \neq 0$ is for large $T$

$$\mathbb{P}[V_2(T) \neq 0] = \mathbb{P}[rT > \mu T + \sigma W_T] = \Phi\left(\frac{r - \mu}{\sigma} \sqrt{T}\right) \approx 0. \quad (103)$$
One can see that the probability of a payout, that $V_2(T) \neq 0$ is for large $T$

$$\mathbb{P}[V_2(T) \neq 0] = \mathbb{P}[rT > \mu T + \sigma W_T] = \Phi\left(\frac{r - \mu}{\sigma \sqrt{T}}\right) \approx 0. \quad (103)$$

Since it is a low probability event that we have to prepare for a payout $V_2(T)$ and since we can directly replicate the payoff $V_1(T)$ by initially purchasing $\frac{(1-\alpha)P}{S_0}$ units of the index fund, an actuary may be tempted to not reserve for the uncertain portion of the guarantee, $V_2(T)$, if the contract has a relatively long term $T$. 
One can see that the probability of a payout, that $V_2(T) \neq 0$ is for large $T$

$$\mathbb{P}[V_2(T) \neq 0] = \mathbb{P}[rT > \mu T + \sigma W_T] = \Phi\left(\frac{r - \mu}{\sigma} \sqrt{T}\right) \approx 0. \quad (103)$$

Since it is a low probability event that we have to prepare for a payout $V_2(T)$ and since we can directly replicate the payoff $V_1(T)$ by initially purchasing $\left(1 - \alpha\right)P_{S_0}$ units of the index fund, an actuary may be tempted to not reserve for the uncertain portion of the guarantee, $V_2(T)$, if the contract has a relatively long term $T$. Is this a wise decision?
Return Measures

A bank, insurance company, hedge fund, or pension sponsor can have their efficiency measured by how they return profits over a time horizon. For example, if the value of a portfolio at time $t$ is denoted by $X(t)$, then returns are measured at times $\{t_0, t_1, t_2, ..., t_n\}$ and the **Loss** in period $(t_{k-1}, t_k)$ is

$$L[X](t_k) := -(X(t_k) - X(t_{k-1})) \quad (104)$$

and the **Loss Distribution** is

$$r[X](t_k) := -\ln \frac{X(t_k)}{X(t_{k-1})} \quad (105)$$
Comparing to a benchmark $r_B(t)$ for returns, a manager can be rated via

$$RoB := \frac{1}{t_n - t_0} \sum_{k=1}^{n} (r[X](t_k) - r_B(t_k))$$ (106)

as well as the deviation away from the benchmark returns:

$$Dev := \sqrt{\frac{1}{t_n - t_0} \sum_{k=1}^{n} (r[X](t_k) - r_B(t_k))^2}$$ (107)

Notice that RoB is a \textit{linear} measure.
Recall from (financial) economics that there is a measure of return vs risk, known as the **Sharpe Ratio**. This measures the reward gained versus the risk ventured to return the gains won.
Recall from (financial) economics that there is a measure of return vs risk, known as the **Sharpe Ratio**. This measures the reward gained versus the risk ventured to return the gains won. If $\alpha$ is the realized return on an asset over a *risk free rate* $r$, then for a standard deviation $\sigma$ of the returns on the asset we have the informational measure

\[
\text{Sharpe Ratio} := \frac{\alpha - r}{\sigma}
\]  

(108)
Sharpe Ratio

Going back to our periodic returns, we can enact a similar measure of performance via

\[
S := \frac{1}{t_n - t_0} \sum_{k=1}^{n} (r[X](t_k) - r_B(t_k)) \sqrt{\frac{1}{t_n - t_0} \sum_{k=1}^{n} (r[X](t_k) - r_B(t_k))^2}
\]  

(109)

The numerical values of \( S \) can be used to rank manager performance, but as with any models there are criticisms. First off is the usefulness of the denominator as a risk measure.
Define a probability space \((\Omega, \mathcal{F}, \mathbb{P})\). Define the set \(\mathcal{A} \subset L^0(\Omega, \mathcal{F}, \mathbb{P})\) of possible risks, where

- \(X, Y \in \mathcal{A} \Rightarrow X + Y \in \mathcal{A}\).
- \(X \in \mathcal{A} \Rightarrow \alpha X \in \mathcal{A}\) for all \(\alpha \in [0, 1]\).
- \(X \in \mathcal{A} \Rightarrow X + a \in \mathcal{A}\) for all \(a \in \mathbb{R}\).
A risk measure $\rho : \mathcal{A} \rightarrow \mathbb{R} \cup \{\infty\}$ may be required to satisfy some of the following properties:

- **Law Invariance:** $\forall X, Y \in \mathcal{A}, x \in \mathbb{R}, \mathbb{P}[X \leq x] = \mathbb{P}[Y \leq x] \Rightarrow \rho[X] = \rho[Y]$.

- **Monotonicity:** $\forall X, Y \in \mathcal{A}, \mathbb{P}[X \leq Y] = 1 \Rightarrow \rho[X] \leq \rho[Y]$.

- **Subadditivity:** $\forall X, Y \in \mathcal{A}, \rho[X + Y] \leq \rho[X] + \rho[Y]$.

- **Positive Homogeneity:** $\forall X \in \mathcal{A}, \alpha \geq 0, \rho[\alpha X] = \alpha \rho[X]$.

- **Translation Invariance:** $\forall X \in \mathcal{A}, a \in \mathbb{R}, \rho[X + a] = \rho[X] + a$.

- **Convexity:** $\forall X, Y \in \mathcal{A}, \lambda \in [0, 1], \rho[\lambda X + (1 - \lambda) Y] \leq \lambda \rho[X] + (1 - \lambda) \rho[Y]$.

A risk measure that satisfies Monotonicity, Subadditivity, Positive Homogeneity, and Translation Invariance is said to be **Coherent**.
A risk measure \( \rho : \mathcal{A} \rightarrow \mathbb{R} \cup \{\infty\} \) may be required to satisfy some of the following properties:

- **Law Invariance:** \( \forall X, Y \in \mathcal{A}, x \in \mathbb{R}, \mathbb{P}[X \leq x] = \mathbb{P}[Y \leq x] \Rightarrow \rho[X] = \rho[Y] \).

- **Monotonicity:** \( \forall X, Y \in \mathcal{A}, \mathbb{P}[X \leq Y] = 1 \Rightarrow \rho[X] \leq \rho[Y] \).

- **Subadditivity:** \( \forall X, Y \in \mathcal{A}, \rho[X + Y] \leq \rho[X] + \rho[Y] \).

- **Positive Homogeneity:** \( \forall X \in \mathcal{A}, \alpha \geq 0, \rho[\alpha X] = \alpha \rho[X] \).

- **Translation Invariance:** \( \forall X \in \mathcal{A}, a \in \mathbb{R}, \rho[X + a] = \rho[X] + a \).

- **Convexity:** \( \forall X, Y \in \mathcal{A}, \lambda \in [0, 1], \rho[\lambda X + (1 - \lambda)Y] \leq \lambda \rho[X] + (1 - \lambda) \rho[Y] \).

A risk measure that satisfies Monotonicity, Subadditivity, Positive Homogeneity, and Translation Invariance is said to be **Coherent**.
Note that Translation Invariance ⇒ \( \rho \left[ X - \rho[X] \right] = 0 \).

In the insurance field, \( \rho[X] \) is seen as the **solvency capital** required by a regulating agency for a company exposed to loss (risk) \( X \).

The capital is a requirement imposed on the firm so that the surplus of assets over liabilities for the company is at least \( \rho[X] \).

\( \rho[X] \) should be chosen to ensure that \( \{ X > \rho[X] \} \) is a sufficiently rare event.

Regulators should, and do, take into consideration the costs incurred by a company that is required to hold excessive solvency capital. Hence, a balance must be sought between solvency and onerous capital requirements. The choice of \( \rho \) should be decided upon with these thoughts in mind.
Companies or individuals who wish to exchange a future liability for an up-front payment seek out willing partners. They find them in the insurance industry.

However, insurance firms pricing the risk of being liable for a future payment is coupled must also meet the demands of regulators to remain solvent.

One of the most common methods is the Expected Value Premium Principle (EPP) which returns $\rho[X] = \mathbb{E}[X]$. 
Companies or individuals who wish to exchange a future liability for an up-front payment seek out willing partners. They find them in the insurance industry.

However, insurance firms pricing the risk of being liable for a future payment is coupled must also meet the demands of regulators to remain solvent.

One of the most common methods is the **Expected Value Premium Principle** (EPP) which returns \( \rho[X] = \mathbb{E}[X] \).

However, such a principle doesn’t take into account possible (large) deviations from the average loss. Some corrections:
- **EPP** with loading: $\rho[X] = (1 + \alpha)\mathbb{E}[X]$ for some $\alpha > 0$.

- **Standard Deviation Premium Principle**: $\rho[X] = \mathbb{E}[X] + \alpha \sqrt{\mathbb{V}[X]}$ for some $\alpha \geq 0$.

- **Variance Premium Principle**: $\rho[X] = \mathbb{E}[X] + \alpha \mathbb{V}[X]$ for some $\alpha \geq 0$.

- **Dutch Premium Principle**: $\rho[X] = \mathbb{E}[X] + \theta \mathbb{E}[(X - \alpha \mathbb{E}[X])^+]$ for some $\alpha \geq 1, \theta \in [0, 1]$.
Let $X \in L^0(\Omega, \mathcal{F}, \mathbb{P})$ and $\alpha \in [0, 1]$. Define

$$\text{VaR}_\alpha[X] := Q_\alpha[X] = \inf \{x \mid \mathbb{P}[X \leq x] \geq \alpha\}$$

(110)

The VaR is the amount of cash a company requires to set the probability of the company becoming insolvent equal to $\alpha$. The Basel Committee on Banking has in the past set VaR as a measure to set capital requirements. VaR is also used in the **Portfolio Premium Pricing Principle** in actuarial settings.
Example: Value at Risk

Consider now the distribution of **Losses** for accident claims from randomly selected policies is as follows:

<table>
<thead>
<tr>
<th>Claim Amount L</th>
<th>Number of Policies</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>400</td>
<td>0.607</td>
</tr>
<tr>
<td>100</td>
<td>200</td>
<td>0.303</td>
</tr>
<tr>
<td>200</td>
<td>50</td>
<td>0.076</td>
</tr>
<tr>
<td>500</td>
<td>8</td>
<td>0.012</td>
</tr>
<tr>
<td>1000</td>
<td>1</td>
<td>0.002</td>
</tr>
</tbody>
</table>

Compute the Value at Risk for the 90%, 95% and 99% levels.
Consider now the distribution of **Losses** for accident claims from randomly selected policies is as follows:

<table>
<thead>
<tr>
<th>Claim Amount $L$</th>
<th>$\mathbb{P}[L \leq x]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000</td>
<td>1</td>
</tr>
<tr>
<td>500</td>
<td>0.998</td>
</tr>
<tr>
<td>200</td>
<td>0.986</td>
</tr>
<tr>
<td>100</td>
<td>0.910</td>
</tr>
<tr>
<td>0</td>
<td>0.607</td>
</tr>
</tbody>
</table>

It follows that

$$( \text{VaR}_{0.90}[L], \text{VaR}_{0.95}[L], \text{VaR}_{0.99}[L]) = (100, 200, 500).$$

(111)
Why Choose VaR as a Risk Measure?

Consider that a firm is allowed to penalize for the cost of capital by a pre-determined factor $0 < \epsilon \leq 1$. 

Define the (penalized) Expected Shortfall via

$$ES(\rho[X]) := E\left[(X - \rho[X])^+ + \epsilon \rho[X]\right] \quad (112)$$

Question: What capital $u = \rho[X]$ level solves $\min_{u \in \mathbb{R}^+} ES(\epsilon)$? (113)
Consider that a firm is allowed to penalize for the cost of capital by a pre-determined factor $0 < \epsilon \leq 1$.

Define the (penalized) **Expected Shortfall** via

$$ES_\epsilon(\rho[X]) := \mathbb{E}\left[(X - \rho[X])_+\right] + \epsilon \rho[X]$$  \hspace{1cm} (112)
Consider that a firm is allowed to penalize for the cost of capital by a pre-determined factor $0 < \epsilon \leq 1$.

Define the (penalized) Expected Shortfall via

$$ES_\epsilon(\rho[X]) := \mathbb{E}[ (X - \rho[X])_+ ] + \epsilon \rho[X]$$  \hspace{1cm} (112)

**Question**: What capital $u = \rho[X]$ level solves

$$\min_{u \in \mathbb{R}^+} ES_\epsilon(u)?$$  \hspace{1cm} (113)
Theorem

The smallest capital $\rho[X]$ that solves

$$\min_{u \in \mathbb{R}_+} ES_\epsilon(u)$$

is given by $\rho[X] = Q_{1-\epsilon}[X]$.

There is a nice geometric proof for this, given by the authors.
(Sketch:) Assume that $X$ has a smooth distribution. The critical point is obtained via

$$0 = \frac{d}{du} \left( \mathbb{E}[(X - u)_+] + \epsilon u \right)$$

$$= \frac{d}{du} \left( \int_0^\infty \mathbb{P}[(X - u)_+ > x] dx + \epsilon u \right)$$

$$= \frac{d}{du} \left( \int_u^\infty \mathbb{P}[X > x] dx + \epsilon u \right)$$

$$= \frac{d}{du} \left( \int_u^\infty (1 - F_X(x)) dx + \epsilon u \right)$$

$$= F_X(u) - 1 + \epsilon$$
Proof.

(Sketch:) Assume that $X$ has a smooth distribution. The critical point is obtained via

$$0 = \frac{d}{du} \left( \mathbb{E}[(X - u)_+] + \epsilon u \right)$$

$$= \frac{d}{du} \left( \int_0^\infty \mathbb{P}[(X - u)_+ > x] \, dx + \epsilon u \right)$$

$$= \frac{d}{du} \left( \int_u^\infty \mathbb{P}[X > x] \, dx + \epsilon u \right)$$

$$= \frac{d}{du} \left( \int_u^\infty (1 - F_X(x)) \, dx + \epsilon u \right)$$

$$= F_X(u) - 1 + \epsilon$$

$$\rightarrow \rho[X] = u = F_X^{-1}(1 - \epsilon) = Q_{1-\epsilon}[X].$$
Note that \( \text{VaR}_\alpha \left[ X - \text{VaR}_\alpha [X] \right] = 0 \)

This implies that the risk of holding an asset with uncertain outcome \( X \) can be neutralized by holding a reserve amount \( \text{VaR}_\alpha [X] \).

But VaR doesn’t account for how big the loss for the company in the event of insolvency.

And regarding diversification..See Example 3.14 in our textbook and example on page 19 of Hardy Study Note
From the *European Commission FAQ on Solvency II*:

- ”The aim of a solvency regime is to ensure the financial soundness of insurance undertakings, and in particular to ensure that they can survive difficult periods. This is to protect policyholders (consumers, businesses) and the stability of the financial system as a whole. ”

- ”..insurers must have available resources sufficient to cover both a Minimum Capital Requirement (MCR) and a Solvency Capital Requirement (SCR). The SCR is based on a Value-at-Risk measure calibrated to a 99.5% confidence level over a 1-year time horizon. The SCR covers all risks that an insurer faces (e.g. insurance, market, credit and operational risk) and will take full account of any risk mitigation techniques applied by the insurer (e.g. reinsurance and securitisation) .. ”
Some other (related) Risk Measures

The quantile measure VaR doesn’t give us information on the tail behavior of $X$ beyond $Q_p[X]$. One way to account for this is to define

$$TVaR_p[X] := \frac{1}{1 - p} \int_p^1 Q_q[X] dq.$$  \hspace{1cm} (116)

Furthermore, in the tail region, we can also determine how much the loss $X$ exceeds $Q_p[X]$, on average, by

$$CTE_p[X] := \mathbb{E}[X \mid X > Q_p[X]].$$  \hspace{1cm} (117)

Finally, regulators are interested in the average shortfall

$$ESF_p[X] := \mathbb{E}[(X - Q_p[X])_+] .$$  \hspace{1cm} (118)
Equivalence of Risk Measures [DGK (2006)]

Theorem

\[ TVaR_p[X] = Q_p[X] + \frac{1}{1 - p} ESF_p[X] \] (119)
Equivalence of Risk Measures [DGK (2006)]

Proof.

Under the transformation $x \rightarrow q$, where $x = F_X^{-1}(q) = Q_q[X]$ and

$$(q, dq) = \left( F_X(x), f_X(x)dx \right)$$
Proof.

Under the transformation $x \to q$, where $x = F_X^{-1}(q) = Q_q[X]$ and

$$(q, dq) = \left( F_X(x), f_X(x)dx \right)$$

$$ESF_p[X] = \int_{-\infty}^{\infty} (x - Q_p[X])_+ f_X(x) dx$$

$$= \int_{0}^{1} (Q_q[X] - Q_p[X])_+ dq$$

$$= \int_{0}^{1} Q_q[X] dq - (1 - p)Q_p[X]$$
Equivalence of Risk Measures [DGK (2006)]

Proof.

Under the transformation $x \rightarrow q$, where $x = F_X^{-1}(q) = Q_q[X]$ and

$$(q, dq) = \left( F_X(x), f_X(x)dx \right)$$

$$ESF_p[X] = \int_{-\infty}^{\infty} (x - Q_p[X])_+ f_X(x)dx$$

$$= \int_{0}^{1} (Q_q[X] - Q_p[X])_+ dq$$

$$= \int_{p}^{1} Q_q[X]dq - (1 - p)Q_p[X]$$

$$\Rightarrow TVaR_p[X] = \frac{1}{1 - p} \int_{p}^{1} Q_q[X]dq = \frac{ESF_p[X]}{1 - p} + Q_p[X].$$

(120)
Note that if $\mathbb{E}[X] < \infty$ and if $F_X$ is continuous, then

$$\lim_{p \to 0} TVaR_p[X] = \mathbb{E}[X]$$

$$TVaR_p[X] = CTE_p[X]$$

(121)
Normally Distributed Risk

For an $X \sim N(\mu, \sigma^2)$, it can be shown for $0 < p < 1$ that

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

$$CTE_p[X] = \mathbb{E}[X \mid X > Q_p[X]] = \frac{1}{1-p} \int_{Q_p}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} ye^{-\frac{(y-\mu)^2}{2\sigma}} dy$$

$$= \frac{1}{1-p} \int_{Q_p}^{\infty} \frac{\mu + \sigma z}{\sqrt{2\pi}\sigma} e^{-\frac{z^2}{2}} dz$$

$$= \frac{\mu}{1-p} \int_{\Phi^{-1}(p)}^{\infty} \frac{e^{-\frac{z^2}{2}}}{\sqrt{2\pi}\sigma} dz + \frac{\sigma}{1-p} \int_{\Phi^{-1}(p)}^{\infty} \frac{ze^{-\frac{z^2}{2}}}{\sqrt{2\pi}\sigma} dz$$

$$= \mu + \frac{\sigma}{1-p} \phi(\Phi^{-1}(p))$$

$$\mathbb{E}[(X - \rho[X])_+] = \sigma \phi\left(\frac{\rho[X] - \mu}{\sigma}\right) - (\rho[X] - \mu)\left[1 - \Phi\left(\frac{\rho[X] - \mu}{\sigma}\right)\right]$$

$$ESF_p[X] = \sigma \phi\left(\Phi^{-1}(p)\right) - \sigma (1-p) \phi(\Phi^{-1}(p))$$
For a $\ln X \sim N(\mu, \sigma^2)$, it can be shown for $0 < p < 1$ that

$$Q_p[X] = e^{\mu + \sigma \Phi^{-1}(p)}$$

$$\mathbb{E}[(X - \rho[X])_+] = e^{\mu + \frac{\sigma^2}{2}} \Phi \left( \sigma + \frac{\mu - \ln \rho[X]}{\sigma} \right) - \rho[X] \Phi \left( \frac{\mu - \ln \rho[X]}{\sigma} \right)$$

$$ESF_p[X] = e^{\mu + \frac{\sigma^2}{2}} \Phi \left( \sigma - \Phi^{-1}(p) \right) - e^{\mu + \sigma \Phi^{-1}(p)} (1 - p)$$

$$CTE_p[X] = e^{\mu + \frac{\sigma^2}{2}} \frac{\Phi \left( \sigma - \Phi^{-1}(p) \right)}{1 - p}$$

(123)
If \( S_X(x) = \left( \frac{\theta}{\theta + x} \right)^\alpha \), then
\[
Q_p[X] = \theta [(1 - p)^{-\frac{1}{\alpha}} - 1]
\]
\[
CTE_p[X] = Q_p[X] + \frac{Q_p[X] + \theta}{\alpha - 1}.
\]

HW:
- Compute \( ESF_p[X] \).
- 3.31, 3.32, 3.35, 3.36.
- Show that \( CTE_p[X] \) is, in general, not subadditive.
- How about \( ESF_p[X] \) - is it subadditive?
- How about \( \rho[X] := E[X] \) - is it subadditive?
Consider again the empirically observed distribution of **Losses** for accident claims from randomly selected policies is as follows:

<table>
<thead>
<tr>
<th>Claim Amount $L$</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.607</td>
</tr>
<tr>
<td>100</td>
<td>0.303</td>
</tr>
<tr>
<td>200</td>
<td>0.076</td>
</tr>
<tr>
<td>500</td>
<td>0.012</td>
</tr>
<tr>
<td>1000</td>
<td>0.002</td>
</tr>
</tbody>
</table>

Compute the $CTE_p[X]$ for $p \in \{0.90, 0.95, 0.99\}$. Consult the Hardy Study Note for cases where the VaR levels $Q_p = Q_{p+\epsilon}$ for some $\epsilon > 0$. 
Thoughts on $CTE_p[X]$ 

- $CTE_p[X] \geq Q_p[X]$, making it more preferable in the eyes of a regulator who always desires higher solvency capital.
- $CTE_p[X]$ is used for equity-linked life contingent insurance in Canada.
- **Minimum Capital Test For Federally Regulated Property and Casualty Insurance Companies** report from the Government of Canada: 
  
  "...(u)nder the MCT, regulatory capital requirements for various risks are set directly at a pre-determined target confidence level. OSFI has elected 99% of the expected shortfall (conditional tail expectation or CTE 99%) over a one-year time horizon as a target confidence level", with the footnote "As an alternative, a value at risk (VaR) at 99.5% confidence level or expert judgement was used when it was not practical to use the CTE approach."

- $CTE_p$ is not the only non-subadditive measure, apparently.
- What to do? Is there a "better" property than sub-additivity?
Distortion Risk Measures

- Recalling actuarial terminology, for the loss distribution $F : \mathbb{R}_+ \rightarrow [0, 1]$, define the **Survival Function**
  \[ S(x) = 1 - F(x). \] (125)

- Define the **Distortion Function** $g : [0, 1] \rightarrow [0, 1]$ with $g(0) = 0$, $g(1) = 1$, and $g$ increasing.

- Define the **Distortion Risk Measure** that inflates the probability of large losses via
  \[ \rho_g[X] = \int_0^\infty g(S(x))dx \]
  \[ = \mathbb{E}_g[X]. \] (126)

**Theorem**

DVTGKV (2004) *If $g$ is also concave, then for any losses $X, Y \in \mathcal{A}$,*

\[ \mathbb{P}[(X, Y) \in \mathbb{R}_+^2] = 1 \Rightarrow \rho_g[X + Y] \leq \rho_g[X] + \rho_g[Y]. \] (127)
Distortion Risk Measures


- **Hint:** Try

\[
g(x) = 1_{\{x \geq 1 - p\}}
\]

\[
g(x) = \min \left\{ \frac{x}{1 - p}, 1 \right\}
\]

(128)

and use the relationships obtained for the above risk measures. What about probability masses?
For $\alpha \geq 1$, define
\[ g(s) = s^{\frac{1}{\alpha}} \]  
(129)

If $X \sim \text{Pareto}(\gamma, \theta)$, then $\mathbb{E}[X] = \frac{\theta}{\gamma - 1}$ and (assuming $\gamma > \alpha$),
\[ S(x) = \mathbb{P}[X > x] = \left( \frac{\theta}{\theta + x} \right)^\gamma \]
\[ g(S(x)) = \left( \frac{\theta}{\theta + x} \right)^{\frac{\gamma}{\alpha}} \]
\[ \Rightarrow \rho_g[X] = \int_0^\infty \left( \frac{\theta}{\theta + x} \right)^{\frac{\gamma}{\alpha}} \, dx \]
\[ = \frac{\theta}{\frac{\gamma}{\alpha} - 1} \]  
(130)
For $\alpha \geq 1$, define

$$g(s) = s^{\frac{1}{\alpha}} \quad (131)$$

If $X \sim \text{exp}(\lambda)$, then $E[X] = \frac{1}{\lambda}$ and (assuming $\gamma > \alpha$,)

$$S(x) = \mathbb{P}[X > x] = e^{-\lambda x}$$

$$g(S(x)) = e^{-\frac{\lambda}{\alpha}x}$$

$$\Rightarrow \rho_g[X] = \int_0^\infty e^{-\frac{\lambda}{\alpha}x} \, dx$$

$$= \frac{\alpha}{\lambda} \quad (132)$$

**HW:** How about if $X \sim N(\mu, \sigma^2)$?
Define $g(s) = 1 - (1 - s)^N$. Note that if $N \in \mathbb{N}$, then for $N$ i.i.d. copies of $X$: i.e. $\{X_1, X_2, \ldots, X_N\}$ it follows that

$$g(S(x)) = 1 - (1 - S(x))^N = \mathbb{P}[\max \{X_1, X_2, \ldots, X_N\} > x]$$

$$\therefore \rho_g[X] = \int_0^\infty g(S(x))dx = \mathbb{E}[\max \{X_1, X_2, \ldots, X_N\}].$$

**HW:** Compute $\rho_g[X]$ for $X \sim \text{exp}(\lambda)$. Use $(\lambda, N) = (0.1, 10)$ and whatever numerical tools you feel comfortable with.
The Wang Transform

Wang, in A Universal Framework for Pricing Financial and Insurance (2002) presents the following idea: Define $\lambda \in \mathbb{R}$ and a distortion function $g$ such that

$$g(S(x)) = \Phi\left(\Phi^{-1}(S(x)) + \lambda\right).$$  

(134)

Then there is a universal pricing method based on this transform for financial assets or liabilities over a time horizon $[0, T]$. In this distortion measure, $\lambda$ is the market price of risk, reflecting the level of systematic risk.

- **HW:** Consider an $X \sim N(\mu, \sigma^2)$. How does the Wang Transform act on $X$? Specifically, compute $\rho_g[X]$.
- **HW:** Consider an $X \sim LN(\mu, \sigma^2)$. How does the Wang Transform act on $X$? Specifically, compute $\rho_g[X]$. How does $\rho_g[X]$ relate to the quantile measure $Q_p[X]$ in this case?
Spectral Risk Measure

Assume a non-negative, non-increasing, right-continuous, integrable function $\gamma : [0, 1] \rightarrow \mathbb{R}_+$ such that

$$\int_0^1 \gamma(p) dp = 1 \quad (135)$$

and the corresponding risk measure

$$\rho_\gamma[X] := \int_0^1 \gamma(p) Q_p[X] dp. \quad (136)$$

Then $\rho_\gamma[X]$ is a **Spectral Risk Measure**.

- **HW:** Consider $X \sim \text{exp}(\lambda)$, and $\gamma(p) = 2p$. Compute $\rho_\gamma[X]$.

- Is Expected Shortfall a Spectral Risk Measure? How about Value at Risk? How about just Expected Value?

- Are Distortion and Spectral Risk Measures related? Check out Theorem 3.1 here for a nice connection.
In the case of choosing an adequate capital requirement, a regulator is charged with making the event \( \{ X > \rho[X] \} \) "unlikely." A selection process can entail first choosing the a **shortfall measure** \( M \), where \( M \) satisfies

\[
\rho_1[X] \leq \rho_2[X] \Rightarrow M \left( X - \rho_1[X] \right)_+ \geq M \left( X - \rho_2[X] \right)_+. \quad (137)
\]

This reflects the view of the regulator that an increase of solvency capital implies a reduction of the shortfall risk that \( M \) measures. A choice of a monotonic \( M \) is enough to guarantee the above condition.
Note that this condition relates the view of the regulator that increasing regulatory capital decreases the likelihood of insolvency.

As mentioned earlier, this view must compete with the cost of holding excess capital.

Given the choice of regulatory measure $M$, the optimal solution $u = \rho[X]$ solves

$$
\min_{u \in \mathbb{R}^+} \left( \mathbb{E} \left[ M[(X - u)_+] \right] + \epsilon u \right).
$$

(138)
It was also shown earlier that for $M[x] = x$, the optimal solvency capital is $\rho[X] = Q_{1-\epsilon}[X]$. 

It follows from our Equivalence Of Risk Measures Theorem that the minimized expected penalized shortfall using this capital is also the measure 

$$
\min_{u \in \mathbb{R}^+} \left( \mathbb{E}\left[(X - u)_{+}\right] + \epsilon u \right) = \epsilon TVaR_{1-\epsilon}[X]. \tag{139}
$$

From the view of a company that must self-determine the optimal solvency capital, the problem mentioned above is also known as determining the \textbf{economic capital} required.

Please refer to Economic Capital Allocation Derived From Risk Measures [DGK (2003)] for more on economic capital from an actuarial perspective.

Great overall review paper is Risk Measures and Comonotonicity: A Review [DGK (2006)]

state and prove the foundational theorem:

Theorem

For a finite set of outcomes, \( \rho \) is a coherent risk measure if and only if \( \exists \) a family \( \mathcal{P} \) of probability measures on the set of outcomes such that

\[
\rho[X] = \sup \left\{ \mathbb{E}^Q[X] : Q \in \mathcal{P} \right\}.
\] (140)

To see how this relates to the margin system SPAN (Standard Portfolio Analysis of Risk) developed by the Chicago Mercantile Exchange, please look through Section 3.1 here and the references within.
An estimator is a quantity, rule, or formula that estimates properties of a set of values. For example, if our set is \( A = \{x_1, x_2, \ldots, x_n\} \), then a common estimator is

\[
\bar{x} := \frac{\sum_{j=1}^{n} x_j}{n}.
\] (141)

A single estimate that results in a value for a unknown parameter \( \theta \) is called a point estimator \( \hat{\theta} \). Such an estimator is called unbiased if

\[
bias(\hat{\theta}) := E[\hat{\theta} - \theta] = 0.\] (142)
A population consists of the values 1, 3, 5, and 9. We want to estimate the mean of the population \( \mu \). A random sample of two values from this population is taken without replacement, and the mean of the sample \( \hat{\mu} \) is used as an estimator of the population mean \( \mu \).

- Find the probability distribution of \( \hat{\mu} \)
- Is \( \hat{\mu} \) an unbiased estimator?
We find that there are six possible outcomes for this experiment, consisting of 6 pairs of values drawn:

$$S = \{\{1, 3\}, \{1, 5\}, \{1, 9\}, \{3, 5\}, \{3, 9\}, \{5, 9\}\}.$$  \hspace{1cm} (143)

Each of these pairs is equally likely, and so we assign probability $$\frac{1}{6}$$ to each of the possible values of $$\hat{\mu}$$:

$$\hat{\mu} \in \{2, 3, 4, 5, 6, 7\}.$$  \hspace{1cm} (144)

and so it follows that

$$\mu = \frac{1 + 3 + 5 + 9}{4} = 4.5$$

$$\mathbb{E}[\hat{\mu}] = \frac{1}{6}(2 + 3 + 4 + 5 + 6 + 7) = 4.5$$

$$\therefore \text{bias}(\hat{\theta}) = \mathbb{E}[\hat{\mu} - \mu] = 0.$$  \hspace{1cm} (145)
Asymptotically Unbiased and Consistent Estimators

If, as the sample size increases, the bias vanishes we say that the estimator \( \hat{\theta}_n \) is asymptotically unbiased if for all \( \theta \):

\[
\lim_{n \to \infty} \mathbb{E}[\hat{\theta}_n - \theta] = 0.
\] (146)

If, for all \( \delta > 0 \),

\[
\lim_{n \to \infty} \mathbb{P}[|\hat{\theta}_n - \theta| > \delta] = 0
\] (147)

then \( \hat{\theta}_n \) is a consistent estimator.
Theorem

If \( \lim_{n \to \infty} \mathbb{E}[\hat{\theta}_n - \theta] = 0 = \lim_{n \to \infty} \text{Var}[\hat{\theta}_n], \) then \( \hat{\theta}_n \) is consistent.

Proof.

Fix a \( \delta > 0 \). By Markov’s Inequality,

\[
0 \leq \mathbb{P}[|\hat{\theta}_n - \theta| > \delta] \leq \frac{\mathbb{E}[(\hat{\theta}_n - \theta)^2]}{\delta^2} = \frac{\mathbb{E} \left[ \left( (\hat{\theta}_n - \mathbb{E}[\hat{\theta}_n]) + (\mathbb{E}[\hat{\theta}_n] - \theta) \right)^2 \right]}{\delta^2} \]

\[
= \frac{\text{Var}[\hat{\theta}_n] - 2\mathbb{E}[\hat{\theta}_n - \theta] \cdot \mathbb{E}[\hat{\theta}_n - \mathbb{E}[\hat{\theta}_n]] + \mathbb{E}[(\mathbb{E}[\hat{\theta}_n] - \theta)^2]}{\delta^2} \to 0.
\]
Assume that \( \{X_1, ..., X_n\} \) are i.i.d., and for all \( i \in \{1, 2, ..n\} \), \( X_i \sim \text{exp}(\lambda) \).

- Are there any values \( \lambda > 0 \) such that
  \[
  \hat{\lambda}_n = \max \{X_1, .., X_n\} \\
  \bar{\lambda}_n = \min \{X_1, .., X_n\}
  \]
  as estimators of \( \mathbb{E}[X] = \frac{1}{\lambda} \) are Asymptotically Unbiased? Consistent?

- If not, can we multiply them by scaling functions of \( n \) to get these desired properties?
The mean-squared error associated with an estimator \( \hat{\theta} \) is

\[
\text{MSE}(\hat{\theta}) = \mathbb{E}[ (\hat{\theta} - \theta)^2 ] = \text{Var}[\hat{\theta}] + \text{bias}^2(\hat{\theta}).
\] (150)

- If two unbiased estimators \( \hat{\theta}_1, \hat{\theta}_2 \) are such that \( \text{MSE}(\hat{\theta}_1) < \text{MSE}(\hat{\theta}_2) \), then we say that \( \hat{\theta}_1 \) is more efficient than \( \hat{\theta}_2 \).
- If \( \hat{\theta}_1 \) is more efficient that any other unbiased estimator \( \hat{\theta} \), then we say that \( \hat{\theta}_1 \) is a uniformly minimum variance unbiased estimator.

HW:

- Find \( \text{MSE}(\hat{\lambda}) \) for the previous HW problem.
- **HW**: Practice problems 46.1 – 46.4, 46.7 in Finan.
Assume that \( \{X_1, ..., X_n\} \) are i.i.d., and for all \( i \in \{1, 2, ..n\} \), \( X_i \sim U(0, L) \). Define \( \hat{L}_n \) as an estimator of \( L \) such that

\[
\hat{L}_n = \max \{ X_1, .., X_n \}.
\]  
(151)

Is \( \hat{L}_n \) Asymptotically Unbiased? Consistent? Asympotically zero -MSE?
By definition, if we set $F_{\hat{L}_n}(y) := \mathbb{P}[\hat{L}_n \leq y]$, then

$$F_{\hat{L}_n}(y) = \left( \frac{y}{L} \right)^n$$

$$f_{\hat{L}_n}(y) = \frac{ny^{n-1}}{L^n}. \quad (152)$$
It follows that

\[ \text{Var}(\hat{L}_n) + \text{bias}^2(\hat{L}_n) = \text{MSE}(\hat{L}_n) \]

\[ = \mathbb{E}[(\hat{L}_n - L)^2] = \int_0^L (y - L)^2 \frac{ny^{n-1}}{L^n} \, dy \]

\[ = \frac{n}{L^n} \int_0^L (y^2 - 2Ly + L^2)y^{n-1} \, dy \]

\[ = \frac{3L^2}{(n + 1)(n + 2)} \rightarrow 0. \quad (153) \]

So not only is \( \hat{L}_n \) asymptotically zero -MSE, but also \( \text{Var}(\hat{L}_n) \rightarrow 0 \) and \( \text{bias}(\hat{L}_n) \rightarrow 0 \) and so \( \hat{L}_n \) is consistent by our previous theorem.
For a random variable $X$ with parameters $\{\theta_1, ..., \theta_p\}$, define $\theta = (\theta_1, ..., \theta_p)$ and

$$F(x \mid \theta) = \mathbb{E}[1\{X \leq x\} \mid \theta]$$

$$\mu'_k(\theta) = \mathbb{E}[X^k \mid \theta]$$

$$\bar{\mu}'_k = \frac{1}{n} \sum_{i=1}^{n} (x_i)^k$$

for a sequence of realizations $\{\{X_1 = x_1\}, ..., \{X_n = x_n\}\}$. 

(154)
A method-of-moments estimate of the vector $\theta$ is any solution to the system of $p$ equations

$$\mu'_k(\theta) = \bar{\mu}'_k.$$  \hfill (155)

The smoothed empirical estimate of the $100g^{th}$ percentile is obtained by arranging the observed values $\{x_1, \ldots, x_n\}$ from smallest to largest:

$$\{x_1, \ldots, x_n\} \rightarrow \{x(1), \ldots, x(n)\}$$  \hfill (156)

and finding the integer $p$ such that $\frac{p}{n+1} \leq g \leq \frac{p+1}{n+1}$. 
From this, we define

\[ \hat{\text{VaR}}_g[X] := [p + 1 - (n + 1)g]x(p) + [(n + 1)g - p]x(p+1). \quad (157) \]

and the method of matching percentiles corresponds to matching

\[ F(\hat{\text{VaR}}_{g_k}[X] | \theta) = g_k \quad \forall k \in \{1, ..., p\} \quad (158) \]

and a set of arbitrarily chosen percentiles \( \{g_1, ..., g_p\} \).

- See Examples 58.5 and 58.8 in Finan for an application of matching percentiles.
- See Examples 58.7 and 58.10 in Finan for an application of method of moments.
- **HW:** Practice problems 58.6 – 58.10 in Finan.
You are given:

- Losses $X$ follow a loglogistic distribution with cumulative distribution function characterized by parameters $(\theta, \gamma)$:

$$F_X(x) = \frac{\left(\frac{x}{\theta}\right)^\gamma}{1 + \left(\frac{x}{\theta}\right)^\gamma}, \forall x \geq 1$$  \hspace{1cm} (159)

- A random sample of losses is

$$\{\text{Sample Losses}\} = \{10, 35, 80, 86, 90, 120, 158, 180, 200, 210, 1500\}.$$ \hspace{1cm} (160)

Calculate the estimate of $\theta$ by percentile matching, using the $40^{th}$ and $80^{th}$ empirically smoothed percentile estimates.
The 40\textsuperscript{th} percentile is the .4(12) = 4.8\textsuperscript{th} smallest observation. By interpolation, it is .2(86) + .8(90) = 89.2. The 80\textsuperscript{th} percentile is the .8(12) = 9.6\textsuperscript{th} smallest observation. By interpolation it is .4(200) + .6(210) = 206. The equations to solve are

\[
0.4 = \frac{(\frac{89.2}{\theta})^\gamma}{1 + (\frac{89.2}{\theta})^\gamma}
\]

\[
0.8 = \frac{(\frac{206}{\theta})^\gamma}{1 + (\frac{206}{\theta})^\gamma}
\]

\[
\Rightarrow (\hat{\gamma}, \hat{\theta}) = (2.1407, 107.08).
\]
You are given:

- A sample \( \{x_1, x_2, \ldots, x_{10}\} \) is drawn from a distribution with probability density function (with \( \theta > \sigma \)):

\[
f_X(x) = \frac{1}{2} \left( \frac{1}{\theta} e^{-\frac{x}{\theta}} + \frac{1}{\sigma} e^{-\frac{x}{\sigma}} \right), \; \forall x \in (0, \infty) \tag{162}
\]

- Summed values

\[
\sum_{i=1}^{10} x_i = 150
\]

\[
\sum_{i=1}^{10} x_i^2 = 5000. \tag{163}
\]

Estimate \( \theta \) by matching the first two sample moments to the corresponding population quantities.
By definition of exponential distributions, we obtain the pair of equations:

\[
\begin{align*}
\frac{150}{10} &= \mathbb{E}[X] = \frac{1}{2}(\theta + \sigma) \\
\frac{5000}{10} &= \mathbb{E}[X^2] = \frac{1}{2}(2\theta^2 + 2\sigma^2) \\
\Rightarrow (\hat{\theta}, \hat{\sigma}) &= (20, 10).
\end{align*}
\] (164)

The solution can be found via

\[
\begin{align*}
30 &= \theta + \sigma \Rightarrow \sigma = 30 - \theta \\
500 &= \theta^2 + \sigma^2 = \theta^2 + (30 - \theta)^2 = 2\theta^2 - 60\theta + 900 \\
\Rightarrow \theta &\in \{10, 20\}.
\end{align*}
\] (165)

We choose \(\theta = 20\) to fix \(\sigma = 10\).
Define the empirically observed quantities

\[
\bar{X} := \frac{1}{n} \sum_{k=1}^{n} x_k
\]

\[
\bar{X^2} := \frac{1}{n} \sum_{k=1}^{n} x_k^2
\]

(166)

for a sequence of \( n \) observed values \( \{X_1 = x_1, \ldots, X_n = x_n\} \). Assume that the \( \{X_i\}_{i=1}^{n} \) are i.i.d. and drawn from a population \( X \sim U[a, b] \). Use the method of moments to determine \( a, b \) in terms of \( \bar{X}, \bar{X^2} \).
Two-sided Uniform Interval Estimation

By definition,

\[ E[X] = \frac{a + b}{2} \]
\[ E[X^2] = \frac{a^2 + b^2 + ab}{3}. \] (167)

Estimation via matching of the first two moments:

\[ \bar{X} = \frac{\hat{a} + \hat{b}}{2} \]
\[ \bar{X^2} = \frac{\hat{a^2} + \hat{b^2} + \hat{a}\hat{b}}{3}. \] (168)

Algebra leads to

\[ (\hat{a}, \hat{b}) = \left( \bar{X} - \sqrt{3} \sqrt{\bar{X^2} - \bar{X^2}}, \bar{X} + \sqrt{3} \sqrt{\bar{X^2} - \bar{X^2}} \right). \] (169)
Assume again a probability space \((\Omega, \mathcal{F}, \mathbb{P})\), where a subset \(\{A_i\}_{i=1}^n\) of \(\mathcal{F}\) is contained within \(\Omega\), and

\[
\bigcup_{i=1}^n A_i \subseteq \Omega. \tag{170}
\]

Here, the \(A_i\) represent observations obtained from an experiment consisting of a sequence of i.i.d. random variables \(\{X_i\}_{i=1}^n\), pulled from a common distribution \(F(\cdot \mid \theta)\).
The idea now is to use information obtained by the outcomes observed \( \{X_i \in A_i\} \). In fact, we determined that the likelihood function

\[
L(\theta) := \prod_{i=1}^{n} \mathbb{P}[X_i \in A_i \mid \theta]
\]  

(171)

should be maximized, due to our observation that this independent sequence of events has occurred!
Maximum Likelihood Estimation for Complete Data

For ease of computation, we may wish to maximize the log-likelihood function

$$I(\theta) := \ln [L(\theta)] = \sum_{i=1}^{n} \ln (P[X_i \in A_i \mid \theta]).$$

(172)
You are given:

- Losses $X$ follow a Single-parameter Pareto distribution with density function characterized by a parameter $\alpha \in (0, \infty)$

$$f_X(x) = \frac{\alpha}{x^{\alpha+1}}, \quad \forall x > 1$$  \hspace{1cm} (173)

- A random sample of size five produced three losses with values 3, 6 and 14, and two losses exceeding 25.

Determine the maximum likelihood estimate of $\alpha$. 
By definition, $S_X(x) = \frac{1}{x^\alpha}$ and so

$$L(\alpha) = f_X(3)f_X(6)f_X(14)S_X(25)^2 = \frac{\alpha}{3^{\alpha+1}} \frac{\alpha}{6^{\alpha+1}} \frac{\alpha}{14^{\alpha+1}} \frac{1}{25^{2\alpha}}$$

$$= \frac{\alpha^3}{252 \cdot 157500^\alpha}$$

$$l(\alpha) = \ln[L(\alpha)] = 3 \ln \alpha - \alpha \ln (157500) - \ln (252)$$

$$\Rightarrow l'(\alpha) = \frac{3}{\alpha} - \ln (157500)$$

$$\Rightarrow \hat{\alpha} = \frac{3}{\ln (157500)} = 0.2507.$$
You are given:

- Losses $X$ follow an exponential distribution with mean $\theta$
- A random sample of 20 losses is distributed as follows:

<table>
<thead>
<tr>
<th>Loss Range</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>[0,1000]</td>
<td>7</td>
</tr>
<tr>
<td>(1000,2000]</td>
<td>6</td>
</tr>
<tr>
<td>(2000,\infty)</td>
<td>7</td>
</tr>
</tbody>
</table>

Determine the maximum likelihood estimate of $\theta$. 
Our loss function is, for \( p := e^{-\frac{1000}{\theta}} \),

\[
L(\theta) = F(1000)^7 [F(2000) - F(1000)]^6 [1 - F(2000)]^7
= \left(1 - e^{-\frac{1000}{\theta}}\right)^7 \left(e^{-\frac{1000}{\theta}} - e^{-\frac{2000}{\theta}}\right)^6 \left(e^{-\frac{2000}{\theta}}\right)^7
\]

\[
L(p) = (1 - p)^7 (p - p^2)^6 (p^2)^7 = p^{20} (1 - p)^{13}
\]

\[
l(p) = 20 \ln p + 13 \ln (1 - p)
\]

\[
l'(\hat{p}) = \frac{20}{\hat{p}} - \frac{13}{1 - \hat{p}} = 0
\]

\[
\Rightarrow \hat{p} = \frac{20}{33} = 1 - e^{-\frac{1000}{1996.90}}
\]

\[
\Rightarrow \hat{\theta} = 1996.90.
\]
See Examples 59.1 – 59.5 in Finan for an application of maximum likelihood estimators to loss models.

**HW:** Practice problems 59.2 – 59.7 in Finan.

**HW:** Read Section 60 in Finan on Maximum Likelihood Estimation for Incomplete Data.

**HW:** Read Section 67 in Finan on Estimation of Class \((a, b, 0)\).

**HW:** Practice problems 67.2.2 – 67.4 in Finan.

**Note:** Take-home final exam may have problems very similar in scope to those above!
Assume now that there is a history of claims \( \{X_1, \ldots, X_n\} \), all with a common risk parameter \( \Theta \) such that the \( \{X_k \mid \Theta\}_{k=1}^n \) are i.i.d.

Furthermore, based on the history we have observed, we would like to use it to estimate the (common) expected loss

\[
\mu(\Theta) := \mathbb{E}[X_{n+1} \mid \Theta].
\]  

(176)
The question to ask is how Credible the data \( \{X_1, \ldots, X_n\} \) is. As a risk manager, there is the option to charge the book or manual premium

\[
\mu := \mathbb{E}[\mu(\Theta)] \quad (177)
\]

or charge the weighted average premium for some sequence of weights \( \{\beta_1, \ldots, \beta_n\} \):

\[
\bar{X} := \sum_{k=1}^{n} \beta_k X_k. \quad (178)
\]
As in the section on risk measures, we seek a metric $Z$ for the credibility of the observed claims $\{X_1, \ldots, X_n\}$ so far.

This value $Z$ should depend on everything, and measure the likelihood that the observed data reflects reality, and should be used for premiums over the assumed average $\mu$:

$$Z : (\Theta, X_1, \ldots, X_n) \to [0, 1]. \quad (179)$$

The balance reached between history and assumed $\mu$ results in an optimal premium at time $n + 1$:

$$P_{n+1} = (1 - Z)\mu + Z\bar{X}. \quad (180)$$
A risk manager may be tempted to give full credibility, $Z = 1$, to observed data. In order to justify this, there must be some stability to the observed data.

By assigning equal weight to the observed claims, $\alpha_k := \frac{1}{n}$, we seek to determine whether, by the Central Limit Theorem

$$p \leq \mathbb{P} \left[ \left| \frac{\bar{X} - \mu}{\mu} \right| \leq r \right] = \mathbb{P} \left[ \left| \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \right| \leq \frac{r\mu\sqrt{n}}{\sigma} \right] \approx 2 \Phi \left( \frac{r\mu\sqrt{n}}{\sigma} \right) - 1 \ (181)$$

for company-determined values of $r$ and $p$, and assumed parameters $(\mu, \sigma^2) = (\mathbb{E}[X_j], \text{Var}[X_j])$. 
Hence, the question of full-credibility now reduces to a large enough number of exposure units:

\[ \frac{r\mu\sqrt{n}}{\sigma} \geq \Phi^{-1} \left( \frac{1 + p}{2} \right) =: y_p. \]  \hspace{1cm} (182)

Rewritten in terms of \( n \), the inequality is now

\[ n \geq \lambda_0 \left( \frac{\sigma}{\mu} \right)^2 \]

\[ \lambda_0 := \left( \frac{y_p}{r} \right)^2. \]  \hspace{1cm} (183)
Consider now the joint sequence \( \{(N_i, X_i)\}_{i=1}^n \), where \( N_i \) is the total number of claims in year \( i \), and \( X_i \) is the total loss in year \( i \).

Assume that the \( X_i \) and \( N_i \) are independent of each other, and that the \( N_i \sim \text{Poisson}(\lambda) \) are i.i.d. as well as \( X_i \) independent with common distribution \( F_X \).

Let \( Y_{ij} \) denote the \( j^{th} \) claim in year \( i \). Assume that the \( Y_{ij} \) are i.i.d. with mean \( \theta_Y \) and variance \( \sigma_Y^2 \). It follows that

\[
X_i = \sum_{j=1}^{N_i} Y_{ij}. \tag{184}
\]
Based on this set-up, the independence of the claim frequency and claim size leads to

\[
\mathbb{E}[X_i] = \mathbb{E}[N_i] \cdot \mathbb{E}[Y_{ij}] = \lambda \theta_Y
\]

\[
\text{Var}[X_i] = \text{Var}[N_i] \cdot \mathbb{E}[Y_{ij}]^2 + \mathbb{E}[N_i] \cdot \text{Var}[Y_{ij}]
\]

\[
= \lambda (\theta_Y^2 + \sigma_Y^2).
\]

It follows that for full credibility, we need

\[
n \geq \lambda_0 \frac{\sigma_{X_i}^2}{\mathbb{E}[X_i]^2} = \lambda_0 \frac{\theta_Y^2 + \sigma_Y^2}{\lambda \theta_Y^2}
\]
You are given:

- The number of claims has a Poisson distribution.
- Claim sizes $Y_{ij}$ have a Pareto distribution with common values $(\theta_Y, \sigma_Y^2) = (0.10, 0.015)$.
- The number of claims and claim sizes are independent.
- The observed pure premium should be within 2% of the expected pure premium 90% of the time.

Determine the expected number of claims needed for full credibility.
The expected number of claims is equal to $\lambda n$. From the information provided, $r = 0.02$ and $y_{0.90} = 1.645$, and so for full credibility, we require

$$\lambda n \geq \lambda \cdot \left( \lambda_0 \frac{\sigma_i^2}{\mathbb{E}[X_i] \lambda_0^2} \right)$$

$$= \lambda_0 \frac{\theta^2_Y + \sigma_Y^2}{\theta^2_Y}$$

$$= \left( \frac{1.645}{0.02} \right)^2 \left( 1 + \frac{0.015^2}{0.10^2} \right)$$

$$= 16913.$$
One way of determining $Z \in (0, 1)$ is by optimizing over all linear functions of past data:

$$\begin{align*}
\min_{\alpha_1, \ldots, \alpha_n} \mathbb{E} \left[ \left( \mu(\Theta) - \left[ \mu(1 - \sum_{k=1}^{n} \alpha_k) + \sum_{k=1}^{n} \alpha_k X_k \right] \right)^2 \right].
\end{align*}$$

(188)
Using Hilbert space theory, Shiu and Sing show the optimal weights to be

\[ \hat{\alpha}_k = \frac{\mathbb{E}[(\mu - \mu(\Theta))^2]}{\mathbb{E}[(X_k - \mu(\Theta))^2]} = \frac{\text{Var}[\mu(\Theta)]}{\mathbb{E}[\text{Var}[X_k|\Theta]]} \]

\[ 1 - Z = 1 - \sum_{k=1}^{n} \hat{\alpha}_k = \frac{1}{1 + \sum_{j=1}^{n} \mathbb{E}[(\mu - \mu(\Theta))^2] / \mathbb{E}[(X_j - \mu(\Theta))^2]} = \frac{1}{1 + \sum_{j=1}^{n} \text{Var}[\mu(\Theta)] / \mathbb{E}[\text{Var}[X_j|\Theta]]} \]

(189)
From the previous slide, we have

\[
Z = \frac{\sum_{j=1}^{n} \text{Var}[\mu(\Theta)]}{1 + \sum_{j=1}^{n} \frac{\text{Var}[\mu(\Theta)]}{\mathbb{E}[\text{Var}[X_j|\Theta]]}}. \tag{190}
\]

If, however, we have a common value \( k \) such that, for all \( j \in \{1, \ldots, n\} \),

\[
\frac{1}{k} := \frac{\text{Var}[\mu(\Theta)]}{\mathbb{E}[\text{Var}[X_j|\Theta]]} \tag{191}
\]

then

\[
Z = \frac{n \frac{1}{k}}{1 + n \frac{1}{k}} = \frac{n}{n + k}. \tag{192}
\]

This common value \( k \) is known as Bühlmanns \( k \)-factor. Notice that if \( n \to \infty \), then \( Z \to 1 \).
Bühlmanns $k$–factor

Furthermore, we have that for constant $\beta := \mathbb{E}[\text{Var}[X_k | \Theta]]$, $\beta_k := \frac{1}{\mathbb{E}[\text{Var}[X_k | \Theta]]} \sum_{j=1}^{n} \frac{1}{\mathbb{E}[\text{Var}[X_j | \Theta]]} = \frac{1}{n}$ (193)

and so we can write

$$P_{n+1} = (1 - Z)\mu + Z \left( \frac{X_1 + \ldots + X_n}{n} \right).$$ (194)

In this setting, we can extend to number of claims or other observed variables $\{V_1, \ldots, V_n\}$ and say that our Bühlmann Credible Estimate for $V_{n+1}$ is

$$\hat{V}_{n+1} = (1 - Z)\mu_V + Z \left( \frac{V_1 + \ldots + V_n}{n} \right).$$ (195)
Assume the existence of a claim count random variable such that $N \sim \text{Poisson}(\lambda)$.
Assume that the claim sizes $\{X_1, X_2, \ldots\}$ to be i.i.d. copies of a random variable $X$.
Assume that $X, \{X_1, X_2, \ldots\}$ are independent of $N$.

Based on these assumptions, define the aggregate claim $Y \mid N$ such that

$$Y \mid N := \sum_{k=1}^{N} X_k. \tag{196}$$
By the Tower Property of Conditional Expectations, and the fact that \( N \) is Poisson-distributed, we have

\[
\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y \mid N]] = \mathbb{E}[N \mathbb{E}[X]] = \mathbb{E}[N] \mathbb{E}[X].
\]

\[
\text{Var}(Y) = \mathbb{E}[	ext{Var}(Y \mid N)] + \text{Var}(\mathbb{E}[Y \mid N])
\]
\[
= \mathbb{E}[N \text{Var}(X)] + \text{Var}(N \mathbb{E}[X])
\]
\[
= \mathbb{E}[N] \text{Var}(X) + \text{Var}(N) \mathbb{E}[X]^2
\]
\[
= \mathbb{E}[N] \text{Var}(X) + \mathbb{E}[N] \mathbb{E}[X]^2
\]
\[
= \mathbb{E}[N] \mathbb{E}[X^2]
\]
You are given:

- Claim counts follow a Poisson distribution with mean $\theta$.
- Claim sizes follow an exponential distribution with mean $10\theta$.
- Claim counts and claim sizes are independent, given $\theta$.
- The prior distribution has probability density function

$$f_\Theta(\theta) = \frac{5}{\theta^6} \text{ for } \theta > 1.$$  \hspace{1cm} (198)

Calculate Bühlmanns $k-$factor for aggregate losses.
Define $N$ as the Poisson claim count variable, $X$ as the claim size variable, and $Y$ as the aggregate loss variable. It follows that

\[
\mu(\theta) = \mathbb{E}[Y \mid \Theta = \theta] = \mathbb{E}[N \mid \Theta = \theta] \cdot \mathbb{E}[X \mid \Theta = \theta] = \theta \cdot 10\theta = 10\theta^2.
\]

\[
v(\theta) = \text{Var}[Y \mid \Theta = \theta] = \mathbb{E}[N \mid \Theta = \theta] \cdot \mathbb{E}[X^2 \mid \Theta = \theta] = 200\theta^3.
\]
To compute $k$, we need

$$
\mu = \mathbb{E}[\mu(\Theta)] = \int_1^\infty 10\theta^2 \cdot \frac{5}{\theta^6} \, d\theta = \frac{50}{3}
$$

$$
\text{Var}[\mu(\Theta)] = \text{Var}[10\Theta^2] = \int_1^\infty \left(10\theta^2 - \frac{50}{3}\right)^2 f_{\Theta}(\theta) \, d\theta
$$

$$
= \frac{11111}{50} = 222.22
$$

$$
\mathbb{E}[v(\Theta)] = \int_1^\infty 200\theta^3 \cdot f_{\Theta}(\theta) \, d\theta = 500
$$

$$
\Rightarrow k = \frac{\mathbb{E}[v(\Theta)]}{\text{Var}[\mu(\Theta)]} = \frac{500}{222.22} = 2.25.
$$
Another way to use previously observed data to estimate future claims and losses is via Bayesian Parameter Estimation. In this approach, we assume that we wish to estimate $X$ using observed data by first conditioning the parameter distribution on observed data:

$$E[X_{n+1} | X_1, ..., X_n] = \int_{\theta \in \mathbb{R}} E[X_{n+1} | \theta] f_\Theta(\theta | X_1, ..., X_n) d\theta. \quad (201)$$

**Question:** Is there a connection between Bayesian Parameter Estimation and Credibility Theory?
You are given:

- The size $X$ of a claim is uniformly distributed on the interval $[0, \theta]$.
- The prior distribution has probability density function

$$f_\Theta(\theta) = \frac{500}{\theta^2} \text{ for } \theta > 500.$$  \hspace{1cm} (202)

Two claims, $x_1 = 400$ and $x_2 = 600$, are observed. You calculate the posterior distribution as:

$$f_\Theta(\theta \mid X_1 = 400, X_2 = 600) = 3 \left( \frac{600^3}{\theta^4} \right) \text{ for } \theta > 600.$$ \hspace{1cm} (203)

Calculate the Bayesian premium $\mathbb{E}[X_3 \mid X_1 = x_1, X_2 = x_2]$. 
As \( X \sim U[0, \theta] \), we have \( \mathbb{E}[X | \Theta] = \frac{1}{2} \Theta \). It follows that

\[
\mathbb{E}[X_3 | X_1 = 400, X_2 = 600] = \int_{\theta \in \mathbb{R}} \mathbb{E}[X_3 | \theta] f_{\Theta}(\theta | X_1 = 400, X_2 = 600) d\theta \\
= \int_{600}^{\infty} \frac{1}{2} \theta \cdot 3 \left( \frac{600^3}{\theta^4} \right) d\theta \\
= 3 \frac{600^3}{2} \int_{600}^{\infty} \frac{1}{\theta^3} d\theta \\
= 3 \frac{600^3}{2} \cdot 2 \cdot 600^2 = 450.
\]
Recall that if we take a sequence of \( i.i.d. \) (independent, identically distributed) random variables \( \{X_1, X_2, ..., X_n\} \), called a \textit{random sample}, with common mean and variance:

\[
\mathbb{E}[X_i] = \mu \\
\text{Var}[X_i] = \sigma^2
\] (205)
Recall that if we take a sequence of \( i.i.d. \) (independent, identically distributed) random variables \( \{X_1, X_2, ..., X_n\} \), called a \textit{random sample}, with common mean and variance:

\[
\mathbb{E}[X_i] = \mu \\
\text{Var}[X_i] = \sigma^2
\]

then

\[
\mathbb{E} \left( \frac{X_1 + X_2 + ... + X_n}{n} - \mu \right) = 0 \\
\text{Var} \left( \frac{X_1 + X_2 + ... + X_n}{n} - \mu \right) = 1
\]
Introduction

If the random sample comes from a Normal population, then by the previous slide, we know

\[
\frac{X_1 + X_2 + \ldots + X_n}{\frac{\sigma}{\sqrt{n}}} - \mu \sim N(0, 1) \tag{207}
\]
If the random sample comes from a Normal population, then by the previous slide, we know

\[
\frac{X_1 + X_2 + \ldots + X_n}{\frac{\sigma}{\sqrt{n}}} - \mu \sim N(0, 1) \tag{207}
\]

This implies we can use our **Observed Sample Mean** \( \frac{X_1 + X_2 + \ldots + X_n}{n} \) to estimate the **Population Sample Mean** \( \mu \) as long as we know \( \sigma \) a priori:

\[
P \left[ -z_{\frac{\alpha}{2}} \leq \frac{X_1 + X_2 + \ldots + X_n}{\frac{\sigma}{\sqrt{n}}} - \mu \leq z_{\frac{\alpha}{2}} \right] = P \left[ -z_{\frac{\alpha}{2}} \leq Z \leq z_{\frac{\alpha}{2}} \right] = \Phi \left( \frac{z_{\frac{\alpha}{2}}}{\sigma} \right) - \Phi \left( -\frac{z_{\frac{\alpha}{2}}}{\sigma} \right) = 1 - \alpha \tag{208}
\]
In fact, we can \textit{tune} the accuracy of our prediction by how small we choose \( \alpha > 0 \) and how large \( n \) of a sample we take. Symbolically, if we observe our sample mean to be \( \bar{x} \), then we are \( 100 \cdot (1 - \alpha) \% \) that 
\[
\mu \in \left( \bar{x} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right)
\]
In fact, we can *tune* the accuracy of our prediction by how small we choose $\alpha > 0$ and how large $n$ of a sample we take. Symbolically, if we observe our sample mean to be $\bar{x}$, then we are $100 \cdot (1 - \alpha)\%$ that

$$\mu \in \left( \bar{x} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right)$$

Some values:

<table>
<thead>
<tr>
<th>$100 \cdot (1 - \alpha)%$</th>
<th>$\frac{\alpha}{2}$</th>
<th>$z_{\frac{\alpha}{2}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>90%</td>
<td>0.05</td>
<td>1.645</td>
</tr>
<tr>
<td>95%</td>
<td>0.025</td>
<td>1.960</td>
</tr>
<tr>
<td>99%</td>
<td>0.005</td>
<td>2.576</td>
</tr>
</tbody>
</table>
Interval Width

In general, a confidence interval with width $n$ observations and $100 \cdot (1 - \alpha)\%$ certainty has a total width

$$w = 2z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}.$$  

(209)

Inverting this for $n$, we derive the number $n$ observations required for a given width $w$, confidence level $100 \cdot (1 - \alpha)\%$, and known deviation $\sigma$ to be

$$n = \left(2z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{w}\right)^2.$$  

(210)
In general, a confidence interval with width $n$ observations and $100 \cdot (1 - \alpha)\%$ certainty has a total width

$$w = 2z_\frac{\alpha}{2} \cdot \frac{\sigma}{\sqrt{n}}. \quad (209)$$

Inverting this for $n$, we derive the number $n$ observations required for a given width $w$, confidence level $100 \cdot (1 - \alpha)\%$, and known deviation $\sigma$ to be

$$n = \left(2z_\frac{\alpha}{2} \cdot \frac{\sigma}{w} \right)^2. \quad (210)$$

For example, if $\alpha = 0.05$, $\sigma = 25$, $w = 10$, then $n = \left(2 \cdot 1.96 \cdot \frac{25}{10} \right)^2 = 96$. 
Interval Width

In general, a confidence interval with width $n$ observations and $100 \cdot (1 - \alpha)\%$ certainty has a total width

$$w = 2z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}.$$

(209)

Inverting this for $n$, we derive the number $n$ observations required for a given width $w$, confidence level $100 \cdot (1 - \alpha)\%$, and known deviation $\sigma$ to be

$$n = \left(2z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{w}\right)^2.$$

(210)

For example, if $\alpha = 0.05$, $\sigma = 25$, $w = 10$, then $n = \left(2 \cdot 1.96 \cdot \frac{25}{10}\right)^2 = 96$.

Notice that $n \propto \frac{1}{w^2}$. 
In most cases, one can safely assume that we are also estimating the Variance of the population along with the population mean. By the CLT, we know that for any random sample with mean $\mu$ and variance $\sigma^2$,

$$
\Pr\left[-z_{\frac{\alpha}{2}} \leq \frac{X_1 + X_2 + \ldots + X_n}{n} - \mu \leq z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right] \rightarrow 1 - \alpha. \quad (211)
$$
Unknown $\sigma$

In most cases, one can safely assume that we are also estimating the Variance of the population along with the population mean. By the CLT, we know that for any random sample with mean $\mu$ and variance $\sigma^2$,

$$\Pr \left[ -z_{\frac{\alpha}{2}} \leq \frac{\bar{X}_1 + \bar{X}_2 + \ldots + \bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \leq z_{\frac{\alpha}{2}} \right] \rightarrow 1 - \alpha. \quad (211)$$

If we work only from observed data, then we recall that our **Sample Mean-Variance pair** ($\bar{X}, S^2$) has the form

$$S^2 = \sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{n - 1}$$

$$\bar{X} = \frac{X_1 + X_2 + \ldots + X_n}{n} \quad (212)$$
If we work only from observed data, we can use our modified version of the CLT:

**Theorem**

\[
\Pr \left[ -z_{\frac{\alpha}{2}} \leq \frac{\bar{X} - \mu}{S / \sqrt{n}} \leq z_{\frac{\alpha}{2}} \right] \to 1 - \alpha. \tag{213}
\]

The speed of convergence depends on the underlying population distribution, of course. As a general rule of thumb, when using only observed data, one can consider the above limit achieved if \( n \geq 40 \).

As a computational consideration, recall that

\[
S^2 = \sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{n - 1}
\]
If we work only from observed data, we can use our modified version of the CLT:

**Theorem**

\[
P \left[ -z_{\frac{\alpha}{2}} \leq \frac{\bar{X} - \mu}{S} \leq z_{\frac{\alpha}{2}} \right] \rightarrow 1 - \alpha. \tag{213}
\]

The speed of convergence depends on the underlying population distribution, of course. As a general rule of thumb, when using only observed data, one can consider the above limit achieved if \( n \geq 40 \).

As a computational consideration, recall that

\[
S^2 = \sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{n-1} = \frac{1}{n-1} \sum_{i=1}^{n} X_i^2 - \frac{1}{n} \left( n \cdot (n-1) \left( \sum_{i=1}^{n} X_i \right) \right)^2 \tag{214}
\]
Imagine now that we are polling a group of people whose answer is either yes or no. Assign the value 1 for an answer of yes, 0 for the answer no. Define $X_i$ to be the numerical value of the $i^{th}$ respondent’s answer, and

$$\hat{p} := \frac{\sum_{i=1}^{n} X_i}{n}$$

(215)

the sample proportion of positive respondents. The question remains: how good of an estimation of the total population positive proportion $p$ is $\hat{p}$?
Luckily, we have seen this *Binomial* random variable before, and we know that given a population proportion $p$ of a yes answer, we have

\[
\begin{align*}
\mathbb{E}[\hat{p}] & := \mathbb{E} \left[ \frac{1}{n} \sum_{i=1}^{n} X_i \right] = p \\
\text{Var}(\hat{p}) & := \text{Var} \left( \frac{1}{n} \sum_{i=1}^{n} X_i \right) = \frac{p \cdot (1 - p)}{n} \\
& = \sigma_{\hat{p}}^2
\end{align*}
\]
Luckily, we have seen this *Binomial* random variable before, and we know that given a population proportion $p$ of a yes answer, we have

$$E[\hat{p}] := E \left[ \frac{1}{n} \sum_{i=1}^{n} X_i \right] = p$$

$$Var(\hat{p}) := Var \left( \frac{1}{n} \sum_{i=1}^{n} X_i \right) = \frac{p \cdot (1 - p)}{n}$$

$$= \sigma^2_{\hat{p}} = E[S^2]$$
Luckily, we have seen this *Binomial* random variable before, and we know that given a population proportion $p$ of a yes answer, we have

$$
\mathbb{E}[\hat{p}] := \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^{n} X_i\right] = p
$$

$$
\text{Var}(\hat{p}) := \text{Var}\left(\frac{1}{n} \sum_{i=1}^{n} X_i\right) = \frac{p \cdot (1 - p)}{n}
$$

$$
= \hat{p}^2 = \mathbb{E}[S^2] \quad \text{............... Really ?}
$$
Bionomial Distribution - Polling?

Luckily, we have seen this *Binomial* random variable before, and we know that given a population proportion $p$ of a yes answer, we have

$$
\mathbb{E}[\hat{p}] := \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^{n} X_i\right] = p
$$

$$
\text{Var}(\hat{p}) := \text{Var}\left(\frac{1}{n} \sum_{i=1}^{n} X_i\right) = \frac{p \cdot (1 - p)}{n}
$$

$$
= \sigma^2_{\hat{p}} = \mathbb{E}[S^2] \text{................. Really ?}
$$

It follows that for large $n$, we can make the approximation

$$
\frac{\hat{p} - p}{\sigma_{\hat{p}}} \sim \mathcal{N}(0, 1).
$$
Formally, the modified CLT tells us that

\[
P\left[-z_\frac{\alpha}{2} \leq \frac{\hat{p} - p}{\sqrt{p \cdot (1-p) \cdot \frac{1}{n}}} \leq z_\frac{\alpha}{2}\right] \to 1 - \alpha. \tag{218}
\]
This can be inverted to say that for large $n$,

$$\mathbb{P}[p_- \leq p \leq p_+] \approx 1 - \alpha$$

$$p_- = \hat{p} + \frac{z_{\alpha/2}^2}{n} - z_{\alpha/2} \sqrt{\frac{\hat{p}\hat{q}}{n} + \frac{z_{\alpha/2}^2}{4n^2}}$$

$$1 + \frac{z_{\alpha/2}^2}{n}$$

$$p_+ = \hat{p} + \frac{z_{\alpha/2}^2}{n} + z_{\alpha/2} \sqrt{\frac{\hat{p}\hat{q}}{n} + \frac{z_{\alpha/2}^2}{4n^2}}$$

$$1 + \frac{z_{\alpha/2}^2}{n}$$

(219)
Since

\[
P\left[ \frac{\bar{X}_1 + \bar{X}_2 + \ldots + \bar{X}_n - \mu}{\frac{S}{\sqrt{n}}} \leq z_\alpha \right] \rightarrow 1 - \alpha
\]

we say that

- \( \mu < \bar{X} + z_\alpha \frac{s}{\sqrt{n}} \) is a **Large Sample Upper Confidence Bound for** \( \mu \)
- \( \mu > \bar{X} - z_\alpha \frac{s}{\sqrt{n}} \) is a **Large Sample Lower Confidence Bound for** \( \mu \).
T-distributions

If the population of interest is normal, then $X_1, X_2, ..., X_n$ constitutes a normal random sample with both $\mu, \sigma$ unknown.
If the population of interest is normal, then $X_1, X_2, ... , X_n$ constitutes a normal random sample with both $\mu, \sigma$ unknown. Since we need $\sigma$ to estimate $\mu$, we do the next best thing and substitute $S$ for $\sigma$, as before.

$$T_{n-1} := \frac{\left( \frac{1}{n} \sum_{i=1}^{n} X_i \right) - \mu}{S \sqrt{\frac{n}{n-1}}}$$

$$f_{T_{n-1}}(x) = \frac{1}{\sqrt{(n-1)\pi}} \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} \frac{1}{\left(1 + \frac{x^2}{n-1}\right)^{\frac{n}{2}}}$$
If the population of interest is normal, then $X_1, X_2, \ldots, X_n$ constitutes a normal random sample with both $\mu, \sigma$ unknown. Since we need $\sigma$ to estimate $\mu$, we do the next best thing and substitute $S$ for $\sigma$, as before.

$$T_{n-1} := \frac{\left(\frac{1}{n} \sum_{i=1}^{n} X_i\right) - \mu}{\frac{S}{\sqrt{n}}}$$

$$f_{T_{n-1}}(x) = \frac{1}{\sqrt{(n-1)\pi}} \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} \frac{1}{\left(1 + \frac{x^2}{n-1}\right)^{\frac{n}{2}}}$$

$$\rightarrow \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} =: f_Z(x)$$
T-distributions

If the population of interest is normal, then $X_1, X_2, \ldots, X_n$ constitutes a normal random sample with both $\mu, \sigma$ unknown. Since we need $\sigma$ to estimate $\mu$, we do the next best thing and substitute $S$ for $\sigma$, as before.

$$T_{n-1} := \frac{\left(\frac{1}{n} \sum_{i=1}^{n} X_i\right) - \mu}{\frac{S}{\sqrt{n}}}$$

$$f_{T_{n-1}}(x) = \frac{1}{\sqrt{(n-1)\pi}} \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} \frac{1}{\left(1 + \frac{x^2}{n-1}\right)^{\frac{n}{2}}}$$

$$\rightarrow \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} =: f_Z(x)$$

$$\mathbb{E}[T_{n-1}] = 0$$

$$\text{Var}(T_{n-1}) = \frac{n-1}{n-3}$$
Define $t_{\alpha, n-1}$ to satisfy

$$P[T_{n-1} \geq t_{\alpha, n-1}] = \alpha \quad (222)$$

It follows that

$$P[-t_{\alpha \over 2, n-1} \leq T_{n-1} \leq t_{\alpha \over 2, n-1}] = 1 - \alpha \quad (223)$$

and the corresponding two-sided and one-sided confidence intervals can be derived.
Given a *normal* random sample $X_1, X_2, \ldots, X_n$, we want to predict the next, independent value $X_{n+1}$. One very tempting predictor is the historical, or sample, average $\overline{X} := \frac{1}{n} \sum_{i=1}^{n} X_i$ of the previous $n$ values. The residual error attached with this prediction is $\overline{X} - X_{n+1}$, and we can compute

$$
\mathbb{E}[\overline{X} - X_{n+1}] = \mu - \mu = 0
$$

$$
\text{Var} \left( \overline{X} - X_{n+1} \right) = \text{Var} \left( \overline{X} \right) + \text{Var}(X_{n+1})
$$

$$
= \frac{\sigma^2}{n} + \sigma^2 = \sigma^2 \cdot \left( 1 + \frac{1}{n} \right)
$$
With the above information, it follows that

\[
\frac{\bar{X} - X_{n+1}}{\sqrt{\sigma^2 \cdot (1 + \frac{1}{n})}} \sim N(0, 1)
\]

and

\[
\frac{\bar{X} - X_{n+1}}{\sqrt{S^2 \cdot (1 + \frac{1}{n})}} \sim T_{n-1}
\]

and we say that if the sample mean is \( \bar{X} \), and sample deviation is \( s \), then the prediction interval for a single observation from a normal population is

\[
P \left[ \left| X_{n+1} - \bar{X} \right| \leq t_{\frac{\alpha}{2}, n-1} \cdot s \cdot \sqrt{1 + \frac{1}{n}} \right] = 1 - \alpha
\]
Given a *normal* random sample $X_1, X_2, ..., X_n$, we see that

$$\frac{n - 1}{\sigma^2} S^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{X})^2}{\sigma^2}$$  \hspace{1cm} (227)$$

has a chi-squared ($\chi^2_{n-1}$) distribution with $n - 1$ degrees of freedom. Because $S^2$ is always positive, and because the chi-squared distribution is asymmetric, we compute the confidence interval

$$P \left[ \chi^2_{1 - \frac{\alpha}{2}, n-1} < \frac{n - 1}{\sigma^2} S^2 < \chi^2_{\frac{\alpha}{2}, n-1} \right] = 1 - \alpha$$  \hspace{1cm} (228)$$

where $P \left[ \chi^2_{n-1} > \chi^2_{\alpha, n-1} \right] = \alpha$