# TABLE OF CONTENTS - VOLUME 1

## INTRODUCTORY COMMENTS

## MODELING

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>SECTION 1 - PROBABILITY REVIEW</td>
<td>LM-1</td>
</tr>
<tr>
<td>PROBLEM SET 1</td>
<td>LM-9</td>
</tr>
<tr>
<td>SECTION 2 - REVIEW OF RANDOM VARIABLES - PART I</td>
<td>LM-19</td>
</tr>
<tr>
<td>PROBLEM SET 2</td>
<td>LM-29</td>
</tr>
<tr>
<td>SECTION 3 - REVIEW OF RANDOM VARIABLES - PART II</td>
<td>LM-35</td>
</tr>
<tr>
<td>PROBLEM SET 3</td>
<td>LM-43</td>
</tr>
<tr>
<td>SECTION 4 - REVIEW OF RANDOM VARIABLES - PART III</td>
<td>LM-51</td>
</tr>
<tr>
<td>PROBLEM SET 4</td>
<td>LM-59</td>
</tr>
<tr>
<td>SECTION 5 - PARAMETRIC DISTRIBUTIONS AND TRANSFORMATIONS</td>
<td>LM-63</td>
</tr>
<tr>
<td>PROBLEM SET 5</td>
<td>LM-71</td>
</tr>
<tr>
<td>SECTION 6 - DISTRIBUTION TAIL BEHAVIOR</td>
<td>LM-75</td>
</tr>
<tr>
<td>PROBLEM SET 6</td>
<td>LM-79</td>
</tr>
<tr>
<td>SECTION 7 - MIXTURE OF TWO DISTRIBUTIONS</td>
<td>LM-81</td>
</tr>
<tr>
<td>PROBLEM SET 7</td>
<td>LM-87</td>
</tr>
<tr>
<td>SECTION 8 - MIXTURE OF ( n ) DISTRIBUTIONS</td>
<td>LM-93</td>
</tr>
<tr>
<td>PROBLEM SET 8</td>
<td>LM-99</td>
</tr>
<tr>
<td>SECTION 9 - CONTINUOUS MIXTURES</td>
<td>LM-107</td>
</tr>
<tr>
<td>PROBLEM SET 9</td>
<td>LM-113</td>
</tr>
<tr>
<td>SECTION 10 - POLICY LIMITS AND THE LIMITED LOSS RANDOM VARIABLE</td>
<td>LM-121</td>
</tr>
<tr>
<td>PROBLEM SET 10</td>
<td>LM-125</td>
</tr>
<tr>
<td>SECTION 11 - POLICY DEDUCTIBLE (1), THE COST PER LOSS RANDOM VARIABLE</td>
<td>LM-127</td>
</tr>
<tr>
<td>PROBLEM SET 11</td>
<td>LM-133</td>
</tr>
</tbody>
</table>
MODELING

SECTION 12 - POLICY DEDUCTIBLE (2),
THE COST PER PAYMENT RANDOM VARIABLE
PROBLEM SET 12

SECTION 13 - POLICY DEDUCTIBLES APPLIED TO THE UNIFORM,
EXPONENTIAL AND PARETO DISTRIBUTIONS
PROBLEM SET 13

SECTION 14 - COMBINED LIMIT AND DEDUCTIBLE
PROBLEM SET 14

SECTION 15 - ADDITIONAL POLICY ADJUSTMENTS
PROBLEM SET 15

SECTION 16 - MODELS FOR THE NUMBER OF CLAIMS
AND THE \((a, b, 0)\) and \((a, b, 1)\) CLASSES
PROBLEM SET 16

SECTION 17 - MODELS FOR THE AGGREGATE LOSS,
COMPOUND DISTRIBUTIONS (1)
PROBLEM SET 17

SECTION 18 - COMPOUND DISTRIBUTIONS (2)
PROBLEM SET 18

SECTION 19 - MORE PROPERTIES OF THE AGGREGATE
LOSS RANDOM VARIABLE
PROBLEM SET 19

SECTION 20 - STOP LOSS INSURANCE
PROBLEM SET 20

SECTION 21 - RISK MEASURES
PROBLEM SET 21
# MODEL ESTIMATION

<table>
<thead>
<tr>
<th>Section</th>
<th>Problem Set</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SECTION 1 - REVIEW OF MATHEMATICAL STATISTICS (1)</strong></td>
<td></td>
</tr>
<tr>
<td>ESTIMATORS</td>
<td>ME-1</td>
</tr>
<tr>
<td>PROBLEM SET 1</td>
<td>ME-7</td>
</tr>
<tr>
<td><strong>SECTION 2 - REVIEW OF MATHEMATICAL STATISTICS (2)</strong></td>
<td></td>
</tr>
<tr>
<td>CONFIDENCE INTERVALS AND HYPOTHESIS TESTS</td>
<td>ME-11</td>
</tr>
<tr>
<td>PROBLEM SET 2</td>
<td>ME-17</td>
</tr>
<tr>
<td><strong>SECTION 3 - NON-PARAMETRIC EMPRICAL POINT ESTIMATION</strong></td>
<td></td>
</tr>
<tr>
<td>PROBLEM SET 3</td>
<td>ME-23</td>
</tr>
<tr>
<td><strong>SECTION 4 - KERNEL SMOOTHING ESTIMATORS</strong></td>
<td></td>
</tr>
<tr>
<td>PROBLEM SET 4</td>
<td>ME-37</td>
</tr>
<tr>
<td><strong>SECTION 5 - EMPIRICAL ESTIMATION FROM GROUPED DATA</strong></td>
<td></td>
</tr>
<tr>
<td>PROBLEM SET 5</td>
<td>ME-61</td>
</tr>
<tr>
<td><strong>SECTION 6 - ESTIMATION FROM CENSORED AND TRUNCATED DATA</strong></td>
<td></td>
</tr>
<tr>
<td>PROBLEM SET 6</td>
<td>ME-75</td>
</tr>
<tr>
<td><strong>SECTION 7 - PROPERTIES OF SURVIVAL PROBABILITY ESTIMATORS</strong></td>
<td></td>
</tr>
<tr>
<td>PROBLEM SET 7</td>
<td>ME-103</td>
</tr>
<tr>
<td><strong>SECTION 8 - MOMENT AND PERCENTILE MATCHING</strong></td>
<td></td>
</tr>
<tr>
<td>PROBLEM SET 8</td>
<td>ME-127</td>
</tr>
<tr>
<td><strong>SECTION 9 - MAXIMUM LIKELIHOOD ESTIMATION</strong></td>
<td></td>
</tr>
<tr>
<td>PROBLEM SET 9</td>
<td>ME-155</td>
</tr>
<tr>
<td><strong>SECTION 10 - MAXIMUM LIKELIHOOD ESTIMATION FOR THE EXPONENTIAL DISTRIBUTION</strong></td>
<td></td>
</tr>
<tr>
<td>PROBLEM SET 10</td>
<td>ME-177</td>
</tr>
<tr>
<td><strong>SECTION 11 - MAXIMUM LIKELIHOOD ESTIMATION FOR PARETO AND WEIBULL DISTRIBUTIONS</strong></td>
<td></td>
</tr>
<tr>
<td>PROBLEM SET 11</td>
<td>ME-189</td>
</tr>
</tbody>
</table>
MODEL ESTIMATION

SECTION 12 - MAXIMUM LIKELIHOOD ESTIMATION FOR DISTRIBUTIONS IN THE EXAM C TABLE
PROBLEM SET 12

SECTION 13 - PROPERTIES OF MAXIMUM LIKELIHOOD ESTIMATORS
PROBLEM SET 13

SECTION 14 - GRAPHICAL EVALUATION OF ESTIMATED MODELS
PROBLEM SET 14

SECTION 15 - HYPOTHESIS TESTS FOR FITTED MODELS
PROBLEM SET 15
INTRODUCTORY COMMENTS

This study guide is designed to help in the preparation for the Society of Actuaries Exam C. The exam covers the topics of modeling (including risk measure), model estimation, construction and selection, credibility and simulation.

The study manual is divided into two volumes. The first volume consists of a summary of notes, illustrative examples and problem sets with detailed solutions on the modeling and model estimation topics. The second volume consists of notes examples and problem sets on the credibility and simulation topics, as well as 14 practice exams.

The practice exams all have 35 questions. The level of difficulty of the practice exams has been designed to be similar to that of the past 4-hour exams. Some of the questions in the problem sets are taken from the relevant topics on SOA exams that have been released prior to 2009 but the practice exam questions are not from old SOA exams.

I have attempted to be thorough in the coverage of the topics upon which the exam is based, and consistent with the notation and content of the official reference text for the exam, "Loss Models" by Klugman, Panjer and Willmot. I have been, perhaps, more thorough than necessary on a couple of topics, such as maximum likelihood estimation, Bayesian credibility and applying simulation to hypothesis testing.

Because of the time constraint on the exam, a crucial aspect of exam taking is the ability to work quickly. I believe that working through many problems and examples is a good way to build up the speed at which you work. It can also be worthwhile to work through problems that have been done before, as this helps to reinforce familiarity, understanding and confidence. Working many problems will also help in being able to more quickly identify topic and question types. I have attempted, wherever possible, to emphasize shortcuts and efficient and systematic ways of setting up solutions. There are also occasional comments on interpretation of the language used in some exam questions. While the focus of the study guide is on exam preparation, from time to time there will be comments on underlying theory in places that I feel those comments may provide useful insight into a topic.

The notes and examples are divided into sections anywhere from 4 to 14 pages, with suggested time frames for covering the material. There are over 330 examples in the notes and over 800 exercises in the problem sets, all with detailed solutions. The 14 practice exams have 35 questions each, also with detailed solutions. Some of the examples and exercises are taken from previous SOA exams. Questions in the problem sets that have come from previous SOA exams are identified as such. Some of the problem set exercises are more in depth than actual exam questions, but the practice exam questions have been created in an attempt to replicate the level of depth and difficulty of actual exam questions. In total there are over 1600 examples/problems/sample exam questions with detailed solutions. ACTEX gratefully acknowledges the SOA for allowing the use of their exam problems in this study guide.

I suggest that you work through the study guide by studying a section of notes and then attempting the exercises in the problem set that follows that section. My suggested order for covering topics is (1) modeling (includes risk measures), (2) model estimation, (Volume 1), (3) credibility theory, and (4) simulation, (Volume 2).
It has been my intention to make this study guide self-contained and comprehensive for all Exam C topics, but there are occasional references to the Loss Models reference book (4th edition) listed in the SOA catalog. While the ability to derive formulas used on the exam is usually not the focus of an exam question, it is useful in enhancing the understanding of the material and may be helpful in memorizing formulas. There may be occasional references in the review notes to a derivation, but you are encouraged to review the official reference material for more detail on formula derivations. In order for the review notes in this study guide to be most effective, you should have some background at the junior or senior college level in probability and statistics. It will be assumed that you are reasonably familiar with differential and integral calculus. The prerequisite concepts to modeling and model estimation are reviewed in this study guide. The study guide begins with a detailed review of probability distribution concepts such as distribution function, hazard rate, expectation and variance.

Of the various calculators that are allowed for use on the exam, I am most familiar with the BA II PLUS. It has several easily accessible memories. The TI-30X IIS has the advantage of a multi-line display. Both have the functionality needed for the exam.

There is a set of tables that has been provided with the exam in past sittings. These tables consist of some detailed description of a number of probability distributions along with tables for the standard normal and chi-squared distributions. The tables can be downloaded from the SOA website www.soa.org.

If you have any questions, comments, criticisms or compliments regarding this study guide, please contact the publisher ACTEX, or you may contact me directly at the address below. I apologize in advance for any errors, typographical or otherwise, that you might find, and it would be greatly appreciated if you would bring them to my attention. ACTEX will be maintaining a website for errata that can be accessed from www.actexmadriver.com.

It is my sincere hope that you find this study guide helpful and useful in your preparation for the exam. I wish you the best of luck on the exam.

Samuel A. Broverman

Department of Statistical Sciences www.sambroverman.com
University of Toronto E-mail: sam@utstat.toronto.edu or 2brove@rogers.com

April, 2016
MODELING
Exam C applies probability and statistical methods to various aspects of loss modeling and model estimation. A good background in probability and statistics is necessary to fully understand models and the modeling that is done. In this section of the study guide, we will review fundamental probability rules.

**Sample point and probability space**
A sample point is the simple outcome of a random experiment. The probability space (also called sample space) is the collection of all possible sample points related to a specified experiment. When the experiment is performed, one of the sample points will be the outcome. An experiment could be observing the loss that occurs on an automobile insurance policy during the course of one year, or observing the number of claims arriving at an insurance office in one week. The probability space is the "full set" of possible outcomes of the experiment. In the case of the automobile insurance policy, it would be the range of possible loss amounts that could occur during the year, and in the case of the insurance office weekly number of claims, the probability space would be the set of integers \( \{0, 1, 2, \ldots\} \).

**Event**
Any collection of sample points, or any subset of the probability space is referred to as an event. We say "event \( A \) has occurred" if the experimental outcome was one of the sample points in \( A \).

**Union of events \( A \) and \( B \)**
\( A \cup B \) denotes the union of events \( A \) and \( B \), and consists of all sample points that are in either \( A \) or \( B \).

**Union of events \( A_1, A_2, \ldots, A_n \)**
\( A_1 \cup A_2 \cup \cdots \cup A_n = \bigcup_{i=1}^{n} A_i \) denotes the union of the events \( A_1, A_2, \ldots, A_n \), and consists of all sample points that are in at least one of the \( A_i \)'s. This definition can be extended to the union of infinitely many events.
**Intersection of events** $A_1, A_2, \ldots, A_n$

$A_1 \cap A_2 \cap \cdots \cap A_n = \bigcap_{i=1}^{n} A_i$ denotes the intersection of the events $A_1, A_2, \ldots, A_n$, and consists of all sample points that are simultaneously in all of the $A_i$'s.

![Intersection of events](image)

**Mutually exclusive events** $A_1, A_2, \ldots, A_n$

Two events are mutually exclusive if they have no sample points in common, or equivalently, if they have **empty intersection**. Events $A_1, A_2, \ldots, A_n$ are mutually exclusive if $A_i \cap A_j = \emptyset$ for all $i \neq j$, where $\emptyset$ denotes the empty set with no sample points. Mutually exclusive events cannot occur simultaneously.

**Exhaustive events** $B_1, B_2, \ldots, B_n$

If $B_1 \cup B_2 \cup \cdots \cup B_n = S$, the entire probability space, then the events $B_1, B_2, \ldots, B_n$ are referred to as exhaustive events.

**Complement of event** $A$

The complement of event $A$ consists of all sample points in the probability space that are **not in** $A$. The complement is denoted $\overline{A}$, $\sim A$, $A'$, or $A^c$ and is equal to $\{ x : x \notin A \}$. When the underlying random experiment is performed, to say that the complement of $A$ has occurred is the same as saying that $A$ has not occurred.

**Subevent (or subset)** $A$ of event $B$

If event $B$ contains all the sample points in event $A$, then $A$ is a subevent of $B$, denoted $A \subset B$. The occurrence of event $A$ implies that event $B$ has occurred.

**Partition of event** $A$

Events $C_1, C_2, \ldots, C_n$ form a partition of event $A$ if $A = \bigcup_{i=1}^{n} C_i$ and the $C_i$'s are mutually exclusive.

**DeMorgan's Laws**

(i) $(A \cup B)' = A' \cap B'$, to say that $A \cup B$ has not occurred is to say that $A$ has not occurred and $B$ has not occurred; this rule generalizes to any number of events,

$$\left( \bigcup_{i=1}^{n} A_i \right)' = (A_1 \cup A_2 \cup \cdots \cup A_n)' = A_1' \cap A_2' \cap \cdots \cap A_n' = \bigcap_{i=1}^{n} A_i'$$

(ii) $(A \cap B)' = A' \cup B'$, to say that $A \cap B$ has not occurred is to say that either $A$ has not occurred or $B$ has not occurred (or both have not occurred); this rule generalizes to any number of events,

$$\left( \bigcap_{i=1}^{n} A_i \right)' = (A_1 \cap A_2 \cap \cdots \cap A_n)' = A_1' \cup A_2' \cup \cdots \cup A_n' = \bigcup_{i=1}^{n} A_i'$$
Indicator function for event $A$

The function $I_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$ is the indicator function for event $A$, where $x$ denotes a sample point. $I_A(x)$ is 1 if event $A$ has occurred.

Some important rules concerning probability are given below.

(i) $P[S] = 1$ if $S$ is the entire probability space (when the underlying experiment is performed, some outcome must occur with probability 1).

(ii) $P[\emptyset] = 0$ (the probability of no face turning up when we toss a die is 0).

(iii) If events $A_1, A_2, \ldots, A_n$ are mutually exclusive (also called disjoint) then

$$P[\bigcup_{i=1}^{n} A_i] = P[A_1 \cup A_2 \cup \cdots \cup A_n] = P[A_1] + P[A_2] + \cdots + P[A_n] = \sum_{i=1}^{n} P[A_i].$$  

This extends to infinitely many mutually exclusive events.

(iv) For any event $A$, $0 \leq P[A] \leq 1$.

(v) If $A \subset B$ then $P[A] \leq P[B]$.


(viii) For any events $A$ and $B$, $P[A] = P[A \cap B] + P[A \cap B']$.

(ix) For exhaustive events $B_1, B_2, \ldots, B_n$, $P[\bigcup_{i=1}^{n} B_i] = 1$.

If $B_1, B_2, \ldots, B_n$ are exhaustive and mutually exclusive, they form a partition of the entire probability space, and for any event $A$,

$$P[A] = P[A \cap B_1] + P[A \cap B_2] + \cdots + P[A \cap B_n] = \sum_{i=1}^{n} P[A \cap B_i].$$  

(x) The words "percentage" and "proportion" are used as alternatives to "probability". As an example, if we are told that the percentage or proportion of a group of people that are of a certain type is 20%, this is generally interpreted to mean that a randomly chosen person from the group has a 20% probability of being of that type. This is the "long-run frequency" interpretation of probability. As another example, suppose that we are tossing a fair die. In the long-run frequency interpretation of probability, to say that the probability of tossing a 1 is $\frac{1}{6}$ is the same as saying that if we repeatedly toss the coin, the proportion of tosses that are 1's will approach $\frac{1}{6}$.
LM-1.2 Conditional Probability and Independence of Events

Conditional probability arises throughout the Exam C material. It is important to be familiar and comfortable with the definitions and rules of conditional probability.

**Conditional probability of event $A$ given event $B$**

If $P(B) > 0$, then $P(A|B) = \frac{P(A \cap B)}{P(B)}$ is the conditional probability that event $A$ occurs given that event $B$ has occurred. By rewriting the equation we get $P(A \cap B) = P(A|B) \cdot P(B)$.

**Partition of a Probability Space**

Events $B_1, B_2, ..., B_n$ are said to form a partition of a probability space $S$ if

(i) $B_1 \cup B_2 \cup \cdots \cup B_n = S$ and (ii) $B_i \cap B_j = \emptyset$ for any pair with $i \neq j$.

A partition is a disjoint collection of events which combines to be the full probability space. A simple example of a partition is any event $B$ and its complement $B'$.

If $A$ is any event in probability space $S$ and $\{B_1, B_2, ..., B_n\}$ is a partition of probability space $S$, then

$$P(A) = P(A \cap B_1) + P(A \cap B_2) + \cdots + P(A \cap B_n).$$

A special case of this rule is $P(A) = P(A \cap B) + P(A \cap B')$ for any two events $A$ and $B$.

**Bayes rule and Bayes Theorem**

For any events $A$ and $B$ with $P(A) > 0$, $P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A)}$. (1.7)

If $B_1$, $B_2$, ..., $B_n$ form a partition of the entire sample space $S$, then

$$P(B_j|A) = \frac{P(A|B_j) \cdot P(B_j)}{\sum_{i=1}^{n} P(A|B_i) \cdot P(B_i)}$$

for each $j = 1, 2, ..., n$. (1.8)

The values of $P(B_j)$ are called prior probabilities, and the value of $P(B_j|A)$ is called a posterior probability. Variations on this rule are very important in Bayesian credibility.
Independent events $A$ and $B$

If events $A$ and $B$ satisfy the relationship $P(A \cap B) = P(A) \cdot P(B)$, then the events are said to be independent or stochastically independent or statistically independent. The independence of (non-empty) events $A$ and $B$ is equivalent to $P(A|B) = P(A)$ or $P(B|A) = P(B)$.

Mutually independent events $A_1, A_2, \ldots, A_n$

The events are mutually independent if

(i) for any $A_i$ and $A_j$, $P(A_i \cap A_j) = P(A_i) \times P(A_j)$, and
(ii) for any $A_i, A_j$ and $A_k$, $P(A_i \cap A_j \cap A_k) = P(A_i) \times P(A_j) \times P(A_k)$, and so on for any subcollection of the events, including all events:

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1) \times P(A_2) \times \cdots \times P(A_n) = \prod_{i=1}^{n} P(A_i).$$  \hspace{1cm} (1.9)

Here are some rules concerning conditional probability and independence. These can be verified in a fairly straightforward way from the definitions given above.

(i) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ for any events $A$ and $B$ \hspace{1cm} (1.10)

(ii) $P(A \cap B) = P(A|B) \cdot P(A) = P(A|B) \cdot P(B)$ for any events $A$ and $B$ \hspace{1cm} (1.11)

(iii) If $B_1, B_2, \ldots, B_n$ form a partition of the sample space $S$, then for any event $A$,

$$P(A) = \sum_{i=1}^{n} P(A \cap B_i) = \sum_{i=1}^{n} P(A|B_i) \cdot P(B_i);$$ \hspace{1cm} (1.12)

as a special case, for any events $A$ and $B$, we have

$$P(A) = P(A \cap B) + P(A \cap B') = P(A|B) \cdot P(B) + P(A|B') \cdot P(B')$$ \hspace{1cm} (1.13)

(iv) If $P(A_1 \cap A_2 \cap \cdots \cap A_{n-1}) > 0$, then

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2) \cdots P(A_n|A_1 \cap A_2 \cap \cdots \cap A_{n-1})$$

(v) $P(A') = 1 - P(A)$ and $P(A'|B) = 1 - P(A|B)$ \hspace{1cm} (1.14)

(vi) if $A \subset B$ then $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)}{P(B)}$, and $P(B|A) = 1$

(vii) if $A$ and $B$ are independent events then $A'$ and $B$ are independent events, $A$ and $B'$ are independent events, and $A'$ and $B'$ are independent events

(viii) since $P(\emptyset) = P(\emptyset \cap A) = 0 = P(\emptyset) \cdot P(A)$ for any event $A$, it follows that $\emptyset$ is independent of any event $A$
Example LM1-1:
Suppose a fair six-sided die is tossed. We define the following events:

\[ A = "\text{the number tossed is } \leq 3" = \{1, 2, 3\}, \ B = "\text{the number tossed is even}" = \{2, 4, 6\}, \ C = "\text{the number tossed is a 1 or a 2}" = \{1, 2\}, \ D = "\text{the number thrown doesn't start with the letters 'f' or 't'}" = \{1, 6\}. \]

The conditional probability of \( A \) given \( B \) is

\[ P(A|B) = \frac{P(\{1,2,3\} \cap \{2,4,6\})}{P(\{2,4,6\})} = \frac{P(\{2\})}{P(\{2,4,6\})} = \frac{1/6}{1/2} = \frac{1}{3}. \]

Events \( A \) and \( B \) are not independent, since \( \frac{1}{6} = P(A \cap B) \neq P(A) \cdot P(B) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} \),
or alternatively, events \( A \) and \( B \) are not independent since \( P(A|B) \neq P(A) \).

\[ P(A|C) = 1 \neq \frac{1}{2} = P(A), \] so that \( A \) and \( C \) are not independent.

\[ P(B|C) = \frac{1}{2} = P(B), \] so that \( B \) and \( C \) are independent

(alternatively, \( P(B \cap C) = P(\{2\}) = \frac{1}{6} = \frac{1}{2} \cdot \frac{1}{3} = P(B) \cdot P(C) \).

It is not difficult to check that both \( A \) and \( B \) are independent of \( D \).

**IMPORTANT NOTE:** The following manipulation of event probabilities arises from time to time:

\[ P(A) = P(A|B) \cdot P(B) + P(A|B') \cdot P(B'). \]

If we know the conditional probabilities for event \( A \) given some other event \( B \) and its complement \( B' \), and if we know the (unconditional) probability of event \( B \), then we can find the probability of event \( A \). One of the important aspects of applying this relationship is the determination of the appropriate events \( A \) and \( B \).

Example LM1-2:
Urn I contains 2 white and 2 black balls and Urn II contains 3 white and 2 black balls. An Urn is chosen at random, and a ball is randomly selected from that Urn. Find the probability that the ball chosen is white.

**Solution:**

Let \( B \) be the event that Urn I is chosen and \( B' \) is the event that Urn II is chosen. The implicit assumption is that both Urns are equally likely to be chosen (this is the meaning of "an Urn is chosen at random"). Therefore, \( P(B) = \frac{1}{2} \) and \( P(B') = \frac{1}{2} \). Let \( A \) be the event that the ball chosen is white. If we know that Urn I was chosen, then there is \( \frac{1}{2} \) probability of choosing a white ball (2 white out of 4 balls, it is assumed that each ball has the same chance of being chosen); this can be described as \( P(A|B) = \frac{1}{2} \).

In a similar way, if Urn II is chosen, then \( P(A|B') = \frac{3}{5} \) (3 white out of 5 balls). We can now apply the relationship described prior to this example. \( P(A \cap B) = P(A|B) \cdot P(B) = \left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = \frac{1}{4} \), and

\[ P(A \cap B') = P(A|B') \cdot P(B') = \left(\frac{3}{5}\right)\left(\frac{1}{2}\right) = \frac{3}{10}. \]

Finally,

\[ P(A) = P(A \cap B) + P(A \cap B') = \frac{1}{4} + \frac{3}{10} = \frac{11}{20}. \]

The order of calculations can be summarized in the following table

<table>
<thead>
<tr>
<th>( A )</th>
<th>( B )</th>
<th>( B' )</th>
<th>( A )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P(A \cap B) = P(A</td>
<td>B) \cdot P(B) )</td>
<td>( P(A \cap B') = P(A</td>
<td>B') \cdot P(B') )</td>
</tr>
</tbody>
</table>
Example LM1-3:
Urn I contains 2 white and 2 black balls and Urn II contains 3 white and 2 black balls. One ball is chosen at random from Urn I and transferred to Urn II, and then a ball is chosen at random from Urn II. The ball chosen from Urn II is observed to be white. Find the probability that the ball transferred from Urn I to Urn II was white.

Solution:
Let $B$ denote the event that the ball transferred from Urn I to Urn II was white and let $A$ denote the event that the ball chosen from Urn II is white. We are asked to find $P(B|A)$.

From the simple nature of the situation (and the usual assumption of uniformity in such a situation, meaning all balls are equally likely to be chosen from Urn I in the first step), we have $P(B) = \frac{1}{2}$ (2 of the 4 balls in Urn I are white), and by implication, it follows that $P(B') = \frac{1}{2}$.

If the ball transferred is white, then Urn II has 4 white and 2 black balls, and the probability of choosing a white ball out of Urn II is $\frac{2}{6}$; this is $P(A|B) = \frac{2}{3}$.

If the ball transferred is black, then Urn II has 3 white and 3 black balls, and the probability of choosing a white ball out of Urn II is $\frac{1}{3}$; this is $P(A|B') = \frac{1}{2}$.

All of the information needed has been identified. We do calculations in the following order:

1. $P(A \cap B) = P(A|B) \cdot P(B) = \left(\frac{2}{3}\right)\left(\frac{1}{2}\right) = \frac{1}{3}$
2. $P(A \cap B') = P(A|B') \cdot P(B') = \left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = \frac{1}{4}$
3. $P(A) = P(A \cap B) + P(A \cap B') = \frac{1}{3} + \frac{1}{4} = \frac{7}{12}$
4. $P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{1/3}{7/12} = \frac{4}{7}.

Example LM1-4:
Three dice have the following probabilities of throwing a "six": $p$, $q$, $r$, respectively. One of the dice is chosen at random and thrown (each is equally likely to be chosen). A "six" appeared. What is the probability that the die chosen was the first one?

Solution:
The event "a 6 is thrown" is denoted by "6"

$$P[\text{die 1}|\text{6}] = \frac{P[\text{die 1} \cap \{\text{6}\}]}{P[\{\text{6}\}]} = \frac{P[\text{6}|\text{die 1}] \cdot P[\text{die 1}]}{P[\{\text{6}\}]} = \frac{p}{2} \cdot \frac{1}{6} = \frac{p}{12}.$$ But

$$P[\{\text{6}\}] = P[\{\text{6}\} \cap \{\text{die 1}\}] + P[\{\text{6}\} \cap \{\text{die 2}\}] + P[\{\text{6}\} \cap \{\text{die 3}\}]$$
$$= P[\{\text{6}\}|\text{die 1}] \cdot P[\text{die 1}] + P[\{\text{6}\}|\text{die 2}] \cdot P[\text{die 2}] + P[\{\text{6}\}|\text{die 3}] \cdot P[\text{die 3}]$$
$$= p \cdot \frac{1}{6} + q \cdot \frac{1}{6} + r \cdot \frac{1}{6} = \frac{p+q+r}{3} \Rightarrow P[\text{die 1}|\{\text{6}\}] = \frac{\frac{p}{12}}{\frac{p+q+r}{3}} = \frac{p}{p+q+r}.$$
1. A survey of 1000 people determines that 80% like walking and 60% like biking, and all like at least one of the two activities. How many people in the survey like biking but not walking?

A) 0        B) .1        C) .2        D) .3        E) .4

2. A life insurer classifies insurance applicants according to the following attributes:
   - the applicant is male
   - the applicant is a homeowner
   Out of a large number of applicants the insurer has identified the following information:
   40% of applicants are male, 40% are homeowners and
   20% are female homeowners.

Find the percentage of applicants who are male and do not own a home.

A) .1        B) .2        C) .3        D) .4        E) .5

3. Let $A$, $B$, $C$ and $D$ be events such that $B = A'$, $C \cap D = \emptyset$, and

$$P[A] = \frac{1}{4}, \quad P[B] = \frac{3}{4}, \quad P[C|A] = \frac{1}{2}, \quad P[C|B] = \frac{3}{4}, \quad P[D|A] = \frac{1}{4}, \quad P[D|B] = \frac{1}{8}$$

Calculate $P[C \cup D]$.

A) $\frac{5}{32}$        B) $\frac{1}{4}$        C) $\frac{27}{32}$        D) $\frac{3}{4}$        E) 1

4. You are given that $P[A] = .5$ and $P[A \cup B] = .7$.
   Actuary 1 assumes that $A$ and $B$ are independent and calculates $P[B]$ based on that assumption.
   Actuary 2 assumes that $A$ and $B$ mutually exclusive and calculates $P[B]$ based on that assumption.

Find the absolute difference between the two calculations.

A) 0        B) .05        C) .10        D) .15        E) .20

5. A test for a disease correctly diagnoses a diseased person as having the disease with probability .85. The test incorrectly diagnoses someone without the disease as having the disease with a probability of .10. If 1% of the people in a population have the disease, what is the chance that a person from this population who tests positive for the disease actually has the disease?

A) .0085        B) .0791        C) .1075        D) .1500        E) .9000

6. Two bowls each contain 5 black and 5 white balls. A ball is chosen at random from bowl 1 and put into bowl 2. A ball is then chosen at random from bowl 2 and put into bowl 1. Find the probability that bowl 1 still has 5 black and 5 white balls.

A) $\frac{2}{3}$        B) $\frac{3}{5}$        C) $\frac{6}{11}$        D) $\frac{1}{2}$        E) $\frac{6}{13}$
7. People passing by a city intersection are asked for the month in which they were born. It is assumed that the population is uniformly divided by birth month, so that any randomly passing person has an equally likely chance of being born in any particular month. Find the minimum number of people needed so that the probability that no two people have the same birth month is less than .5.

A) 2  B) 3  C) 4  D) 5  E) 6

8. In a T-maze, a laboratory rat is given the choice of going to the left and getting food or going to the right and receiving a mild electric shock. Assume that before any conditioning (in trial number 1) rats are equally likely to go the left or to the right. After having received food on a particular trial, the probability of going to the left and right become .6 and .4, respectively on the following trial. However, after receiving a shock on a particular trial, the probabilities of going to the left and right on the next trial are .8 and .2, respectively. What is the probability that the animal will turn left on trial number 2?

A) .1  B) .3  C) .5  D) .7  E) .9

9. In the game show "Let's Make a Deal", a contestant is presented with 3 doors. There is a prize behind one of the doors, and the host of the show knows which one. When the contestant makes a choice of door, at least one of the other doors will not have a prize, and the host will open a door (one not chosen by the contestant) with no prize. The contestant is given the option to change his choice after the host shows the door without a prize. If the contestant switches doors, what is the probability that he gets the door with the prize?

A) 0  B) \( \frac{1}{6} \)  C) \( \frac{1}{3} \)  D) \( \frac{1}{2} \)  E) \( \frac{2}{3} \)

10. A supplier of a testing device for a type of component claims that the device is highly reliable, with 

\[ P[A|B] = P[A'|B'] = .95 \]

where

\[ A = \text{device indicates component is faulty, and} \]

\[ B = \text{component is faulty.} \]

You plan to use the testing device on a large batch of components of which 5% are faulty. Find the probability that the component is faulty given that the testing device indicates that the component is faulty.

A) 0  B) .05  C) .15  D) .25  E) .50
11. An insurer classifies flood hazard based on geographical areas, with hazard categorized as low, medium and high. The probability of a flood occurring in a year in each of the three areas is

<table>
<thead>
<tr>
<th>Area</th>
<th>Hazard</th>
<th>low</th>
<th>medium</th>
<th>high</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prob. of Flood</td>
<td>.001</td>
<td>.02</td>
<td>.25</td>
<td></td>
</tr>
</tbody>
</table>

The insurer's portfolio of policies consists of a large number of policies with 80% low hazard policies, 18% medium hazard policies and 2% high hazard policies. Suppose that a policy had a flood claim during a year. Find the probability that it is a high hazard policy.

A) .50        B) .53        C) .56        D) .59        E) .62

12. One of the questions asked by an insurer on an application to purchase a life insurance policy is whether or not the applicant is a smoker. The insurer knows that the proportion of smokers in the general population is .30, and assumes that this represents the proportion of applicants who are smokers. The insurer has also obtained information regarding the honesty of applicants:

40% of applicants that are smokers say that they are non-smokers on their applications, none of the applicants who are non-smokers lie on their applications.

What proportion of applicants who say they are non-smokers are actually non-smokers?

A) 0        B) \frac{6}{11}        C) \frac{12}{11}        D) \frac{35}{41}        E) 1

13. When sent a questionnaire, 50% of the recipients respond immediately. Of those who do not respond immediately, 40% respond when sent a follow-up letter. If the questionnaire is sent to 4 persons and a follow-up letter is sent to any of the 4 who do not respond immediately, what is the probability that at least 3 never respond?

A) \left( .3 \right)^4 + 4 \left( .3 \right)^3 \left( .7 \right) 
B) 4 \left( .3 \right)^3 \left( .7 \right) 
C) \left( .1 \right)^4 + 4 \left( .1 \right)^3 \left( .9 \right) 
D) .4 \left( .3 \right) \left( .7 \right)^3 + \left( .7 \right)^4 
E) \left( .9 \right)^4 + 4 \left( .9 \right)^3 \left( .1 \right) 

14. A fair coin is tossed. If a head occurs, 1 fair die is rolled; if a tail occurs, 2 fair dice are rolled. If \( Y \) is the total on the die or dice, then \( P[Y = 6] = \)

A) \frac{1}{9} 
B) \frac{5}{36} 
C) \frac{11}{72} 
D) \frac{1}{6} 
E) \frac{11}{36}

15. In Canada's national 6-49 lottery, a ticket has 6 numbers each from 1 to 49, with no repeats. Find the probability of matching exactly 4 of the 6 winning numbers if the winning numbers are all randomly chosen.

A) .00095 
B) .00097 
C) .00099 
D) .00101 
E) .00103
1. Let $A = \text{"like walking"}$ and $B = \text{"like biking"}$. We use the interpretation that "percentage" and "proportion" are taken to mean "probability".

We are given $P(A) = .8$, $P(B) = .6$ and $P(A \cup B) = 1$.

From the diagram below we can see that since $A \cup B = A \cup (B \cap A')$ we have $P(A \cup B) = P(A) + P(A' \cap B) \rightarrow P(A' \cap B) = .2$ is the proportion of people who like biking but (and) not walking. In a similar way we get $P(A \cap B') = .4$

![Diagram showing A and B with intersections and unions]

An algebraic approach is the following. Using the rule $P(A \cup B) = P(A) + P(B) - P(A \cap B)$, we get $1 = .8 + .6 - P(A \cap B) \rightarrow P(A \cap B) = .4$. Then, using the rule $P(B) = P(B \cap A) + P(B \cap A')$, we get $P(B \cap A') = .6 - .4 = .2$. Answer: C

2. $P[M] = .4$, $P[M'] = .6$, $P[H] = .4$, $P[H'] = .6$, $P[M \cap H] = .2$, $P[M' \cap H] = .2$.

We wish to find $P[M \cap H']$. From probability rules, we have

$\begin{align*}
.6 &= P[H'] = P[M' \cap H'] + P[M \cap H'], \\
.6 &= P[M'] = P[M' \cap H] + P[M' \cap H'] = .2 + P[M' \cap H'].
\end{align*}$

Thus, $P[M' \cap H'] = .4$ and then $P[M \cap H'] = .2$. The following diagram identifies the component probabilities.

![Diagram showing M, M', H, and H' with intersections and unions]

The calculations above can also be summarized in the following table. The events across the top of the table categorize individuals as male ($M$) or female ($M'$), and the events down the left side of the table categorize individuals as homeowners ($H$) or non-homeowners ($H'$).

<table>
<thead>
<tr>
<th>$P(H')$</th>
<th>$P(M') = 1 - .4 = .6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(H)$</td>
<td>$P(M) = .4$, given $P(M' \cap H) = .2$, given $P(H) - P(M' \cap H) = .4 - .2 = .2$.</td>
</tr>
</tbody>
</table>

Answer: B
3. Since \( C \) and \( D \) have empty intersection, \( P[C \cup D] = P[C] + P[D] \).

Also, since \( A \) and \( B \) are "exhaustive" events (since they are complementary events, their union is the entire sample space, with a combined probability of 
\[
\]

We use the rule \( P[C] = P[C \cap A] + P[C \cap A'] \), and the rule \( P[C|A] = \frac{P[A \cap C]}{P[A]} \) to get
\[
P[C] = P[C|A] \cdot P[A] + P[C|A'] \cdot P[A'] = \frac{1}{2} \cdot \frac{1}{4} + \frac{3}{4} \cdot \frac{3}{4} = \frac{11}{16} \text{ and}
P[D] = P[D|A] \cdot P[A] + P[D|A'] \cdot P[A'] = \frac{1}{4} \cdot \frac{1}{4} + \frac{1}{8} \cdot \frac{3}{4} = \frac{5}{32}.
\]

Then, \( P[C \cup D] = P[C] + P[D] = \frac{27}{32} \).

Answer: C.

4. Actuary 1: Since \( A \) and \( B \) are independent, so are \( A' \) and \( B' \).
\[
P(A' \cap B') = 1 - P[A \cup B] = .3.
\]

But .3 = \( P[A' \cap B'] = P[A'] \cdot P[B'] = (.5)P[B'] \rightarrow P[B'] = .6 \rightarrow P[B] = .4 \).

Actuary 2: .7 = \( P[A \cup B] = P[A] + P[B] = .5 + P[B] \rightarrow P[B] = .2 \).

Absolute difference is \( |.A - .2| = .2 \).

Answer: E

5. We define the following events: \( D \) - a person has the disease, \( TP \) - a person tests positive for the disease. We are given \( P[TP|D] = .85 \) and \( P[TP|D'] = .10 \) and \( P[D] = .01 \). We wish to find \( P[D|TP] \).

Using the formulation for conditional probability we have \( P[D|TP] = \frac{P[D \cap TP]}{P[TP]} \).

But \( P[D \cap TP] = P[TP|D] \cdot P[D] = (.85)(.01) = .0085 \), and \( P[D' \cap TP] = P[TP|D'] \cdot P[D'] = (.10)(.99) = .099 \). Then,
\[
P[TP] = P[D \cap TP] + P[D' \cap TP] = .1075 \rightarrow P[D|TP] = \frac{.0085}{.1075} = .0791.
\]

The following table summarizes the calculations.

| \( P[D] \) = .01 , given | \( ⇒ \) | \( P[D'] = 1 - P[D] = .99 \) |
| \( \downarrow \) | \( \downarrow \) |
| \( P[D \cap TP] \) | \( \downarrow \) |
| \( = P[TP|D] \cdot P[D] = .0085 \) | \( = P[TP|D'] \cdot P[D'] = .099 \) |
| \( P[TP] = P[D \cap TP] + P[D' \cap TP] = .1075 \) |
| \( P[D|TP] = \frac{P[D \cap TP]}{P[TP]} = \frac{.0085}{.1075} = .0791 \). Answer: B |
6. Let $C$ be the event that bowl 1 has 5 black balls after the exchange.
Let $B_1$ be the event that the ball chosen from bowl 1 is black, and let $B_2$ be the event that the ball chosen from bowl 2 is black.

Event $C$ is the disjoint union of $B_1 \cap B_2$ and $B'_1 \cap B'_2$ (black-black or white-white picks), so that $P[C] = P[B_1 \cap B_2] + P[B'_1 \cap B'_2]$.

The black-black combination has probability $(\frac{6}{11})(\frac{1}{2})$, since there is a $\frac{5}{10}$ chance of picking black from bowl 1, and then (with 6 black in bowl 2, which now has 11 balls) $\frac{6}{11}$ is the probability of picking black from bowl 2. This is

$$P[B_1 \cap B_2] = P[B_2|B_1] \cdot P[B_1] = \left(\frac{6}{11}\right)\left(\frac{1}{2}\right).$$

In a similar way, the white-white combination has probability $\left(\frac{6}{11}\right)\left(\frac{1}{2}\right)$.

Then $P[C] = \left(\frac{6}{11}\right)\left(\frac{1}{2}\right) + \left(\frac{6}{11}\right)\left(\frac{1}{2}\right) = \frac{6}{11}$. Answer: C

7. $A_2 =$ event that second person has different birth month from the first.

$P(A_2) = \frac{11}{12} = .9167$.

A$_3$ = event that third person has different birth month from first and second.

Then, the probability that all three have different birthdays is

$$P[A_3 \cap A_2] = P[A_3|A_2] \cdot P(A_2) = \left(\frac{10}{12}\right)\left(\frac{11}{12}\right) = .7639.$$  

A$_4$ = event that fourth person has different birth month from first three.

Then, the probability that all four have different birthdays is

$$P[A_4 \cap A_3 \cap A_2] = P[A_4|A_3 \cap A_2] \cdot P[A_3 \cap A_2] = P[A_4|A_3 \cap A_2] \cdot P(A_3|A_2) \cdot P(A_2) = \left(\frac{9}{12}\right)\left(\frac{10}{12}\right)\left(\frac{11}{12}\right) = .5729.$$  

A$_5$ = event that fifth person has different birth month from first four.

Then, the probability that all five have different birthdays is

$$P[A_5 \cap A_4 \cap A_3 \cap A_2] = P[A_5|A_4 \cap A_3 \cap A_2] \cdot P[A_4 \cap A_3 \cap A_2] = P[A_5|A_4 \cap A_3 \cap A_2] \cdot P[A_4|A_3 \cap A_2] \cdot P[A_3|A_2] \cdot P(A_2)$$

$$= \left(\frac{8}{12}\right)\left(\frac{9}{12}\right)\left(\frac{10}{12}\right)\left(\frac{11}{12}\right) = .3819.$$  

Answer: D

8. $L_1 =$ turn left on trial 1, $R_1 =$ turn right on trial 1, $L_2 =$ turn left on trial 2.

We are given that $P[L_1] = P[R_1] = .5$.

$P[L_2] = P[L_2 \cap L_1] + P[L_2 \cap R_1]$ since $L_1, R_1$ form a partition.

$P[L_2|L_1] = .6$ (if the rat turns left on trial 1 then it gets food and has a .6 chance of turning left on trial 2). Then $P[L_2 \cap L_1] = P[L_2|L_1] \cdot P[L_1] = (.6)(.5) = .3$.

In a similar way, $P[L_2 \cap R_1] = P[L_2|R_1] \cdot P[R_1] = (.8)(.5) = .4$.

Then, $P[L_2] = .3 + .4 = .7$. Answer: D
9. We define the events $A =$ prize door is chosen after contestant switches doors, $B =$ prize door is initial one chosen by contestant. Then $P[B] = \frac{1}{3}$, since each door is equally likely to hold the prize initially. To find $P[A]$ we use the Law of Total Probability.

$$P[A] = P[A|B] \cdot P[B] + P[A|B'] \cdot P[B'] = (0)(\frac{1}{3}) + (1)(\frac{2}{3}) = \frac{2}{3}.$$ If the prize door is initially chosen, then after switching, the door chosen is not the prize door, so that $P[A|B] = 0$. If the prize door is not initially chosen, then since the host shows the other non-prize door, after switching the contestant definitely has the prize door, so that $P[A|B'] = 1$.

Answer: E

10. We are given $P[B] = .05$. We can calculate entries in the following table in the order indicated.

<table>
<thead>
<tr>
<th></th>
<th>$A$</th>
<th>$A'$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$B$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$P[B] = .05$ (given)</td>
<td>$P[A</td>
<td>B] = .95$ (given)</td>
</tr>
<tr>
<td>1.</td>
<td>$P[A \cap B] = P[A</td>
<td>B] \cdot P[B] = .0475$</td>
</tr>
<tr>
<td>$B'$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$P[B']$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$1 - P[B]$ = .95</td>
<td>$P[A \cap B']$</td>
<td>$P[A'</td>
</tr>
<tr>
<td>2.</td>
<td>$P[A'</td>
<td>B'] = P[B'] - P[A \cap B']$</td>
</tr>
<tr>
<td>3.</td>
<td>$P[A \cap B']$</td>
<td>$=.95^2 = .9025$</td>
</tr>
<tr>
<td>4.</td>
<td>$P[A] = P[A \cap B] + P[A \cap B'] = .095$</td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td>$P[B</td>
<td>A] = \frac{P[B \cap A]}{P[A]} = \frac{.0475}{.095} = .5$</td>
</tr>
</tbody>
</table>

11. This is a classical Bayesian probability situation. Let $C$ denote the event that a flood claim occurred. We wish to find $P(H|C)$.

We can summarize the information in the following table, with the order of calculations indicated.

<table>
<thead>
<tr>
<th></th>
<th>$L$ , $P(L) = .8$ (given)</th>
<th>$M$ , $P(M) = .18$ (given)</th>
<th>$H$ , $P(H) = .02$ (given)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C$</td>
<td>$P(C</td>
<td>L) = .001$ (given)</td>
<td>$P(C</td>
</tr>
<tr>
<td>1.</td>
<td>$P(C \cap L)$ = $P(C</td>
<td>L) \cdot P(L) = .0008$</td>
<td>$P(C \cap M)$ = $P(C</td>
</tr>
<tr>
<td>2.</td>
<td>$P(C \cap M)$ = $P(C</td>
<td>M) \cdot P(M) = .0036$</td>
<td>$P(C \cap H)$ = $P(C</td>
</tr>
<tr>
<td>3.</td>
<td>$P(C \cap H)$ = $P(C</td>
<td>H) \cdot P(H) = .005$</td>
<td>$P(C \cap H)$ = $P(C</td>
</tr>
<tr>
<td>4.</td>
<td>$P(C) = P(C \cap L) + P(C \cap M) + P(C \cap H) = .0094$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td>$P(H</td>
<td>C) = \frac{P(H \cap C)}{P(C)} = \frac{.005}{.0094} = .532$</td>
<td>Answer: B</td>
</tr>
</tbody>
</table>
12. We identify the following events:

- the applicant is a smoker, \( S \)
- the applicant is a non-smoker, \( S' \)
- the applicant declares to be a smoker on the application, \( DS \)
- the applicant declares to be non-smoker on the application, \( DN \)

The information we are given is \( P[S] = .3, \ P[NS] = .7, \ P[DN|S] = .4, \ P[DS|NS] = 0 \). We wish to find \( P[NS|DN] = \frac{P[NS \cap DN]}{P[DN]} \).

We calculate \( A = P[DN|S] = \frac{P[DN \cap S]}{P[S]} = \frac{.3}{.3} = P[DN \cap S] = .12 \), and

\( 0 = P[DS|NS] = \frac{P[DS \cap NS]}{P[NS]} = \frac{.7}{.7} = P[DS \cap NS] = 0 \).

Using the rule \( P[A] = P[A \cap B] + P[A \cap B'] \), and noting that \( DS = DN' \) and \( S = NS' \) we have

\[
P[DS \cap S] = P[S] - P[DN \cap S] = .3 - .12 = .18 \), and
\[
P[DN \cap NS] = P[NS] - P[DS \cap NS] = .7 - 0 = .7 \), and
\[
P[DN] = P[DN \cap NS] + P[DN \cap S] = .7 + .12 = .82 .
\]

Then, \( P[NS|DN] = \frac{P[NS \cap DN]}{P[DN]} = \frac{.7}{.82} = \frac{35}{44} \).

These calculations can be summarized in the order indicated in the following table.

<table>
<thead>
<tr>
<th>Step</th>
<th>Expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>( P(NS) = 1 - P(S) = .7 )</td>
</tr>
<tr>
<td>2.</td>
<td>( P(DS</td>
</tr>
<tr>
<td>3.</td>
<td>( P(DN \cap NS) = )</td>
</tr>
<tr>
<td>4.</td>
<td>( P(DN</td>
</tr>
<tr>
<td>5.</td>
<td>( P(DS \cap S) = P(S) - P(DN \cap S) = .3 - .12 = .18 )</td>
</tr>
<tr>
<td>6.</td>
<td>( DS \leftarrow P(DS) = P(DS \cap S) + P(DS \cap NS) = .18 + 0 = .18 )</td>
</tr>
<tr>
<td>7.</td>
<td>( DN \leftarrow P(DN) = 1 - P(DS) = 1 - .18 = .82 )</td>
</tr>
<tr>
<td>8.</td>
<td>( P[NS</td>
</tr>
</tbody>
</table>
13. The probability that an individual will not respond to either the questionnaire or the follow-up letter is \((.5)(.6) = .3\). The probability that all 4 will not respond to either the questionnaire or the follow-up letter is \((.3)^4\).

\[
P[3 \text{ don't respond}] = P[1 \text{ response on 1st round, no additional responses on 2nd round}] \\
+ P[\text{no responses on 1st round, 1 response on 2nd round}] \\
= 4[(.5)^4(.6)^3] + 4[(.5)^4(.6)^3(.4)] = 4(.3)^3(.7) . \quad \text{Then,} \\
P[\text{at least 3 don't respond}] = (.3)^4 + 4(.3)^3(.7) . \quad \text{Answer: A}
\]

14. If 1 fair die is rolled, the probability of rolling a 6 is \(\frac{1}{6}\), and if 2 fair dice are rolled, the probability of rolling a 6 is \(\frac{5}{36}\) (of the 36 possible rolls from a pair of dice, the rolls 1-5, 2-4, 3-3, 4-2 and 5-1 result in a total of 6). Since the coin is fair, the probability of rolling a head or tail is .5. Thus, the probability that \(Y = 6\) is \((.5)(\frac{1}{6}) + (.5)(\frac{5}{36}) = \frac{11}{72}\) .

Answer: C

15. Suppose you have bought a lottery ticket. There are \(\binom{6}{4} = 15\) ways of picking 4 numbers from the 6 numbers on your ticket. Suppose we look at one of those subsets of 4 numbers from your ticket. In order for the winning ticket number to match exactly those 4 of your 6 numbers, the other 2 winning ticket numbers must come from the 43 numbers between 1 and 49 that are not numbers on your ticket. There are \(\binom{43}{2} = \frac{43 \times 42}{2 \times 1} = 903\) ways of doing that, and since there are 15 subsets of 4 numbers on your ticket, there are \(15 \times 903 = 13,545\) ways in which the winning ticket numbers match exactly 3 of your ticket numbers. Since there are a total of 13,983,816 ways of picking 6 out of 49 numbers, your chance of matching exactly 4 of the winning numbers is \(\frac{13,545}{13,983,816} = .00096862\) .

Answer: B
MODELING SECTION 2 - REVIEW OF RANDOM VARIABLES - PART I

Probability, Density and Distribution Functions

This section relates to Chapter 2 of "Loss Models". The suggested time frame for covering this section is two hours. A brief review of some basic calculus relationships is presented first.

LM-2.1 Calculus Review

Natural logarithm and exponential functions

\[ \ln(x) = \log_e(x) \] is the logarithm to the base \( e \);

\[
\begin{align*}
\ln(e) &= 1, & \ln(1) &= 0, & e^0 &= 1, \\
\ln(e^y) &= y, & e^{\ln(x)} &= x, & \ln(a^y) &= y \cdot \ln(a) , \\
\ln(y \cdot z) &= \ln(y) + \ln(z), & \ln\left(\frac{y}{z}\right) &= \ln(y) - \ln(z) , \\
e^{x^2}e^z &= e^{x^2+z} , & (e^x)^w &= e^{xw} .
\end{align*}
\] (2.1)

Differentiation

For the function \( f(x) \), \( f'(x) = \frac{df}{dx} = \lim_{h \to 0} \frac{f(x+h)-f(x)}{h} ; \) (2.2)

Product rule: \( \frac{d}{dx} [g(x)h(x)] = g'(x)h(x) + g(x)h'(x) ; \) (2.3)

Quotient rule: \( \frac{d}{dx} \left[ \frac{g(x)}{h(x)} \right] = \frac{h(x)g'(x) - g(x)h'(x)}{[h(x)]^2} ; \) (2.4)

Chain rule: \( \frac{d}{dx} \ln[g(x)] = \frac{g'(x)}{g(x)} , \quad \frac{d}{dx} [g(x)]^n = n[g(x)]^{n-1} \cdot g'(x) , \quad \frac{d}{dx} a^x = a^x \cdot \ln(a) \) (2.5)

Integration

\[
\int x^n \, dx = \frac{x^{n+1}}{n+1} + c , \quad \int a^x \, dx = \frac{a^x}{\ln(a)} + c , \quad \int \frac{1}{a+bx} \, dx = \frac{1}{b} \cdot \ln[a + bx] + c
\] (2.6)

Integration by parts: \( \int_a^b u(t) \, dv(t) = u(b)v(b) - u(a)v(a) - \int_a^b v(t) \, du(t) \)

for definite integrals, and

\[
\int u \, dv = uv - \int v \, du
\]

for indefinite integrals (this is derived by integrating both sides of the product rule); note that

\[
\frac{dv(t)}{dt} = v'(t) \quad \text{and} \quad \frac{du(t)}{dt} = u'(t) \quad dt ;
\]

\[
\frac{d}{dx} \int_a^x g(t) \, dt = g(x) , \quad \frac{d}{dx} \int_x^b g(t) \, dt = - g(x) \; ,
\] (2.7)

\[
\frac{d}{dx} \int_{h(x)}^{f(x)} g(t) \, dt = g(f(x)) \cdot f'(x) - g(h(x)) \cdot h'(x) ,
\] (2.8)

\[
\int_0^\infty \alpha^n e^{-\kappa x} \, dx = \frac{n!}{\kappa^{n+1}} \quad \text{if } \kappa > 0 \quad \text{and} \quad n \text{ is an integer } \geq 0 .
\] (2.9)
The word "model" used in the context of a loss model, usually refers to the distribution of a loss random variable. Random variables are the basic components used in actuarial modeling. In this section we review the definitions and illustrate the variety of random variables that we will encounter in the Exam C material.

A random variable is a numerical quantity that is related to the outcome of some random experiment on a probability space. For the most part, the random variables we will encounter are the numerical outcomes of some loss related event such as the dollar amount of claims in one year from an auto insurance policy, or the number of tornados that touch down in Kansas in a one year period.

**LM-2.2 Discrete Random Variable**

The random variable $X$ is discrete and is said to have a **discrete distribution** if it can take on values only from a finite or countable infinite sequence (usually the integers or some subset of the integers). As an example, consider the following two random variables related to successive tosses of a coin:

- $X = 1$ if the first head occurs on an even-numbered toss, $X = 0$ if the first head occurs on an odd-numbered toss;
- $Y = n$, where $n$ is the number of the toss on which the first head occurs.

Both $X$ and $Y$ are discrete random variables, where $X$ can take on only the values 0 or 1, and $Y$ can take on any positive integer value.

**Probability function of a discrete random variable**

The probability function (pf) of a discrete random variable is usually denoted $p(x)$ (or $f(x)$), and is equal to $P(X = x)$. As its name suggests, the probability function describes the probability of individual outcomes occurring.

The probability function must satisfy

\[
(i) \quad 0 \leq p(x) \leq 1 \quad \text{for all } x, \quad \text{and} \quad (ii) \quad \sum_x p(x) = 1. \tag{2.10}
\]

For the random variable $X$ above, the probability function is $p(0) = \frac{2}{3}$, $p(1) = \frac{1}{3}$, and for $Y$ it is $p(k) = \frac{1}{2^k}$ for $k = 1, 2, 3, \ldots$.

An event $A$ is a subset of the set of all possible outcomes of $X$, and the probability of event $A$ occurring is $P[A] = \sum_{x \in A} p(x)$.

For $Y$ above, $P[Y \text{ is even}] = P[Y = 2, 4, 6, \ldots] = \frac{1}{2^2} + \frac{1}{2^4} + \frac{1}{2^6} + \cdots = \frac{1}{3} = P(X = 1)$. 


**LM-2.3 Continuous Random Variable**

A continuous random variable usually can assume numerical values from an interval of real numbers, perhaps the whole set of real numbers. As an example, the length of time between successive streetcar arrivals at a particular (in service) streetcar stop could be regarded as a continuous random variable (assuming that time measurement can be made perfectly accurate).

**Probability density function**

A continuous random variable $X$ has a probability density function (pdf) denoted $f(x)$ or $f_X(x)$ (or sometimes denoted $p(x)$), which is a continuous function (except possibly at a finite or countably infinite number of points). For a continuous random variable, we do not describe probability at single points. We describe probability in terms of intervals. In the streetcar example, we would not define the probability that the next street car will arrive in exactly 1.23 minutes, but rather we would define a probability such as the probability that the streetcar will arrive between 1 and 1.5 minutes from now.

Probabilities related to $X$ are found by integrating the density function over an interval.

$$P[X \in (a, b)] = P[a < X < b] \text{ is defined to be equal to } \int_a^b f(x) \, dx.$$  \hspace{1cm} (2.11)

$f(x)$ must satisfy (i) $f(x) \geq 0$ for all $x$, and (ii) $\int_{-\infty}^{\infty} f(x) \, dx = 1$.

Often, the region of non-zero density is a finite interval, and $f(x) = 0$ outside that interval. If $f(x)$ is continuous except at a finite number of points, then probabilities are defined and calculated as if $f(x)$ was continuous everywhere (the discontinuities are ignored).

For example, suppose that $X$ has density function

$$f(x) = \begin{cases} 2x & \text{for } 0 < x < 1 \\ 0 & \text{elsewhere} \end{cases}.$$  

Then $f$ satisfies the requirements for a density function, since

$$\int_{-\infty}^{\infty} f(x) \, dx = \int_0^1 2x \, dx = 1.$$  

Then, for example

$$P[.5 < X < 1] = \int_{.5}^{1} 2x \, dx = x^2 \bigg|_{.5}^{1} = .75.$$  

This is illustrated in the shaded area in the graph below.

```
0.5 1 1.5
0 1 2

\text{Graph:}
\text{\textcolor{blue}{f(x) = 2x}}
```

For a continuous random variable $X$,

$$P[a < X < b] = P[a \leq X < b] = P[a < X \leq b] = P[a \leq X \leq b],$$

so that when calculating the probability for a continuous random variable on an interval, it is irrelevant whether or not the endpoints are included. For a continuous random variable, $P[X = a] = 0$; non-zero probabilities only exist over an interval, not at a single point.
LM-2.4 Mixed Distribution
A random variable may have some points with non-zero probability mass and with a continuous pdf elsewhere. Such a distribution may be referred to as a mixed distribution, but the general notion of mixtures of distributions will be covered later. The sum of the probabilities at the discrete points of probability plus the integral of the density function on the continuous region for $X$ must be 1. For example, suppose that $X$ has probability of .5 at $X = 0$, and $X$ is a continuous random variable on the interval $(0, 1)$ with density function $f(x) = x$ for $0 < x < 1$, and $X$ has no density or probability elsewhere. This satisfies the requirements for a random variable since the total probability is

$$P[X = 0] + \int_0^1 f(x) \, dx = .5 + \int_0^1 x \, dx = .5 + .5 = 1.$$  

Then,

$$P[0 < X < .5] = \int_0^5 x \, dx = .125,$$

and

$$P[0 \leq X < .5] = P[X = 0] + P[0 < X < .5] = .5 + .125 = .625.$$  

Notice that for this random variable $P[0 < X < .5] \neq P[0 \leq X < .5]$ because there is a probability mass at $X = 0$.

LM-2.5 Cumulative Distribution Function, Survival Function and Hazard Function
Given a random variable $X$, the cumulative distribution function of $X$ (also called the distribution function, or cdf) is $F(x) = P[X \leq x]$ (also denoted $F_X(x)$).

The cdf $F(x)$ is the "left-tail" probability, or the probability to the left of and including $x$.

The survival function is the complement of the distribution function,

$$S(x) = 1 - F(x) = P[X > x].$$  \hspace{1cm} (2.12)

The event $X > x$ is referred to as a "tail" or right tail of the distribution.

For any cdf $P[a < X \leq b] = F(b) - F(a)$, $\lim_{x \to \infty} F(x) = 1$, $\lim_{x \to -\infty} F(x) = 0$.  \hspace{1cm} (2.13)

For a discrete random variable with probability function $p(x)$, $F(x) = \sum_{w \leq x} p(w)$, and in this case $F(x)$ is a "step function" (see Example LM2-1 below); it has a jump (or step increase) at each point with non-zero probability, while remaining constant until the next jump. Note that for a discrete random variable, $F(x)$ includes the probability at the point $x$ as well as the total probabilities of all the points to the left of $x$.

If $X$ has a continuous distribution with density function $f(x)$, then

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(t) \, dt \text{ and } S(x) = P(X > x) = \int_x^{\infty} f(t) \, dt$$  \hspace{1cm} (2.14)

and $F(x)$ is a continuous, differentiable, non-decreasing function such that

$$\frac{d}{dx} F(x) = F'(x) = - S'(x) = f(x).$$  

Also, for a continuous random variable, the hazard rate or failure rate is

$$h(x) = \frac{f(x)}{1 - F(x)} = \frac{f(x)}{S(x)} = - \frac{d}{dx} \ln S(x).$$  \hspace{1cm} (2.15)

If $X$ is continuous and $X \geq 0$, then the survival function satisfies $S(0) = 1$ and $S(x) = e^{-\int_0^x h(t) \, dt}$. The cumulative hazard function is

$$H(x) = \int_0^x h(t) \, dt.$$  \hspace{1cm} (2.16)
If $X$ has a mixed distribution with some discrete points and some continuous regions, then $F(x)$ is continuous except at the points of non-zero probability mass, where $F(x)$ will have a jump.

The region of positive probability of a random variable is called the **support** of the random variable.

**LM-2.6 Examples of Distribution Functions**

The following examples illustrate the variety of distribution functions that can arise from random variables. The support of a random variable is the set of points over which there is positive probability or density.

**Example LM2-1:**

Finite Discrete Random Variable (finite support)

$W =$ number turning up when tossing one fair die, so $W$ has probability function

$$p_W(w) = P[W = w] = \frac{1}{6} \text{ for } w = 1, 2, 3, 4, 5, 6 .$$

$$F_W(w) = P[W \leq w] = \begin{cases} 
0 & \text{if } w < 1 \\
\frac{1}{6} & \text{if } 1 \leq w < 2 \\
\frac{2}{6} & \text{if } 2 \leq w < 3 \\
\frac{3}{6} & \text{if } 3 \leq w < 4 \\
\frac{4}{6} & \text{if } 4 \leq w < 5 \\
\frac{5}{6} & \text{if } 5 \leq w < 6 \\
1 & \text{if } w \geq 6 
\end{cases}$$

The graph of the cdf is a step-function that increases at each point of probability by the amount of probability at that point (all 6 points have probability $\frac{1}{6}$ in this example). Since the support of $W$ is finite (the support is the set of integers from 1 to 6), $F_W(w)$ reaches 1 at the largest point $W = 6$ (and stays at 1 for all $w \geq 6$).
Example LM2-2:
Infinite Discrete Random Variable (infinite support)

\( X = \) number of successive independent tosses of a fair coin until the first head turns up.

\( X \) can be any integer \( \geq 1 \), and the probability function of \( X \) is \( p_X(x) = \frac{1}{2^x} \).

The cdf is 
\[
F_X(x) = \sum_{k=1}^{x} \frac{1}{2^k} = 1 - \frac{1}{2^x} \quad \text{for} \quad x = 1, 2, 3, \ldots.
\]

The graph of the cdf is a step-function that increases at each point of probability by the amount of probability at that point. Since the support of \( X \) is infinite (the support is the set of integers \( \geq 1 \) ) \( F_X(x) \) never reaches 1, but approaches 1 as a limit as \( x \to \infty \). The graph of \( F_X(x) \) is
Example LM2-3:
Continuous Random Variable on a Finite Interval

\( Y \) is a continuous random variable on the interval \((0, 1)\) with density function

\[
 f_Y(y) = \begin{cases} 
 3y^2 & \text{for } 0 < y < 1 \\
 0, & \text{elsewhere}
\end{cases}
\]

Then \( F_Y(y) = \begin{cases} 
 0 & \text{if } y < 0 \\
 y^3 & \text{if } 0 \leq y < 1 \\
 1 & \text{if } y \geq 1
\end{cases} \)

![Graph of f(y) and F(y) for Example LM2-3](image)

Example LM2-4:
Continuous Random Variable on an Infinite Interval

\( U \) is a continuous random variable on the interval \((0, \infty)\) with density function

\[
 f_U(u) = \begin{cases} 
 ae^{-au} & \text{for } a > 0 \\
 0, & \text{for } u \leq 0
\end{cases}
\]

Then \( F_U(u) = \begin{cases} 
 0 & \text{for } u \leq 0 \\
 1 - (1 + u)e^{-au}, & \text{for } a > 0
\end{cases} \)

![Graph of f(u) and F(u) for Example LM2-4](image)
Example LM2-5:
Mixed Random Variable
\( Z \) has a mixed distribution on the interval \([0, 1]\). \( Z \) has probability of .5 at \( Z = 0 \), and \( Z \) has density function \( f_Z(z) = z \) for \( 0 < z < 1 \), and \( Z \) has no density or probability elsewhere. Then,
\[
F_Z(z) = \begin{cases} 
0 & \text{if } z < 0 \\
.5 & \text{if } z = 0 \\
.5 + \frac{1}{2}z^2 & \text{if } 0 < z < 1 \\
1 & \text{if } z \geq 1 
\end{cases}
\]

LM-2.7 The Empirical Distribution For a Data Sample
Given a sample of \( n \) observations from a random variable \( X \), say \( x_1, x_2, \ldots, x_n \), the empirical distribution or empirical model for the sample is the discrete random variable which takes on the values \( x_1, x_2, \ldots, x_n \) with a probability of \( \frac{1}{n} \) for each point. This is material in Section 4.2.4 of "Loss Models".

If \( x_i = c \) for \( k \) distinct \( x \)'s (\( k \) repeated values), then the probability function of the empirical distribution at \( c \) is \( p_n(c) = \frac{k}{n} \).

For instance, suppose we have the following sample of size \( n = 6 \):
\[
x_1 = 2, \; x_2 = 3, \; x_3 = 5, \; x_4 = 5, \; x_5 = 6, \; x_6 = 8.
\]

The empirical distribution has probability function
\[
p_6(2) = \frac{1}{6}, \; p_6(3) = \frac{1}{6}, \; p_6(5) = \frac{2}{6}, \; p_6(6) = \frac{1}{6}, \; p_6(8) = \frac{1}{6}.
\]

Concepts related to random samples will be considered in detail later in the study guide.
**LM-2.8 Gamma Function, Incomplete Gamma Function and Incomplete Beta Function**

Many of the continuous distributions described in the Exam C Tables make reference to the gamma function and the incomplete gamma function. The definitions of these functions are

- gamma function:  
  \[
  \Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x} \, dx \quad \text{for} \quad \alpha > 0
  \]

- incomplete gamma function:  
  \[
  \Gamma(\alpha; x) = \frac{1}{\Gamma(\alpha)} \cdot \int_x^\infty y^{\alpha-1}e^{-y} \, dy \quad \text{for} \quad \alpha > 0, \ x > 0
  \]

Some important points to note about these functions are the following:

- if \( n \) is an integer and \( n \geq 1 \), then  
  \[
  \Gamma(n) = (n-1)!
  \]

-  
  \[
  \Gamma(\alpha + 1) = \alpha \cdot \Gamma(\alpha) \quad \text{and} \quad \Gamma(\alpha + k) = (\alpha + k - 1)(\alpha + k - 2) \cdots \alpha \cdot \Gamma(\alpha)
  \]
  for any \( \alpha > 0 \) and integer \( k \geq 1 \)

-  
  \[
  \int_0^\infty x^k e^{-cx} \, dx = \frac{\Gamma(k+1)}{c^{k+1}} \quad \text{for} \quad k \geq 0 \quad \text{and} \quad c > 0 \quad \text{(use substitution} \ u = cx) \]

-  
  \[
  \int_0^\infty \frac{1}{x^k} e^{-c/x} \, dx = \frac{\Gamma(k-1)}{c^{k-1}} \quad \text{for} \quad k > 1 \quad \text{and} \quad c > 0 \quad \text{(use substitution} \ u = \frac{c}{x}) \]

Some of the table distributions make reference to the incomplete beta function, which is defined as follows:

- \( \beta(a, b; x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^x t^{a-1}(1-t)^{b-1} \, dt \quad \text{for} \quad 0 \leq x \leq 1, a, b > 0 \).

References to the gamma function have been rare and the incomplete functions have not been referred to on the released exams. It is useful to remember the integral relationship  
\[
\int_0^\infty x^k e^{-cx} \, dx = \frac{\Gamma(k+1)}{c^{k+1}},
\]
particularly in the case in which \( k \) is a non-negative integer.

In that case, we get  
\[
\int_0^\infty x^k e^{-cx} \, dx = \frac{k!}{c^{k+1}},
\]
which can occasionally simplify integral relationships. This relationship is embedded in the definition of the gamma distribution in the Exam C Table.

The pdf of the gamma distribution with parameters \( \alpha \) and \( \theta \) is  
\[
f(t) = \frac{t^{\alpha-1}e^{-t/\theta}}{\theta^\alpha \Gamma(\alpha)},
\]
defined on the interval \( t > 0 \). This means that  
\[
\int_0^\infty t^{\alpha-1}e^{-t/\theta} \, dt = 1,
\]
which can be reformulated as  
\[
\int_0^\infty \theta^\alpha e^{-t/\theta} \, dx = \theta^\alpha \Gamma(\alpha).
\]

If we let \( \theta = \frac{1}{c} \) and \( k = \alpha - 1 \), we get the relationship  
\[
\int_0^\infty x^k e^{-cx} \, dx = \frac{\Gamma(k+1)}{c^{k+1}}.
\]

Looking at the various continuous distributions in the Exam C table gives some hints at calculating a number of integral forms. For instance, the pdf of the beta distribution with parameters \( a, b, \theta = 1 \) is  
\[
f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \cdot x^{a-1}(1-x)^{b-1} \quad \text{for} \quad 0 < x < 1.
\]

Therefore,  
\[
\int_0^1 \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \cdot x^{a-1}(1-x)^{b-1} \, dx = 1,
\]
from which we get  
\[
\int_0^1 x^{a-1}(1-x)^{b-1} \, dx = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}.
\]
1. Let $X$ be a discrete random variable with probability function
   \[ P[X = x] = \frac{2}{3^x} \quad \text{for} \quad x = 1, 2, 3, \ldots \]
   What is the probability that $X$ is even?

   A) $\frac{1}{4}$  B) $\frac{2}{7}$  C) $\frac{1}{3}$  D) $\frac{2}{3}$  E) $\frac{3}{4}$

2. For a certain discrete random variable on the non-negative integers, the probability function satisfies the relationships
   \[ P(0) = P(1) \quad \text{and} \quad P(k + 1) = \frac{1}{k} \cdot P(k) \quad \text{for} \quad k = 1, 2, 3, \ldots \]
   Find $P(0)$.

   A) $\ln e$  B) $e - 1$  C) $(e + 1)^{-1}$  D) $e^{-1}$  E) $(e - 1)^{-1}$

3. Let $X$ be a continuous random variable with density function
   \[ f(x) = \begin{cases} 
   6x(1-x) & \text{for} \quad 0 < x < 1 \\
   0 & \text{otherwise}
   \end{cases} \]
   Calculate $P[|X - \frac{1}{2}| > \frac{1}{4}]$.

   A) .0521  B) .1563  C) .3125  D) .5000  E) .8000

4. Let $X$ be a random variable with distribution function
   \[ F(x) = \begin{cases} 
   0 & \text{for} \quad x < 0 \\
   \frac{x}{8} & \text{for} \quad 0 \leq x < 1 \\
   \frac{1}{4} + \frac{x}{8} & \text{for} \quad 1 \leq x < 2 \\
   \frac{3}{4} + \frac{x}{12} & \text{for} \quad 2 \leq x < 3 \\
   1 & \text{for} \quad x \geq 3
   \end{cases} \]
   Calculate $P[1 \leq X \leq 2]$.

   A) $\frac{1}{8}$  B) $\frac{3}{8}$  C) $\frac{7}{16}$  D) $\frac{13}{24}$  E) $\frac{10}{24}$

5. Let $X_1$, $X_2$ and $X_3$ be three independent continuous random variables each with density function
   \[ f(x) = \begin{cases} 
   \sqrt{2} - x & \text{for} \quad 0 < x < \sqrt{2} \\
   0 & \text{otherwise}
   \end{cases} \]

   What is the probability that exactly 2 of the 3 random variables exceeds 1?

   A) $\frac{3}{2} - \sqrt{2}$  B) $3 - 2\sqrt{2}$  C) $3(\sqrt{2} - 1)(2 - \sqrt{2})^2$
   D) $(\frac{3}{2} - \sqrt{2})(\sqrt{2} - \frac{1}{2})$  E) $3(\frac{3}{2} - \sqrt{2})^2(\sqrt{2} - \frac{1}{2})$
6. Let $X_1$, $X_2$ and $X_3$ be three independent, identically distributed random variables each with density function $f(x) = \begin{cases} 3x^2 & \text{for } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$. Let $Y = \max\{X_1, X_2, X_3\}$. Find $P[Y > \frac{1}{2}]$.
   A) $\frac{1}{64}$  B) $\frac{37}{64}$  C) $\frac{343}{512}$  D) $\frac{7}{8}$  E) $\frac{511}{512}$

7. Let the distribution function of $X$ for $x > 0$ be $F(x) = 1 - \sum_{k=0}^{\infty} \frac{x^k e^{-x}}{k!}$. What is the density function of $X$ for $x > 0$?
   A) $e^{-x}$  B) $\frac{x^2 e^{-x}}{2}$  C) $\frac{x^3 e^{-x}}{6}$  D) $\frac{x^3 e^{-x}}{6} - e^{-x}$  E) $\frac{x^3 e^{-x}}{6} + e^{-x}$

8. Let $X$ have the density function $f(x) = \frac{3x^2}{\theta^3}$ for $0 < x < \theta$, and $f(x) = 0$, otherwise. If $P[X > 1] = \frac{7}{8}$, find the value of $\theta$.
   A) $\frac{1}{2}$  B) $\left(\frac{7}{8}\right)^{1/3}$  C) $\left(\frac{8}{7}\right)^{1/3}$  D) $2^{1/3}$  E) 2

9. A large wooden floor is laid with strips 2 inches wide and with negligible space between strips. A uniform circular disk of diameter 2.25 inches is dropped at random on the floor. What is the probability that the disk touches three of the wooden strips?
   A) $\frac{1}{\sqrt{\pi}}$  B) $\frac{1}{\pi}$  C) $\frac{1}{4}$  D) $\frac{1}{8}$  E) $\frac{1}{\pi^2}$

10. If $X$ has a continuous uniform distribution on the interval from 0 to 10, then what is $P[X + \frac{10}{X} > 7]$?
    A) $\frac{3}{10}$  B) $\frac{31}{70}$  C) $\frac{1}{2}$  D) $\frac{39}{70}$  E) $\frac{7}{10}$

11. For a loss distribution where $x \geq 2$, you are given:
   i) The hazard rate function: $h(x) = \frac{x^2}{2x}$, for $x \geq 2$
   ii) A value of the distribution function: $F(5) = 0.84$

   Calculate $z$.
   A) 2  B) 3  C) 4  D) 5  E) 6
12. A pizza delivery company has purchased an automobile liability policy for its delivery drivers from the same insurance company for the past five years. The number of claims filed by the pizza delivery company as the result of at-fault accidents caused by its drivers is shown below:

<table>
<thead>
<tr>
<th>Year</th>
<th>Claims</th>
</tr>
</thead>
<tbody>
<tr>
<td>2002</td>
<td>4</td>
</tr>
<tr>
<td>2001</td>
<td>1</td>
</tr>
<tr>
<td>2000</td>
<td>3</td>
</tr>
<tr>
<td>1999</td>
<td>2</td>
</tr>
<tr>
<td>1998</td>
<td>15</td>
</tr>
</tbody>
</table>

Calculate the skewness of the empirical distribution of the number of claims per year.

A) Less than 0.50
B) At least 0.50, but less than 0.75
C) At least 0.75, but less than 1.00
D) At least 1.00, but less than 1.25
E) At least 1.25
MODELING - PROBLEM SET 2 SOLUTIONS

1. \[ P[X \text{ is even}] = P[X = 2] + P[X = 4] + P[X = 6] + \cdots \]
   \[ = \frac{2}{3} \cdot \left[ \frac{1}{3} + \frac{1}{3^2} + \frac{1}{3^3} + \cdots \right] = \frac{2}{3} \cdot \frac{1}{1 - \frac{1}{3^2}} = \frac{1}{4} . \]
   Answer: A

2. \[ P(2) = P(1) = P(0) , \quad P(3) = \frac{1}{2} \cdot P(2) = \frac{1}{2} \cdot P(0) , \quad \ldots \quad P(k) = \frac{1}{(k-1)!} \cdot P(0) . \]
   The probability function must satisfy the requirement \[ \sum_{i=0}^{\infty} P(i) = 1 \] so that
   \[ P(0) + \sum_{i=1}^{\infty} \frac{1}{(i-1)!} \cdot P(0) = P(0)(1 + e) = 1 \]
   (this uses the series expansion for \( e^x \) at \( x = 1 \)). Then, \( P(0) = \frac{1}{e+1} . \)
   Answer: C

3. \[ P\left[ \left| X - \frac{1}{2} \right| \leq \frac{1}{4} \right] = P\left[ -\frac{1}{4} \leq X - \frac{1}{2} \leq \frac{1}{4} \right] = P\left[ \frac{1}{4} \leq X \leq \frac{3}{4} \right] = \int_{\frac{1}{4}}^{\frac{3}{4}} 6x(1-x) \, dx \]
   \[ = 0.6875 \to P\left[ \left| X - \frac{1}{2} \right| < \frac{1}{4} \right] = 1 - P\left[ \left| X - \frac{1}{2} \right| \geq \frac{1}{4} \right] = 0.3125 . \]
   Answer: C

4. \[ P[1 \leq X \leq 2] = P[X \leq 2] - P[X < 1] = F(2) - \lim_{x \to 1} F(x) = \frac{11}{12} - \frac{1}{8} = \frac{10}{24} . \]
   Answer: E

5. \[ P[X \leq 1] = \int_{0}^{1} (\sqrt{2} - x) \, dx = \sqrt{2} - \frac{1}{2} , \quad P[X > 1] = 1 - P[X \leq 1] = \frac{3}{2} - \sqrt{2} . \]
   With 3 independent random variables, \( X_1, X_2 \) and \( X_3 \), there are 3 ways in which exactly 2 of the \( X_i \)'s exceed 1 (either \( X_1, X_2 \) or \( X_1, X_3 \) or \( X_2, X_3 \)).
   Each way has probability \( (P[X > 1])^2 \cdot P[X \leq 1] = (\frac{3}{2} - \sqrt{2})^2(\sqrt{2} - \frac{1}{2}) \)
   for a total probability of \( 3 \cdot (\frac{3}{2} - \sqrt{2})^2(\sqrt{2} - \frac{1}{2}) \). Answer: E

6. \[ P[Y > \frac{1}{2}] = 1 - P[Y \leq \frac{1}{2}] = 1 - P[(X_1 \leq \frac{1}{2}) \cap (X_2 \leq \frac{1}{2}) \cap (X_3 \leq \frac{1}{2})] \]
   \[ = 1 - (P[X \leq \frac{1}{2}])^3 = 1 - \left[ \int_{0}^{1/2} 3x^2 \, dx \right]^3 = 1 - \left( \frac{1}{8} \right)^3 = \frac{511}{512} . \]
   Answer: E

7. \[ f(x) = F'(x) = -\sum_{k=0}^{3} \frac{kx^{k-1}e^{-x} - x^k e^{-x}}{k!} = -e^{-x} \cdot \frac{3}{k!} \sum_{k=0}^{3} \left[ x^k - kx^{k-1} \right] \]
   \[ = -e^{-x} \cdot \left[ 1 + \frac{x-1}{1} + \frac{x^2 - 2x}{2} + \frac{x^3 - 3x^2}{6} \right] = e^{-x} \frac{x^3}{6} . \]
   Answer: C

8. Since \( f(x) = 0 \) if \( x > \theta \), and since \( P[X > 1] = \frac{7}{8} \), we must conclude that \( \theta > 1 \).
   Then, \( P[X > 1] = \int_{1}^{\theta} f(x) \, dx = \int_{1}^{\theta} \frac{3x^2}{\theta^2} \, dx = 1 - \frac{1}{\theta^2} = \frac{7}{8} \), or equivalently, \( \theta = 2 \). Answer: E
9. Let us focus on the left-most point \( p \) on the disk. Consider two adjacent strips on the floor. Let the interval \([0, 2]\) represent the distance as we move across the left strip from left to right. If \( p \) is between 0 and 1.75, then the disk lies within the two strips.

If \( p \) is between 1.75 and 2, the disk will lie on 3 strips (the first two and the next one to the right). Since any point between 0 and 2 is equally likely as the left most point \( p \) on the disk (i.e. uniformly distributed between 0 and 2) it follows that the probability that the disk will touch three strips is \( \frac{25}{2} = \frac{1}{8} \). Answer: D

10. Since the density function for \( X \) is \( f(x) = \frac{1}{10} \) for \( 0 < x < 10 \), we can regard \( X \) as being positive. Then

\[
P[X + \frac{10}{X} > 7] = P[X^2 - 7X + 10 > 0] = P[(X - 5)(X - 2) > 0] = P[X > 5] + P[X < 2]
\]

(since \((t - 5)(t - 2) > 0\) if either both \( t - 5, t - 2 > 0\) or both \( t - 5, t - 2 < 0\) = \( \frac{5}{10} + \frac{2}{10} = \frac{7}{10} \)). Answer: E

11. The survival function \( S(y) \) for a random variable can be formulated in terms of the hazard rate function: \( S(y) = e^{xp[- \int_{-\infty}^{y} h(x) \, dx]} \).

In this question, \( S(5) = 1 - F(5) = .16 = e^{xp[- \int_{2}^{5} \frac{2}{2x} \, dx]} = e^{xp[- \frac{2}{2} \ln(\frac{5}{2})]} \).

Taking natural log of both sides of the equation results in \(- \frac{2}{2} \ln(\frac{5}{2}) = \ln(.16)\), and solving for \( z \) results in \( z = 2 \). Answer: A

12. The empirical distribution is a five point random variable with probability function

\[
\]

(each observation in an empirical distribution of a data set is given the same probability as the others).

The skewness is \( \frac{E[(X-E(X))^3]}{(Var[X])^{3/2}} \).

For the empirical distribution,

\[
E[X] = (.2)(1 + 2 + 3 + 4 + 15) = 5, \\
E[X^2] = (.2)(1 + 4 + 9 + 16 + 225) = 51, \\
E[(X - 5)^3] = (.2)((-4)^3 + (-3)^3 + (-2)^3 + (-1)^3 + (10)^3) = 180, \\
\]

Skewness = \( \frac{180}{26^{3/2}} = 1.36 \). Answer: E
MODELING SECTION 3: REVIEW OF RANDOM VARIABLES - PART II

This section relates to Sections 3.1 - 3.3 of "Loss Models". The mean and variance of a random variable are two fundamental distribution parameters. In this section we review those and some other important distribution parameters. Chapter 3 of Loss models also introduces deductibles and policy limits. These topics will be considered in detail later in the study guide.

The suggested time frame for this section is 2 hours.

LM-3.1 Expected Value and Other Moments of a Random Variable

For a random variable \( X \), the expected value is denoted \( E[X] \), or \( \mu_X \) or \( \mu \). The expected value of \( X \) is also called the expectation of \( X \), or the mean of \( X \). The expected value is the "average" over the range of values that \( X \) can be, or the "center" of the distribution.

For a discrete random variable, the expected value of \( X \) is

\[
E[X] = \sum_{x} x \cdot p(x),
\]

where the sum is taken over all points \( x \) at which \( X \) has non-zero probability. For instance, if \( X \) is the result of one toss of a fair die, then \( E[X] = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + \cdots + 6 \cdot \frac{1}{6} = \frac{7}{2} \).

For a continuous random variable, the expected value is

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

Although the integral is written with lower limit \(-\infty\) and upper limit \(\infty\), the interval of integration is the interval of support (non-zero-density) for \( X \). For instance, if \( f(x) = 2x \) for \( 0 < x < 1 \), then the mean is \( E[X] = \int_{0}^{1} x \cdot 2x \, dx = \frac{2}{3} \).

If \( h \) is a function, then \( E[h(X)] \) is equal to \( \sum_{x} h(x) \cdot p(x) \) if \( X \) is a discrete random variable, and it is equal to \( \int_{-\infty}^{\infty} h(x) \cdot f(x) \, dx \) if \( X \) is a continuous random variable. \( E[h(X)] \) is the "average" value of \( h(X) \) based on the possible outcomes of random variable \( X \).

The mean of a random variable \( X \) might not exist, it might be \(+\infty\) or \(-\infty\). For example, the continuous random variable \( X \) with

\[
f(x) = \begin{cases} f & \text{for } x \geq 1 \\ 0, & \text{otherwise} \end{cases}
\]

has expected value \( \int_{1}^{\infty} x \cdot \frac{1}{x^2} \, dx = +\infty \).

For any constants \( a_1, a_2 \) and \( b \) and functions \( h_1 \) and \( h_2 \),

\[
E[a_1h_1(X) + a_2h_2(X) + b] = a_1E[h_1(X)] + a_2E[h_2(X)] + b
\]

If \( X \) is a non-negative random variable (defined on \([0, \infty)\) or \((0, \infty)\)) then

\[
E[X] = \int_{0}^{\infty} [1 - F(x)] \, dx = \int_{0}^{\infty} S(x) \, dx
\]

This relationship is valid for any random variable, discrete, continuous or with a mixed distribution. Using the example on the previous page, if \( f(x) = 2x \) for \( 0 < x < 1 \), then

\[
F(x) = \begin{cases} x^2 & \text{for } 0 < x \leq 1 \\ 1, & \text{for } x > 1 \end{cases}, \quad \text{and} \quad \int_{0}^{\infty} [1 - F(x)] \, dx = \int_{0}^{1} [1 - x^2] \, dx = \frac{2}{3} = E[X].
\]

It tends to be more awkward to apply this rule to discrete random variables.

Jensen's Inequality states that if \( g \) is a function such the \( g''(x) \geq 0 \) on the probability space for \( X \), then \( E[g(X)] \geq g(E[X]) \). For example, \( E[X^2] \geq (E[X])^2 \).
Moments of a random variable
If \( n \geq 1 \) is an integer, then the \( n \text{th (raw) moment of } X \) is \( E[X^n] \) (sometimes denoted \( \mu'_n \)) and is

\[
\sum_{x} x^n \cdot p(x) \quad \text{in the discrete case, and } \int_{-\infty}^{\infty} x^n \cdot f(x) \, dx \quad \text{in the continuous case.} \tag{3.5}
\]

If the mean of \( X \) is \( \mu \), then the \( n \text{th central moment of } X \) (about the mean \( \mu \)) is defined to be \( E[(X - \mu)^n] \) and may be denoted \( \mu_n \).

For instance, the 3rd central moment of the fair die toss random variable \( X \) is

\[
E[(X - \frac{7}{2})^3] = (1 - \frac{7}{2})^3 \cdot \frac{1}{6} + (2 - \frac{7}{2})^3 \cdot \frac{1}{6} + \cdots + (6 - \frac{7}{2})^3 \cdot \frac{1}{6} = 0.
\]

The 2nd central moment of the continuous random variable \( X \) pdf \( f(x) = 2x \) for \( 0 < x < 1 \) is

\[
\int_0^1 (x - \frac{2}{3})^2 \cdot 2x \, dx = \frac{1}{18}.
\]

Variance of \( X \)
The variance of \( X \) is denoted \( Var[X] \), \( V[X] \), \( \sigma^2_X \) or \( \sigma^2 \). It is defined to be equal to

\[
Var[X] = E[(X - \mu_X)^2] = E[X^2] - \mu_X^2 = E[X^2] - (E[X])^2. \tag{3.6}
\]

The representation of \( Var[X] \) as \( E[X^2] - (E[X])^2 \) is often the most convenient one to use. The variance is the 2nd central moment of \( X \); \( \mu_2 = Var[X] = \mu_2^2 - \mu^2 \).

Variance is a measure of the "dispersion" of \( X \) about the mean. As the expected value of \( (X - \mu_X)^2 \), the variance is the average squared deviation of \( X \) from its mean \( \mu_X \).

A large variance indicates significant levels of probability or density for points far from \( E[X] \). The variance is always \( \geq 0 \) (the variance of \( X \) is equal to 0 only if \( X \) has a discrete distribution with a single point and probability 1 at that point; in other words, not random at all).

The random variable \( U = \{ \begin{array}{ll} 4 & \text{prob. .5} \\ 6 & \text{prob. .5} \end{array} \) has mean \( E[U] = 5 \) and variance \( Var[U] = 1 \).

The random variable \( W = \{ \begin{array}{ll} 2 & \text{prob. .5} \\ 8 & \text{prob. .5} \end{array} \) has the same mean as \( U \), \( E[W] = 5 \), but has variance \( Var[W] = 9 \). The higher variance of \( W \) is reflected in the further dispersion of the outcomes of \( W \) from the mean 5 as compared to \( U \).

If \( a \) and \( b \) are constants, then \( Var[aX + b] = a^2Var[X] \).

Here is a useful shortcut for finding the variance of a 2-point discrete random variable.

\[
\text{If } X \text{ is the two-point random variable } \begin{cases} \begin{array}{ll} a & \text{Prob. } p \\ b & \text{Prob. } 1 - p \end{array} \end{cases} \text{ then } Var[X] = (b - a)^2 \times p \times (1 - p). \tag{3.7}
\]

Standard deviation of \( X \)
The standard deviation of the random variable \( X \) is the square root of the variance, and is denoted \( \sigma_X = \sqrt{Var[X]} \).
Coefficient of variation
The coefficient of variation of \( X \) is
\[
\frac{\sigma_X}{\mu_X} = \frac{\sqrt{\text{Var}[X]}}{E[X]}.
\] (3.8)

Skewness and kurtosis
The skewness of \( X \) is \( \frac{E[(X - \mu)^3]}{\sigma^3} \), and the kurtosis is \( \frac{E[(X - \mu)^4]}{\sigma^4} \). (3.9)

Skewness measures the symmetry of a random variable; skewness of 0 indicates a distribution which is symmetric around its mean. The fair dice toss random variable has skewness of 0. Kurtosis is a measure of the "peakedness" of a distribution. Higher kurtosis suggests that more of the variance is due to less frequent large deviations, rather than more frequent smaller deviations.

There have been infrequent references to skewness and kurtosis on Exam C.

LM-3.2 Moment generating function of random variable \( X \)
The moment generating function of \( X \) (mgf) is denoted \( M_X(t) \), \( m_X(t) \), \( M(t) \) or \( m(t) \), and it is defined to be \( M_X(t) = E[e^{tX}] \), which is either
\[
\sum_x e^{tx} p(x) \quad \text{if } X \text{ is discrete, or } \int_{-\infty}^{\infty} e^{tx} f(x) \, dx \quad \text{if } X \text{ is continuous.}
\] (3.10)

It is always true that \( M_X(0) = 1 \).

The moment generating function of \( X \) might not exist for all real numbers, but usually exists on some interval of real numbers. The function \( \ln[M_X(t)] \) is called the cumulant generating function. The function \( M_X(ln t) = E[t^X] \) is called the probability generating function and may be denoted \( P_X(t) = E[t^X] \); the probability generating function is usually used in the case of a discrete random variable.

Some properties of moment and probability generating functions
Suppose that for the random variable \( X \), the moment generating function \( M_X(t) \) exists in an interval containing the point \( t = 0 \). Then
\[
\frac{d^n}{dt^n} M_X(t) \bigg|_{t=0} = M_X^{(n)}(0) = E[X^n] = \mu'_n, \quad \text{the } n\text{-th moment of } X, \quad \text{and} \quad (3.11)
\]
\[
\frac{d}{dt} \ln[M_X(t)] \bigg|_{t=0} = \frac{M'_X(0)}{M_X(0)} = E[X], \quad \text{and} \quad \frac{d^2}{dt^2} \ln[M_X(t)] \bigg|_{t=0} = \text{Var}[X].
\] (3.12)

If \( X_1 \) and \( X_2 \) are random variables, and \( M_{X_1}(t) = M_{X_2}(t) \) for all values of \( t \) in an interval containing \( t = 0 \), then \( X_1 \) and \( X_2 \) have identical probability distributions.

If \( X_1, X_2, \ldots, X_n \) are independent random variables and \( S = \sum_{i=1}^{n} X_i \) then
\[
M_S(t) = M_{X_1}(t) \cdot M_{X_2}(t) \cdot \cdots \cdot M_{X_n}(t) = \prod_{i=1}^{n} M_{X_i}(t).
\] (3.13)

If \( X \) has a discrete non-negative integer distribution with \( p_k = P[X = k] \), then the probability generating function is
\[
P_X(t) = p_0 + p_1 \cdot t + p_2 \cdot t^2 + \cdots = \sum_{k=0}^{\infty} p_k \cdot t^k,
\] (3.14)
and \( P_X(0) = p_0 \).
LM-3.3 Percentiles and Quantiles of a distribution
If $0 < p < 1$, then the $100p$-th percentile of the distribution of $X$ is the number $c_p$ which satisfies the following inequalities:

$$F[c_p] = P[X < c_p] \leq p \leq P[X \leq c_p] = F[c_p].$$

(3.15)

For a continuous random variable, it is sufficient to find the $c_p$ for which $P[X \leq c_p] = p$. If $p = .5$, the 50-th percentile of a distribution is referred to as the median of the distribution, it is the point $M$ for which $P[X \leq M] = .5$. The median $M$ is the 50% probability point, half of the distribution probability is to the left of $M$ and half is to the right. The word "quantile" is a general term for the proportion or percent of a distribution below a certain given point and is often used interchangeably with "percentile".

LM-3.4 The mode of a distribution
The mode is any point at which the probability or density function $f(x)$ is maximized. For the fair die toss random variable, each of $x = 1, 2, 3, 4, 5$ or $6$ would satisfy the requirements of being a mode, since the probability at each point is $\frac{1}{6}$, which is the maximum probability of any individual point. For the continuous random variable with $f(x) = 2x$ for $0 < x < 1$, strictly speaking, there is no mode since the upper bound of the density of 2 is never reached. The mode could be described as occurring at 1 as a limit.

Example LM3-1:
Let $X$ equal the number of tosses of a fair die until the first "1" appears. Find $E[X]$.

Solution:
$X$ is a discrete random variable that can take on any integer value $\geq 1$. The probability that the first 1 appears on the $x$-th toss is $f(x) = \left(\frac{5}{6}\right)^{x-1}\left(\frac{1}{6}\right)$ for $x \geq 1$ (x - 1 tosses that are not 1 followed by a 1). This is the probability function of $X$. Then

$$E[X] = \sum_{k=1}^{\infty} k \cdot f(k) = \sum_{k=1}^{\infty} k \cdot \left(\frac{5}{6}\right)^{k-1}\left(\frac{1}{6}\right) = \left(\frac{1}{6}\right)[1 + 2\left(\frac{5}{6}\right) + 3\left(\frac{5}{6}\right)^2 + \cdots].$$

We use the general increasing geometric series relation $1 + 2r + 3r^2 + \cdots = \frac{1}{(1-r)^2}$, so that $E[X] = \left(\frac{1}{6}\right) \cdot \frac{1}{(1-\frac{5}{6})^2} = 6$.

Example LM3-2:
The moment generating function of $X$ is $\frac{\alpha}{\alpha-t}$ for $t < \alpha$, where $\alpha > 0$. Find $Var[X]$.

Solution:

$$Var[X] = E[X^2] - (E[X])^2. \quad E[X] = M_X'(0) = \frac{\alpha}{(\alpha-t)^2}\bigg|_{t=0} = \frac{1}{\alpha},$$

and $E[X^2] = M_X''(0) = \left.\frac{2\alpha}{(\alpha-t)^3}\right|_{t=0} = \frac{2}{\alpha^2}$ \Rightarrow $Var[X] = \frac{2}{\alpha^2} - \left(\frac{1}{\alpha}\right)^2 = \frac{1}{\alpha^2} \cdot$

Alternatively, $ln M_X(t) = ln\left(\frac{\alpha}{\alpha-t}\right) = ln \alpha - ln(\alpha-t) \Rightarrow \frac{d}{dt} ln[M_X(t)] = \frac{1}{\alpha-t}$
and $\frac{d^2}{dt^2} ln[M_X(t)] = \frac{1}{(\alpha-t)^2}$ so that $Var[X] = \left.\frac{d^2}{dt^2} ln[M_X(t)]\right|_{t=0} = \frac{1}{\alpha^2}$. 


Example LM3-3:
Given that the density function of $X$ is $f(x) = \theta e^{-\theta x}$, for $x > 0$, and 0 elsewhere, find the $n$-th moment of $X$, where $n$ is a non-negative integer (assuming that $\theta > 0$).

Solution:
The $n$-th moment of $X$ is $E[X^n] = \int_0^\infty x^n \cdot \theta e^{-\theta x} \, dx$. Applying integration by parts, this can be written as

$$
\int_0^\infty x^n \, d(-e^{-\theta x}) = -x^n e^{-\theta x}\bigg|_{x=0}^{x=\infty} - \int_0^\infty nx^{n-1}e^{-\theta x} \, dx = \int_0^\infty nx^{n-1}e^{-\theta x} \, dx.
$$

Repeatedly applying integration by parts results in

$$
E[X^n] = \frac{n!}{\theta^n}.
$$

It is worthwhile noting the general form of the integral that appears in this example; if $k \geq 0$ is an integer and $\alpha > 0$, then by repeated applications of integration by parts, we have

$$
\int_0^\infty t^k e^{-\alpha t} \, dt = \frac{k!}{\alpha^{k+1}},
$$

so that in this example

$$
\int_0^\infty x^n \theta e^{-\theta x} \, dx = \theta \int_0^\infty x^n e^{-\theta x} \, dx = \theta \cdot \frac{n!}{\theta^{n+1}} = \frac{n!}{\theta^n}.
$$

An alternative solution uses the moment generating function.

$$
M_X(t) = E[e^{tX}] = \int_0^\infty e^{tx} \theta e^{-\theta x} \, dx = \theta \int_0^\infty e^{-(\theta-t)x} \, dx = \frac{\theta}{\theta-t}
$$

(which will be valid for $t < \theta$). Then

$$
M_X(t) = \frac{\theta}{(\theta-t)^n} \text{ so that } E[X] = M_X'(0) = \frac{1}{\theta}, \text{ and } M_X^{(2)}(t) = \frac{2\theta}{(\theta-t)^{n+1}} \text{ so that } E[X^2] = M_X^{(2)}(0) = \frac{2}{\theta^2}.
$$

It can be shown by induction on $n$ that $M_X^{(n)}(t) = \frac{n!}{(\theta-t)^{n+1}}$ so that $E[X^n] = M_X^{(n)}(0) = \frac{n!}{\theta^n}$.

Example LM3-4:
The continuous random variable $X$ has pdf $f(x) = \frac{1}{2} \cdot e^{-|x|}$ for $-\infty < x < \infty$.

Find the 87.5-th percentile of the distribution.

Solution:
The 87.5-th percentile is the number $b$ for which

$$
P[X \leq b] = \int_{-\infty}^b f(x) \, dx = \int_{-\infty}^b \frac{1}{2} \cdot e^{-|x|} \, dx.
$$

Note that this distribution is symmetric about 0, since $f(-x) = f(x)$, so the mean and median are both 0. Thus, $b > 0$, and so

$$
\int_{-\infty}^{0} \frac{1}{2} \cdot e^{-|x|} \, dx = \int_{0}^{b} \frac{1}{2} \cdot e^{-|x|} \, dx + \int_{0}^{b} \frac{1}{2} \cdot e^{-|x|} \, dx = .5 + \int_{0}^{b} \frac{1}{2} \cdot e^{-x} \, dx = .5 + \frac{1}{2}(1 - e^{-b}) = .875 \rightarrow b = -\ln(.25) = \ln 4.
$$

Example LM3-5:
$X$ is the outcome after 1 toss of a fair die. Find the 25-th percentile of $X$.

Solution:
The distribution function of $X$ was found in Example LM2-1 above. We see that

$$
P(2^-) = P[X < 2] = \frac{1}{6} \leq .25 \leq \frac{2}{6} = P[X \leq 2] = F(2).
$$

Any number other than 2 will not satisfy one side of the inequalities. The 25-th percentile is 2. Note that 2 would also be the 20-th and the 30-th percentile; 2 would be any percentile from 16.7-th to 33.3-rd.
LM-3.5 Normal Distribution and Normal Approximation

The standard normal distribution, $X \sim N(0, 1)$, has a mean of 0 and variance of 1. A table of probabilities for the standard normal distribution is provided on the exam. The density function is $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ for $-\infty < x < \infty$. The density function has the following graph. The shaded area is $P[X \leq x]$, which is denoted $\Phi(x)$.

![Normal Distribution Graph]

Normal Distribution Table

Entries represent the area under the standardized normal distribution from $-\infty$ to $z$, $Pr(Z < z)$. The value of $z$ to the first decimal is given in the left column. The second decimal place is given in the top row.

<table>
<thead>
<tr>
<th>z</th>
<th>0.00</th>
<th>0.01</th>
<th>0.02</th>
<th>0.03</th>
<th>0.04</th>
<th>0.05</th>
<th>0.06</th>
<th>0.07</th>
<th>0.08</th>
<th>0.09</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>0.5000</td>
<td>0.5040</td>
<td>0.5080</td>
<td>0.5120</td>
<td>0.5160</td>
<td>0.5199</td>
<td>0.5239</td>
<td>0.5279</td>
<td>0.5319</td>
<td>0.5359</td>
</tr>
<tr>
<td>0.1</td>
<td>0.5398</td>
<td>0.5438</td>
<td>0.5478</td>
<td>0.5517</td>
<td>0.5557</td>
<td>0.5596</td>
<td>0.5636</td>
<td>0.5675</td>
<td>0.5714</td>
<td>0.5753</td>
</tr>
<tr>
<td>0.2</td>
<td>0.5793</td>
<td>0.5832</td>
<td>0.5871</td>
<td>0.5910</td>
<td>0.5948</td>
<td>0.5987</td>
<td>0.6026</td>
<td>0.6064</td>
<td>0.6103</td>
<td>0.6141</td>
</tr>
<tr>
<td>0.3</td>
<td>0.6179</td>
<td>0.6217</td>
<td>0.6255</td>
<td>0.6293</td>
<td>0.6331</td>
<td>0.6368</td>
<td>0.6406</td>
<td>0.6443</td>
<td>0.6480</td>
<td>0.6517</td>
</tr>
<tr>
<td>0.4</td>
<td>0.6554</td>
<td>0.6591</td>
<td>0.6628</td>
<td>0.6664</td>
<td>0.6700</td>
<td>0.6736</td>
<td>0.6772</td>
<td>0.6808</td>
<td>0.6844</td>
<td>0.6897</td>
</tr>
<tr>
<td>0.5</td>
<td>0.6915</td>
<td>0.6950</td>
<td>0.6985</td>
<td>0.7019</td>
<td>0.7054</td>
<td>0.7086</td>
<td>0.7123</td>
<td>0.7157</td>
<td>0.7190</td>
<td>0.7224</td>
</tr>
<tr>
<td>0.6</td>
<td>0.7257</td>
<td>0.7291</td>
<td>0.7324</td>
<td>0.7357</td>
<td>0.7389</td>
<td>0.7422</td>
<td>0.7454</td>
<td>0.7486</td>
<td>0.7517</td>
<td>0.7549</td>
</tr>
<tr>
<td>0.7</td>
<td>0.7580</td>
<td>0.7611</td>
<td>0.7642</td>
<td>0.7673</td>
<td>0.7704</td>
<td>0.7734</td>
<td>0.7764</td>
<td>0.7794</td>
<td>0.7823</td>
<td>0.7852</td>
</tr>
<tr>
<td>0.8</td>
<td>0.7881</td>
<td>0.7910</td>
<td>0.7939</td>
<td>0.7967</td>
<td>0.7995</td>
<td>0.8023</td>
<td>0.8051</td>
<td>0.8078</td>
<td>0.8106</td>
<td>0.8133</td>
</tr>
<tr>
<td>0.9</td>
<td>0.8159</td>
<td>0.8186</td>
<td>0.8212</td>
<td>0.8238</td>
<td>0.8264</td>
<td>0.8289</td>
<td>0.8315</td>
<td>0.8340</td>
<td>0.8365</td>
<td>0.8389</td>
</tr>
<tr>
<td>1.0</td>
<td>0.8413</td>
<td>0.8438</td>
<td>0.8461</td>
<td>0.8485</td>
<td>0.8508</td>
<td>0.8531</td>
<td>0.8554</td>
<td>0.8577</td>
<td>0.8599</td>
<td>0.8621</td>
</tr>
<tr>
<td>1.1</td>
<td>0.8643</td>
<td>0.8666</td>
<td>0.8686</td>
<td>0.8708</td>
<td>0.8729</td>
<td>0.8749</td>
<td>0.8770</td>
<td>0.8790</td>
<td>0.8810</td>
<td>0.8830</td>
</tr>
<tr>
<td>1.2</td>
<td>0.8849</td>
<td>0.8869</td>
<td>0.8888</td>
<td>0.8907</td>
<td>0.8925</td>
<td>0.8944</td>
<td>0.8962</td>
<td>0.8980</td>
<td>0.8997</td>
<td>0.9015</td>
</tr>
<tr>
<td>1.3</td>
<td>0.9032</td>
<td>0.9049</td>
<td>0.9066</td>
<td>0.9082</td>
<td>0.9099</td>
<td>0.9115</td>
<td>0.9131</td>
<td>0.9147</td>
<td>0.9162</td>
<td>0.9177</td>
</tr>
<tr>
<td>1.4</td>
<td>0.9192</td>
<td>0.9207</td>
<td>0.9222</td>
<td>0.9236</td>
<td>0.9251</td>
<td>0.9265</td>
<td>0.9279</td>
<td>0.9292</td>
<td>0.9306</td>
<td>0.9319</td>
</tr>
</tbody>
</table>

| Pr(\(Z < z\)) | 0.800 | 0.850 | 0.900 | 0.950 | 0.975 | 0.990 | 0.995 |

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

This is an excerpt from the Exam C Table provided at the exam. The entries in the table are probabilities \(\Phi(x) = P[X \leq x]\). The 95-th percentile of \(Z\) is 1.645 (sometimes denoted \(z_{0.05}\)) since \(\Phi(1.645) = 0.950\) (the shaded region to the left of \(x = 1.645\) in the graph above). We use the symmetry of the distribution to find \(\Phi(x)\) for negative values of \(x\). For instance,

\[
\Phi(-1) = P[Z \leq -1] = P[Z \geq 1] = 1 - \Phi(1)
\]

since the two regions have the same area (probability).
Notice also that
\[ P[-1.96 \leq Z \leq 1.96] = 0.95 , \]
since
\[ P[Z > 1.96] = 1 - \Phi(1.96) = 0.025 , \]
and this area is deleted from both ends of the curve.

The general form of the normal distribution has mean \( \mu \) and variance \( \sigma^2 \). This is a continuous distribution with a "bell-shaped" density function similar to that of the standard normal, but symmetric around the mean \( \mu \). The median and the mode are also \( \mu \).

Given any normal random variable \( W \sim N(\mu, \sigma^2) \), it is possible to find \( P[r < W < s] \) by first "standardizing", \( X = \frac{W-\mu}{\sigma} \) and then

\[ P[r < W < s] = P\left[ \frac{r-\mu}{\sigma} < \frac{W-\mu}{\sigma} < \frac{s-\mu}{\sigma} \right] = \Phi\left( \frac{s-\mu}{\sigma} \right) - \Phi\left( \frac{r-\mu}{\sigma} \right) . \]  

(3.16)

As an example, suppose that \( W \) has a normal distribution with mean 1 and variance 4. Then
\[ P[W \leq 2.5] = P\left[ \frac{W-1}{\sqrt{4}} \leq \frac{2.5-1}{\sqrt{4}} \right] = P[X \leq .75] = \Phi(0.75) = 0.7734 . \]

The 95-th percentile of \( W \) is \( c \), where
\[ P[W \leq c] = 0.95 \rightarrow P\left[ \frac{W-1}{\sqrt{4}} \leq \frac{c-1}{\sqrt{4}} \right] = \Phi\left( \frac{c-1}{\sqrt{4}} \right) = 0.95 \rightarrow \frac{c-1}{\sqrt{4}} = 1.645 \rightarrow c = 4.29 . \]

The moment generating function of \( W \) is
\[ M_W(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}} . \]  

(3.17)

**LM-3.6 Approximating a Distribution Using a Normal Distribution**

Given a random variable with mean \( \mu \) and variance \( \sigma^2 \), probabilities related to the distribution of \( X \) are sometimes approximated by assuming the distribution of \( X \) is approximately \( N(\mu, \sigma^2) \). If \( X \) is discrete and integer-valued then an "integer correction" should be applied; the probability \( P[n \leq X \leq m] \) is approximated by assuming that \( X \) is normal and then finding the probability \( P[n - \frac{1}{2} \leq X \leq m + \frac{1}{2}] \). This is also sometimes referred to as the "correction for discontinuity". The integer correction should be used on an exam question when the normal approximation is applied to an integer distribution.

If the random variable \( S \) is (approximately) normal, then the transformed variable \( \frac{S-E[S]}{\sqrt{\text{Var}[S]}} \) has a standard normal distribution; a standard normal random variable has mean 0 and variance 1.

As an example, the 95-th percentile of \( S \), say \( c_{0.95} \), can be found by solving the expression
\[ P[S \leq c_{0.95}] = P\left[ \frac{S-E[S]}{\sqrt{\text{Var}[S]}} \leq \frac{c_{0.95}-E[S]}{\sqrt{\text{Var}[S]}} \right] = 0.95 . \]
It follows that
\[
\frac{c_{0.95} - E[S]}{\sqrt{\text{Var}[S]}} = 1.645
\]

is the 95-th percentile of the standard normal distribution (found from the standard normal table which is part of the Exam C exam tables). Thus, if \( E[S] \) and \( \text{Var}[S] \) are known, then using the normal approximation, the approximate percentiles of \( S \) can be found.

If \( Y_1, Y_2, \ldots, Y_k \) are independent and identically distributed random variables with common mean \( E[Y] \) and common variance \( \text{Var}[Y] \), then the Central Limit Theorem of probability states that \( S = \sum_{i=1}^{k} Y_i \) has a distribution which is approximately normal with mean \( kE[Y] \) and variance \( k\text{Var}[Y] \). As \( k \) gets larger, \( S \) approaches a normal distribution. In practice, a value of 30 or so is regarded as "large enough" for the normal distribution to be a reasonable approximation to the distribution of \( S \). This is a justification for using the normal approximation in some circumstances.

**Example LM3-6:**
\( X_1, X_2, \ldots, X_{100} \) are independent random variables each uniformly distributed on the interval \( (0, 1) \). Use the normal approximation to find the 95-th percentile of \( S = \sum_{i=1}^{100} X_i \).

**Solution:**
It is always true that \( E\left[\sum_{i=1}^{100} X_i\right] = \sum_{i=1}^{100} E[X_i] \), so that \( E[S] = 100E[X] = 50 \).

If \( X_1, X_2, \ldots, X_n \) are independent, then \( \text{Var}\left[\sum_{i=1}^{100} X_i\right] = \sum_{i=1}^{100} \text{Var}[X_i] \), so that
\[
\text{Var}[S] = 100\text{Var}[X] = 100\left(\frac{1}{12}\right) = 8.333.
\]
The 95-th percentile \( c \) must satisfy the relationship \( P[S \leq c] = .95 \).

Standardizing the probability results in
\[
P[S \leq c] = P\left[\frac{S-50}{\sqrt{8.333}} \leq \frac{c-50}{\sqrt{8.333}}\right] = \Phi\left(\frac{c-50}{\sqrt{8.333}}\right) = .95 \rightarrow \frac{c-50}{\sqrt{8.333}} = 1.645 \rightarrow c = 54.75 . \]
1. If $X$ is a random variable with density function $f(x) = \begin{cases} 1.4e^{-2x} + 9e^{-3x} & \text{for } x > 0 \\ 0, & \text{otherwise} \end{cases}$, then $E[X] =$

A) $\frac{9}{20}$ B) $\frac{5}{6}$ C) 1 D) $\frac{230}{126}$ E) $\frac{23}{10}$

2. Let $X$ be a continuous random variable with density function $f(x) = \begin{cases} \frac{1}{30}x(1+3x) & \text{for } 1 < x < 3 \\ 0, & \text{otherwise} \end{cases}$. Find $E\left[\frac{1}{X}\right]$.

A) $\frac{1}{12}$ B) $\frac{7}{15}$ C) $\frac{45}{103}$ D) $\frac{11}{20}$ E) $\frac{14}{15}$

3. Let $X$ be a continuous random variable with density function $f(x) = \begin{cases} 2 & \text{for } 0 \leq x \leq 2 \\ 0, & \text{otherwise} \end{cases}$. Find $E[|X - E[X]|]$.

A) 0 B) $\frac{2}{9}$ C) $\frac{32}{81}$ D) $\frac{64}{81}$ E) $\frac{4}{3}$

4. Let $Y$ be a continuous random variable with cumulative distribution function $F(y) = \begin{cases} 0 & \text{for } y \leq a \\ 1 - e^{-\frac{1}{2}(y-a)^2} & \text{otherwise} \end{cases}$, where $a$ is a constant. Find the 75th percentile of $Y$.

A) $F(0.75)$ B) $a - \sqrt{2 \ln 2}$ C) $a + \sqrt{2 \ln 2}$ D) $a - 2\sqrt{\ln 2}$ E) $a + 2\sqrt{\ln 2}$

5. Let $X$ be a continuous random variable with density function $f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{for } x > 0 \\ 0, & \text{otherwise} \end{cases}$.

If the median of this distribution is $\frac{1}{3}$, then $\lambda =$

A) $\frac{1}{3} \ln \frac{1}{2}$ B) $\frac{1}{3} \ln 2$ C) $2 \ln \frac{3}{2}$ D) $3 \ln 2$ E) 3

6. Let $X$ be a continuous random variable with density function $f(x) = \begin{cases} \frac{1}{2}x(4-x) & \text{for } 0 < x < 3 \\ 0, & \text{otherwise} \end{cases}$.

What is the mode of $X$?

A) $\frac{4}{9}$ B) 1 C) $\frac{3}{2}$ D) $\frac{7}{4}$ E) 2
7. A system made up of 7 components with independent, identically distributed lifetimes will operate until any of 1 of the system's components fails. If the lifetime $X$ of each component has density function $f(x) = \begin{cases} \frac{3}{x^4} & \text{for } 1 < x \\ 0, \text{otherwise} \end{cases}$, what is the expected lifetime until failure of the system?

A) 1.02  B) 1.03  C) 1.04  D) 1.05  E) 1.06

8. Let $X$ have the density function $f(x) = \begin{cases} \frac{2x}{k^2} & \text{for } 0 \leq x \leq k \\ 0, \text{otherwise} \end{cases}$.
For what value of $k$ is the variance of $X$ equal to 2?

A) 2  B) 6  C) 9  D) 18  E) 36

9. Two players put one dollar into a pot. They decide to throw a pair of dice alternately. The first one who throws a total of 5 on both dice wins the pot. How much should the player who starts add to the pot to make this a fair game?

A) $\frac{9}{17}$  B) $\frac{8}{17}$  C) $\frac{1}{8}$  D) $\frac{2}{9}$  E) $\frac{8}{10}$

10. If $f(x) = (k + 1)x^2$ for $0 < x < 1$, find the moment generating function of $X$.

A) $\frac{e^t(6 + 6t + 3t^2)}{t^3}$  B) $\frac{e^t(6 - 6t + 3t^2)}{t^3}$  C) $\frac{e^t(6 + 6t + 3t^2)}{t^3} - \frac{6}{t^3}$
D) $\frac{e^t(6 + 6t + 3t^2)}{t^3} + \frac{6}{t^3}$  E) $\frac{e^t(6 - 6t + 3t^2)}{t^3} - \frac{6}{t^3}$

11. If the moment generating function for the random variable $X$ is $M_X(t) = \frac{1}{1 + t}$, find the third moment of $X$ about the point $x = 2$.

A) $\frac{1}{3}$  B) $\frac{2}{3}$  C) $\frac{3}{2}$  D) $-38$  E) $-\frac{19}{3}$

12. Let $X$ be a random variable with moment generating function $M(t) = \left(\frac{2 + e^t}{3}\right)^9$ for $-\infty < x < \infty$. Find the variance of $X$.

A) 2  B) 3  C) 8  D) 9  E) 11

13. If the mean and variance of random variable $X$ are 2 and 8, find the first three terms in the Taylor series expansion of the moment generating function of $X$ about the point $t = 0$.

A) $2t + 2t^2$  B) $1 + 2t + 6t^2$  C) $1 + 2t + 2t^2$  D) $1 + 2t + 4t^2$  E) $1 + 2t + 12t^2$
14. Let $X$ be a random variable with a continuous uniform distribution on the interval $(1, a)$ where $a > 1$. If $E[X] = 6 \cdot Var[X]$, then $a =$

A) 2      B) 3      C) $3\sqrt{2}$      D) 7      E) 8

15. A student received a grade of 80 in a math final where the mean grade was 72 and the standard deviation was $s$. In the statistics final, he received a 90, where the mean grade was 80 and the standard deviation was 15. If the standardized scores (i.e., the scores adjusted to a mean of 0 and standard deviation of 1) were the same in each case, find $s$.

A) 10      B) 12      C) 16      D) 18      E) 20

16. If $X$ has a standard normal distribution and $Y = e^X$, what is the $k$th moment of $Y$?

A) 0      B) 1      C) $e^{k/2}$      D) $e^{k^2/2}$      E) 1 if $k = 2m - 1$ and $e^{(2m-1)/(2m-3)\ldots} - 1$ if $k = 2m$

17. The random variable $X$ has an exponential distribution with mean $1/b$. It is found that $M_X(-b^2) = 0.2$. Find $b$.

A) 1      B) 2      C) 3      D) 4      E) 5

18. Let $X$ be a continuous random variable with density function

$f(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ for $-\infty < x < \infty$. Calculate $E[X|X \geq 0]$.

A) 0      B) $\frac{1}{\sqrt{2\pi}}$      C) $\frac{1}{2}$      D) $\sqrt{\frac{2}{\pi}}$      E) 1

19. $X$ has a lognormal distribution with parameters with a mean of $e^3$ and variance of $e^{10} - e^6$. Find $P[X^2 \leq e^2]$.

A) .66      B) .69      C) .72      D) .75      E) .78

20. If $X$ has a normal distribution with mean 1 and variance 4, then $P[X^2 - 2X \leq 8] =$?

A) .13      B) .43      C) .75      D) .87      E) .93

21. For watches produced by a certain manufacturer:

(i) Lifetimes follow a single-parameter Pareto distribution with $\alpha > 1$ and $\theta = 4$.

(ii) The expected lifetime of a watch is 8 years. Calculate the probability that the lifetime of a watch is at least 6 years.

A) 0.44      B) 0.50      C) 0.56      D) 0.61      E) 0.67
22. The claim amount random variable $B$ has the following distribution function

$$F(x) = \begin{cases} 
0 & x < 0 \\
x/2,000 & 0 \leq x < 1000 \\
.75 & x = 1000 \\
(x + 11,000)/16,000 & 1000 < x < 5000 \\
1 & x \geq 5000 
\end{cases}$$

What is $E[B] + \sqrt{\text{Var}(B)}$?

A) 2400  B) 2450  C) 2500  D) 2550  E) 2600

23. Given $E[X \mid Y = y] = 3y$ and $\text{Var}[X \mid Y = y] = 2$, where $Y$ has an exponential distribution with a mean of $\frac{1}{3}$, what is $\text{Var}[X]$?

A) $\frac{1}{3}$  B) 1  C) $\frac{3}{2}$  D) 2  E) 3

24. $X$ has a gamma distribution with mean 8 and skewness 1. Find the variance of $X$.

A) 4  B) 8  C) 16  D) 32  E) 64

25. For a random variable $X$ such that its df has $F(10) = 0$, $E[(X - 10)^3] \leq (E[X] - 10)^3$.

A) True  B) False

26. A random variable has pdf $f(x) = 2x$ for $0 < x < 1$. Find the 75th percentile of the distribution, $\pi_{.75}$.

A) .750  B) .777  C) .833  D) .866  E) .902

27. You are given:
   (i) $Z_1$ and $Z_2$ are independent $N(0, 1)$ random variables.
   (ii) $a, b, c, d, e, f$ are constants such that not both $e$ and $f$ are 0.
   (iii) $Y = a + bZ_1 + cZ_2$ and $X = d + eZ_1 + fZ_2$

Determine $E(Y \mid X)$.

A) $a$
B) $a + (b + c)(X - d)$
C) $a + (be + ef)(X - d)$
D) $a + [(be + cf)/(e^2 + f^2)]X$
E) $a + [(be + cf)/(e^2 + f^2)](X - d)$
MODELING - PROBLEM SET 3 SOLUTIONS

1. \[ E[X] = \int_{-\infty}^{\infty} x \cdot f(x) \, dx = \int_{0}^{\infty} (1.4xe^{-2x} + .9xe^{-3x}) \, dx \]
\[ = \left. \left( -.7xe^{-2x} - .35e^{-2x} - 3xe^{-3x} - .1e^{-3x} \right) \right|_{x=0}^{x=\infty} = .45. \]

The integrals were found by integration by parts. Note that we could also have used
\[ \int_{0}^{\infty} x^k e^{-ax} \, dx = \frac{k!}{a^{k+1}} \] if \( k \) is an integer \( \geq 0 \), and \( a > 0 \).

Answer: A

2. \[ E\left[ \frac{1}{X} \right] = \int_{1}^{\infty} \frac{1}{x} \cdot \frac{3}{30} x(1 + 3x) \, dx = \frac{7}{15}. \]

Answer: B

3. \[ E[X] = \int_{0}^{2} \frac{x}{2} \, dx = \frac{4}{3}. \]
Then, \( X - E[X] = X - \frac{4}{3} \), which is negative for \( 0 \leq x \leq \frac{4}{3} \)
and is positive for \( \frac{4}{3} \leq x \leq 2 \). Thus, \( |X - E[X]| = \frac{4}{3} - X \) if \( 0 \leq X \leq \frac{4}{3} \) and \( |X - E[X]| = X - \frac{4}{3} \) if \( \frac{4}{3} \leq x \leq 2 \).

Then, \( E[|X - E[X]|] = \int_{0}^{4/3} \left( \frac{4}{3} - x \right) \cdot \frac{x}{2} \, dx + \int_{4/3}^{2} \left( x - \frac{4}{3} \right) \cdot \frac{x}{2} \, dx = \frac{32}{81} \).

Answer: C

4. Let us denote the 75th percentile of \( Y \) by \( c \). Thus, \( F(c) = .75 \), so that \( 1 - e^{-\frac{1}{2}(c-a)^2} = .75 \).

Solving this equation for \( c \) results in \( e^{-\frac{1}{2}(c-a)^2} = .25 \), or equivalently,
\( \frac{1}{2}(c-a)^2 = \ln 4 \rightarrow c = a + \sqrt{2 \ln 4} = a + 2 \sqrt{\ln 2} \).

Answer: E

5. \[ \int_{0}^{1/3} \lambda e^{-\lambda x} \, dx = 1 - e^{-\lambda/3} = \frac{1}{2} \rightarrow \lambda = 3 \ln 2. \]

Answer: D

6. The mode is the point at which \( f(x) \) is maximized. \( f'(x) = -\frac{1}{9}x + \frac{1}{9}(4 - x) = \frac{4}{9} - \frac{2}{9}x \).

Setting \( f'(x) = 0 \) results in \( x = 2 \). Since \( f''(2) = -\frac{2}{9} < 0 \), that point is a relative maximum.

Answer: E

7. Let \( T \) be the time until failure for the system. In order for the system to not fail by time \( t > 0 \), it must be the case that none of the components have failed by time \( t \).

For a given component, with time until failure of \( W \), \( P[W > t] = \int_{t}^{\infty} \frac{d}{x^2} \, dx = \frac{1}{t^2} \). Thus,
\[ P[T > t] = P[(W_1 > t) \cap (W_2 > t) \cap \cdots \cap (W_7 > t)] = P[W_1 > t] \cdot P[W_2 > t] \cdots P[W_7 > t] = \left( \frac{1}{t^2} \right)^7 \]
(because of independence of the \( W_i \)'s). The cumulative distribution function for \( T \) is
\[ F_T(t) = P[T \leq t] = 1 - P[T > t] = 1 - \frac{1}{t^2} \],
so the density function for \( T \) is \( f_T(t) = \frac{21}{t^2} \).

The expected value of \( T \) is then
\[ E[T] = \int_{1}^{\infty} t \cdot \frac{21}{t^2} \, dt = \frac{21}{20} = 1.05 \text{.} \]

Alternatively, once the cdf of \( T \) is known, since the region of density for \( T \) is \( t > 1 \), the expected value of \( T \) is \( E[T] = 1 + \int_{1}^{\infty} [1 - F_T(t)] \, dt = 1 + \int_{1}^{\infty} \frac{1}{t^2} \, dt = 1 + \frac{1}{20} \).

Answer: D
8. \( E[X] = \int_0^k x \cdot \frac{2x}{k^2} \, dx = \frac{2k^2}{3} \), \( E[X^2] = \int_0^k x^2 \cdot \frac{2x}{k^2} \, dx = \frac{k^2}{2} \)

\( \rightarrow Var[X] = \frac{k^2}{2} - \left( \frac{2k}{3} \right)^2 = \frac{k^2}{18} = 2 \rightarrow k = 6. \) Answer: B

9. Player 1 throws the dice on throws 1, 3, 5, \ldots and the probability that player wins on throw \( 2k + 1 \) is \( \left( \frac{8}{9} \right)^{2k} \cdot \frac{1}{9} \) for \( k = 0, 1, 2, 3, \ldots \) (there is a \( \frac{1}{9} \) probability of throwing a total of 5 on any one throw of the pair of dice). The probability that player 1 wins the pot is \( \frac{1}{9} + \left( \frac{8}{9} \right)^2 \cdot \frac{1}{9} + \left( \frac{8}{9} \right)^4 \cdot \frac{1}{9} + \cdots = \frac{1}{9} \cdot \frac{1}{1-\left( \frac{8}{9} \right)^2} = \frac{9}{17}. \)

Player 2 throws the dice on throws 2, 4, 6, \ldots. The probability that player 2 wins the pot on throw \( 2k \) is \( \left( \frac{8}{9} \right)^{2k-1} \cdot \frac{1}{9} \) for \( k = 1, 2, 3, \ldots \) and the probability that player 2 wins is \( \frac{8}{9} \cdot \frac{1}{9} + \left( \frac{8}{9} \right)^3 \cdot \frac{1}{9} + \left( \frac{8}{9} \right)^5 \cdot \frac{1}{9} + \cdots = \frac{8}{9} \cdot \left( \frac{1}{9} \right) \cdot \frac{1}{1-\left( \frac{8}{9} \right)^2} = \frac{8}{17} = 1 - \frac{9}{17}. \)

If player 1 puts \( 1 + c \) dollars into the pot, then his expected gain is \( 1 \cdot \frac{9}{17} - (1 + c) \cdot \frac{8}{17} \) and player 2's expected gain is \( (1 + c) \cdot \frac{8}{17} - 1 \cdot \frac{9}{17}. \)

In order for the two players to have the same expected gain, we must have \( 1 \cdot \frac{9}{17} - (1 + c) \cdot \frac{8}{17} = 0 \), so that \( c = \frac{1}{8} \). Answer: C

10. Since \( \int_0^1 f(x) \, dx = 1 \), it follows that \( (k + 1) \cdot \frac{1}{3} = 1 \), so that \( k = 2 \), and \( f(x) = 3x^2 \). Then, \( M_X(t) = E[e^{tX}] = \int_0^1 e^{tx} \cdot 3x^2 \, dx \). Applying integration by parts, we have

\[
\int_0^1 e^{tx} \cdot 3x^2 \, dx = \int_0^1 3x^2 \, d\left( \frac{e^{tx}}{t} \right) = \frac{3x^2 e^{tx}}{t} \bigg|_{x=0}^{x=1} - \int_0^1 \frac{6x e^{tx}}{t^2} \, dx.
\]

\[
= 3e^t - \int_0^1 \frac{6x e^{tx}}{t^2} \, dx = 3e^t - \left[ \frac{6x e^{tx}}{t^2} \bigg|_{x=0}^{x=1} - \int_0^1 \frac{6e^{tx}}{t^2} \, dx \right]
\]

\[
= 3e^t - \frac{6e^t}{t^2} + \frac{6(e^t-1)}{t^3} = \frac{e^t(6-6t+3t^2)}{t^3} - \frac{6}{t^3}. \] Answer: E

11. \( E[(X-2)^3] = E[X^3] - 6E[X^2] + 12E[X] - 8 = M_X^{(3)}(0) - 6M_X^{(2)}(0) + 12M_X'(0) - 8. \)

\( M_X(t) = -\frac{1}{(1+t)^4} \rightarrow M_X'(0) = -1, \) \( M_X^{(2)}(t) = \frac{2}{(1+t)^6} \rightarrow M_X''(0) = 2, \)

\( M_X^{(3)}(t) = -\frac{6}{(1+t)^8} \rightarrow M_X^{(3)}(0) = -6. \) Then, \( E[(X-2)^3] = -38. \) Answer: D

12. \( Var[X] = E[X^2] - (E[X])^2, \) and \( E[X] = M'(0), \) \( E[X^2] = M''(0). \)

\( M'(t) = 9 \left( \frac{2+e^t}{3} \right)^8 \cdot \frac{e^t}{3}, \) \( M''(t) = 9 \cdot 8 \left( \frac{2+e^t}{3} \right)^7 \cdot \left( \frac{e^t}{3} \right)^2 + 9 \left( \frac{2+e^t}{3} \right)^8 \cdot \frac{e^t}{3}. \)

Then, \( M'(0) = 3 \) and \( M''(0) = 8 + 3 = 11, \) so that \( Var[X] = 11 - 3^2 = 2. \)

Alternatively, \( Var[X] = \frac{d^2}{dt^2} \ln M(t) \bigg|_{t=0}. \) In this case,

\( \ln M(t) = 9 \ln \left( \frac{2+e^t}{3} \right) = 9 [\ln(2+e^t) - \ln 3], \)

so that \( \frac{d}{dt} \ln M(t) = \frac{9e^t}{2+e^t}, \) and \( \frac{d^2}{dt^2} \ln M(t) = \frac{(2+e^t)(9e^t) - (9e^t)(e^t)}{(2+e^t)^2}, \)

and then \( \frac{d^2}{dt^2} \ln M(t) \bigg|_{t=0} = \frac{3(9) - (9)(1)}{(3)^2} = 2. \) Answer: A
13. \( M(t) = E[e^{tX}] = E[1 + tX + \frac{t^2X^2}{2} + \cdots] = E[1] + t \cdot E[X] + \frac{t^2}{2} \cdot E[X^2] + \cdots \) is given as 2, and \( Var[X] = E[X^2] - (E[X])^2 \) is given to be 8, so that \( E[X^2] = 12 \). Then the first 3 terms of the expansion of \( M(t) \) are \( 1 + 2t + 6t^2 \). Answer: B

14. \( E[X] = \frac{1+a}{2} \) and \( Var[X] = \frac{(a-1)^2}{12} \), so that \( \frac{a+1}{2} = 6 \cdot \frac{(a-1)^2}{12} \rightarrow a^2 - 3a = 0 \rightarrow a = 0, 3 \rightarrow a = 3 \) (since \( a > 0 \)). Answer: B

15. The standardized statistics score is \( \frac{90-80}{15} = \frac{2}{3} \). The standardized math score is \( \frac{80-72}{s} = \frac{8}{s} = \frac{2}{3} \rightarrow s = 12 \). Answer: B

16. The \( k \)th moment of \( Y \) is \( E[Y^k] = E[e^{kX}] = M_X(k) = e^{k^2/2} \) (since \( \mu = 0 \) and \( \sigma^2 = 1 \)). Answer: D

17. \( M_X(t) = \frac{1}{1-t/\mu} \rightarrow M_X(-b^2) = \frac{b}{b^{-\sigma^2}} = \frac{b}{\sigma^2} = \frac{1}{1+b} = .2 \rightarrow b = 4 \). Answer: D

18. \( X \) has a \( N(0, 1) \) distribution, so that the density function of the conditional distribution is \( f(x \mid X > 0) = \frac{f(x)}{P[X > 0]} = \frac{f(x)}{1/2} = 2f(x) \). The conditional expectation is \( \int_0^\infty x \cdot f(x \mid X > 0) \, dx = \int_0^\infty 2x \cdot \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \, dx = -\frac{2}{\sqrt{2\pi}} e^{-x^2/2} \bigg|_{x=0}^{x=\infty} = \frac{2}{\sqrt{2\pi}} = \frac{\sqrt{2}}{\sqrt{\pi}} \). Answer: D

19. A lognormal random variable with parameters \( m \) and \( \sigma^2 \) has mean \( E[X] = e^{m+\frac{1}{2}\sigma^2} = e^3 \), and variance \( Var[X] = (e^{\sigma^2} - 1)e^{m+\sigma^2} = e^{10} - e^6 \).

Then, \( e^{2m+\sigma^2} = (e^{m+\frac{1}{2}\sigma^2})^2 = e^6 \rightarrow e^{\sigma^2} - 1 = e^4 - 1 \rightarrow \sigma^2 = 4 \), \( m = 1 \).

\( W = \ln(X) \sim N(m, \sigma^2) = N(1, 4) \rightarrow P[X \leq e^2] = P[\ln(X) \leq 2] = P[W \leq 2] \).

Then, \( Z = \frac{W-E[W]}{\sqrt{Var[W]}} = \frac{W-1}{2} \) has a standard normal distribution, so that \( P[W \leq 2] = P\left[\frac{W-1}{2} \leq \frac{2-1}{2}\right] = P[Z \leq .5] = .69 \) (from the table of the standard normal distribution). Answer: B

20. Since \( X \sim N(1, 4) \), \( Z = \frac{X-1}{2} \) has a standard normal distribution. The probability in question can be written as

\[
P[X^2 - 2X \leq 8] = P[X^2 - 2X + 1 \leq 9]
= P[(X - 1)^2 \leq 9] = P[-3 \leq X - 1 \leq 3]
= P[-1.5 \leq \frac{X-1}{2} \leq 1.5] = P[-1.5 \leq Z \leq 1.5]
= \Phi(1.5) - [1 - \Phi(1.5)] = .8664 \) (from the standard normal table).

Answer: D
21. \[ E[X] = \frac{\alpha^d}{\alpha - 1} = 4^{\alpha} \frac{\alpha - 1}{\alpha - 1} = 8 \rightarrow \alpha = 2 \rightarrow P[X \geq 6] = 1 - F(6) = (\frac{4}{6})^2 = .44. \] Answer: A

22. The pdf of \( B \) is \( f(x) = 0 \) for \( x < 0 \) and \( f(x) = 0 \) for \( x \geq 5000 \), it is \( f(x) = .0005 \) for \( 0 \leq x \leq 1000 \), and it is \( f(x) = .0000625 \) for \( 1000 < x < 5000 \).

There is a point mass of probability with \( f(x) = .25 \) at \( x = 1000 \) (\( B \) has a mixed distribution).

\[ E[B] = \int_{-\infty}^{\infty} x \cdot (.0005)\,dx + (1000)(.25) + \int_{1000}^{\infty} x \cdot (.0000625)\,dx = 1250, \]
\[ E[B^2] = \int_{-\infty}^{\infty} x^2 \cdot (.0005)\,dx + (1000^2)(.25) + \int_{1000}^{\infty} x^2 \cdot (.0000625)\,dx = 3,000,000 \]
\[ Var[B] = E[B^2] - (E[B])^2 = 1,437,500. \]
\[ E[B] + \sqrt{Var[B]} = 2449. \] Answer: B

23. \[ Var[X] = Var[E[X \mid Y]] + \text{Var}[X \mid Y] = Var[3Y] + E[2] = 9\text{Var}[Y] + 2 = 9 \cdot \frac{1}{9} + 2 = 3, \]

since the variance of an exponential rv is the square of the mean, \( \text{Var}[Y] = \frac{1}{9} \). Answer: E

24. \[ E[X] = \theta \alpha, \text{ Skewness } \frac{E[(X-E[X])^3]}{(E[(X-E[X])]^2)^{3/2}} = \frac{E[X^3]-3E[X^2]E[X]+3E[X]E[X^2]}{(E[X^2]-E[X]^2)^{3/2}} \]
\[ = \frac{\theta^3(\alpha+2)(\alpha+1)\alpha-3\theta^2(\alpha+1)^2\alpha^2+2\alpha^3}{\theta^2\alpha^{3/2}} = \frac{2}{\alpha^{3/2}} = 1 \rightarrow \alpha = 4. \]
\[ E[X] = 8 = \theta \alpha = 4\theta \rightarrow \theta = 2 \rightarrow \text{Var}[X] = \theta^2 \alpha = 16. \] Answer: C

25. Jensen's inequality states that if \( X \) is a random variable and \( h(x) \) is a function such that \( h''(x) \geq 0 \) on the region for which \( X \) has positive probability or density, then \( E[h(X)] \geq h(E[X]) \). Since \( F(10) = P[X \leq 10] = 0 \), it follows that the regions of non-zero density of \( X \) is \( X \geq 10 \). With function \( h(x) = (x-10)^3 \), we have \( h''(x) = 3(x-10)^2 \geq 0 \) (for any \( x \)). From Jensen's inequality it follows that

\[ E[(X-10)^3] = E[h(X)] \geq h(E[X]) = (E[X] - 10)^3. \] Answer: B

26. \[ .75 = P[X < \pi, \tau_5] = \int_0^{\pi, \tau_5} 2\,dx \to \pi, \tau_5 = \sqrt{.75} = .866. \] Answer: D

27. This problem can be solved by eliminating answers based on a careful choice of the constant values. Suppose that \( f = 0 \). Then \( X = d + eZ_1 \), or equivalently, \( Z_1 = (X - d)/e \). Then,

\[ E(Y \mid X) = E[a+bZ_1+eZ_2 \mid X] = E[a+bZ_1+eZ_2 \mid Z_1 = (X-d)/e] = a+b[(X-d)/e] \]

(since \( Z_2 \) does not appear in \( X \), it follows that \( E[Z_2] = 0 \)).

The only answer consistent with this expectation is E, since with \( f = 0 \) Answer E becomes

\[ a + [(be)/(e^2)](X-d) = a + b[(X-d)/e]. \]

There is an alternative solution to this problem. If \( U \) and \( W \) are any two normal random variables with means \( \mu_U \) and \( \mu_W \), and variances \( \sigma_U^2 \) and \( \sigma_W^2 \) and covariance \( \sigma_{UW} \), then it is true that

\[ E[Y \mid X = x] = \mu_Y + \frac{\sigma_{YW}}{\sigma_X} \cdot (x - \mu_X). \]

In this example, \( X \) and \( Y \) are normal with \( \mu_X = d \), \( \mu_Y = a \), \( \sigma_X^2 = e^2 + f^2 \), and \( \sigma_{XY} = Cov[a+bZ_1+cZ_2, d+eZ_1+fZ_2] = be+cf \). Then, \( E[Y \mid X] = a + \frac{be+cf}{e^2+f^2} \cdot (X-d) \).

Answer: E
**MODELING SECTION 4: REVIEW OF RANDOM VARIABLES - PART III**

**Joint, Marginal and Conditional Distributions**

The suggested time frame for this section is 2-3 hours.

**LM-4.1 Joint Distribution of Random Variables X and Y**

A joint distribution of two random variables has a probability function or probability density function \( f(x, y) \) that is a function of two variables (sometimes denoted \( f_{X,Y}(x, y) \)). It is defined over a two-dimensional region. For joint distributions of continuous random variables \( X \) and \( Y \), the region of probability (the probability space) is usually a rectangle or triangle in the \( x-y \) plane.

If \( X \) and \( Y \) are discrete random variables, then \( f(x, y) = P[(X = x) \cap (Y = y)] \) is the joint probability function, and it must satisfy

(i) \( 0 \leq f(x, y) \leq 1 \) and (ii) \( \sum_{x} \sum_{y} f(x, y) = 1 \). \hspace{1cm} (4.1)

If \( X \) and \( Y \) are continuous random variables, then \( f(x, y) \) must satisfy

(i) \( f(x, y) \geq 0 \) and (ii) \( \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \, dy \, dx = 1 \). \hspace{1cm} (4.2)

It is possible to have a joint distribution in which one variable is discrete and one is continuous, or either has a mixed distribution. The joint distribution of two random variables can be extended to a joint distribution of any number of random variables.

If \( A \) is a subset of two-dimensional space, then \( P[(X, Y) \in A] \) is the summation (discrete case) or double integral (continuous case) of \( f(x, y) \) over the region \( A \).

**Example LM4.1:**

\( X \) and \( Y \) are discrete random variables which are jointly distributed with the following probability function \( f(x, y) \) (in table form):

\[
\begin{array}{c|ccc}
X & -1 & 0 & 1 \\
--- & --- & --- & --- \\
1 & \frac{1}{18} & \frac{1}{9} & \frac{1}{6} \\
0 & \frac{1}{9} & 0 & \frac{1}{6} \\
-1 & \frac{1}{6} & \frac{1}{9} & \frac{1}{9} \\
\end{array}
\]

From this table we see, for example, that \( P[X = 0, Y = -1] = f(0, -1) = \frac{1}{9} \).

Find (i) \( P[X + Y = 1] \), (ii) \( P[X = 0] \) and (iii) \( P[X < Y] \).

**Solution:**

(i) We identify the \((x, y)\)-points for which \( X + Y = 1 \), and the probability is the sum of \( f(x, y) \) over those points. The only \( x, y \) combinations that sum to 1 are the points \( (0, 1) \) and \( (1, 0) \). Therefore, \( P[X + Y = 1] = f(0, 1) + f(1, 0) = \frac{1}{9} + \frac{1}{6} = \frac{5}{18} \).

(ii) We identify the \((x, y)\)-points for which \( X = 0 \). These are \((0, -1)\) and \((0, 1)\) (we omit \((0, 0)\) since there is no probability at that point). \( P[X = 0] = f(0, -1) + f(0, 1) = \frac{1}{9} + \frac{1}{9} = \frac{2}{9} \).

(iii) The \((x, y)\)-points satisfying \( X < Y \) are \((−1, 0)\), \((−1, 1)\) and \((0, 1)\).

Then \( P[X < Y] = f(-1, 0) + f(-1, 1) + f(0, 1) = \frac{1}{9} + \frac{1}{18} + \frac{1}{9} = \frac{5}{18} \). \( \mathbb{Q} \)
Example LM4-2:
Suppose that \( f(x, y) = K(x^2 + y^2) \) is the density function for the joint distribution of the continuous random variables \( X \) and \( Y \) defined over the unit square bounded by the points \((0, 0), (1, 0), (1, 1)\) and \((0, 1)\), find \( K \). Find \( P[X + Y \geq 1] \).

Solution:
In order for \( f(x, y) \) to be a properly defined joint density, the (double) integral of the density function over the region of density must be 1, so that \( 1 = \int_0^1 \int_0^1 K(x^2 + y^2) \, dy \, dx = K \cdot \frac{2}{3} \Rightarrow K = \frac{3}{2} \)
\( \Rightarrow f(x, y) = \frac{3}{2}(x^2 + y^2) \) for \( 0 \leq x \leq 1 \) and \( 0 \leq y \leq 1 \).

In order to find the probability \( P[X + Y \geq 1] \), we identify the two dimensional region representing \( X + Y \geq 1 \). This is generally found by drawing the boundary line for the inequality, which is \( x + y = 1 \) (or \( y = 1 - x \)) in this case, and then determining which side of the line is represented in the inequality. We can see that \( x + y \geq 1 \) is equivalent to \( y \geq 1 - x \). This is the shaded region in the graph below.

![Graph showing the region where \( x + y \geq 1 \).]

The probability \( P[X + Y \geq 1] \) is found by integrating the joint density over the two-dimensional region. It is possible to represent two-variable integrals in either order of integration. In some cases one order of integration is more convenient than the other. In this case there is not much advantage of one direction of integration over the other.

\[
P[X + Y \geq 1] = \int_0^1 \int_{1-x}^1 \frac{3}{2}(x^2 + y^2) \, dy \, dx = \int_0^1 \frac{1}{2} (3x^2y + y^3) \bigg|_{y=1-x}^{y=1} \, dx
\]
\[
= \int_0^1 \left(3x^2 + 1 - 3x^2(1-x) - (1-x)^3\right) \, dx = \frac{3}{4}.
\]
Reversing the order of integration, we have \( x \geq 1 - y \), so that \( P[X + Y \geq 1] = \int_0^1 \int_{1-y}^1 \frac{3}{2}(x^2 + y^2) \, dx \, dy = \frac{3}{4} \).

Expectation of a function of jointly distributed random variables
If \( h(x, y) \) is a function of two variables, and \( X \) and \( Y \) are jointly distributed random variables, then the expected value of \( h(X, Y) \) is defined to be
\[
E[h(X, Y)] = \sum_x \sum_y h(x, y) \cdot f(x, y) \quad \text{in the discrete case, and}
\]
\[
E[h(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x, y) \cdot f(x, y) \, dy \, dx \quad \text{in the continuous case.} \quad (4.3)
\]

Example LM4-3:
\( X \) and \( Y \) are discrete random variables which are jointly distributed with the following probability function \( f(x, y) \) (from Example LM4-1):

\[
\begin{array}{c|ccc}
X & -1 & 0 & 1 \\
\hline
1 & 1/8 & 1/9 & 1/6 \\
0 & 1/9 & 0 & 1/6 \\
-1 & 1/6 & 1/9 & 1/9 \\
\end{array}
\]

Find \( E[X \cdot Y] \).
Solution:

\[ E[XY] = \sum_x \sum_y x y \cdot f(x, y) = (-1)(1)(\frac{1}{18}) + (-1)(0)(\frac{1}{9}) + (-1)(-1)(\frac{1}{6}) \\
+ (0)(1)(\frac{1}{9}) + (0)(0)(0) + (0)(-1)(\frac{1}{9}) + (1)(1)(\frac{1}{6}) + (1)(0)(\frac{1}{6}) + (1)(-1)(\frac{1}{6}) = \frac{1}{6}. \]

Note that the summation is taken over all pairs \((x, y)\) in the joint distribution. □

Example LM4-4:

Suppose that \( f(x, y) = \frac{3}{2}(x^2 + y^2) \) is the density function for the joint distribution of the continuous random variables \(X\) and \(Y\) defined over the unit square defined on the region \( 0 \leq x \leq 1 \) and \( 0 \leq y \leq 1 \). Find \( E[X^2 + Y^2] \).

Solution:

\[
E[X^2 + Y^2] = \int_0^1 \int_0^1 (x^2 + y^2) \cdot f(x, y) \, dy \, dx = \int_0^1 \int_0^1 (x^2 + y^2)(\frac{3}{2})(x^2 + y^2) \, dy \, dx \\
= \int_0^1 (1.5x^4 + x^2 + .3) \, dx = \frac{14}{15}. \]

LM-4.2 Marginal distribution of \(X\) found from a joint distribution of \(X\) and \(Y\)

If \(X\) and \(Y\) have a joint distribution with joint density or probability function \(f(x, y)\), then the marginal distribution of \(X\) has a probability function or density function denoted \(f_X(x)\), which is equal to

\[ f_X(x) = \sum_y f(x, y) \quad \text{in the discrete case}, \]

\[ f_X(x) = \int_{-\infty}^{\infty} f(x, y) \, dy \quad \text{in the continuous case}. \] (4.4)

The density function for the marginal distribution of \(Y\) is found in a similar way, \(f_Y(y)\) is equal to either \(f_Y(y) = \sum_x f(x, y)\) or \(f_Y(y) = \int_{-\infty}^{\infty} f(x, y) \, dx\).

For instance, \(f_X(1) = \sum_y f(1, y)\) in the discrete case. What we are doing is "adding up" the probability for all points whose \(x\)-value is 1 to get the overall probability the \(X\) is 1.

Marginal probability functions and marginal density functions must satisfy all the requirements of probability and density functions. A marginal probability function must sum to 1 over all points of probability and a marginal density function must integrate to 1. The marginal distribution of \(X\) describes the behavior of \(X\) alone without reference to \(Y\) (and same for marginal of \(Y\)).

Example LM4-5:

Find the marginal distributions of \(X\) and \(Y\) for the joint distribution in Example LM4-1.

Solution:

The joint distribution was given as

\[
\begin{array}{ccc}
X & -1 & 0 & 1 \\
1 & \frac{1}{18} & \frac{1}{9} & \frac{1}{6} \\
Y & 0 & \frac{1}{9} & \frac{1}{6} \\
-1 & \frac{1}{6} & \frac{1}{9} & \frac{1}{9}
\end{array}
\]

To find the marginal probability function for \(X\), we first note that \(X\) can be \(-1, 0\), or \(1\).

We wish to find \(f_X(-1) = P[X = -1]\), \(f_X(0)\), and \(f_X(1)\).
As noted above, to find \( f_X(x) \) we sum over the other variable \( Y \):

\[
    f_X(-1) = \sum_{y} f(-1, y) = f(-1, -1) + f(-1, 0) + f(-1, 1) = \frac{1}{6} + \frac{1}{9} + \frac{1}{18} = \frac{1}{3}.
\]

In a similar way we get \( f_X(0) = \frac{1}{9} + 0 + \frac{2}{9} = \frac{2}{9} \) and \( f_X(1) = \frac{1}{6} + \frac{1}{9} + \frac{1}{6} = \frac{4}{9} \).

In Example LM4-1 we saw that \( P[X = 0] = \frac{2}{9} \).

What we were finding was the marginal probability \( f_X(0) \) that was just found here.

Note also that \( \sum_{x} f_X(x) = f_X(-1) + f_X(0) + f_X(1) = \frac{1}{3} + \frac{2}{9} + \frac{4}{9} = 1 \).

This verifies that \( f_X(x) \) satisfies the requirements of a probability function. The marginal probability function of \( Y \) is found in the same way, except that we sum over \( x \) (across each row in the table above).

\[
    f_Y(-1) = \frac{1}{6} + \frac{1}{9} + \frac{1}{9} = \frac{7}{18}, \quad f_Y(0) = \frac{1}{9} + 0 + \frac{1}{6} = \frac{5}{18} \quad \text{and} \quad f_Y(1) = \frac{1}{18} + \frac{1}{9} + \frac{1}{6} = \frac{1}{3}.
\]

**Example LM4-6:**

Find the marginal distributions of \( X \) and \( Y \) for the joint distribution in Example LM4-2.

**Solution:**

The joint density function is \( f(x, y) = \frac{3}{2}(x^2 + y^2) \) for \( 0 \leq x \leq 1 \) and \( 0 \leq y \leq 1 \).

The marginal density function of \( X \) is found by integrating out the other variable \( y \):

\[
    f_X(x) = \int_{y} f(x, y) \, dy = \int_{0}^{1} f(x, y) \, dy = \int_{0}^{1} \frac{3}{2}(x^2 + y^2) \, dy = \frac{3}{2}x^2 + \frac{1}{2} \quad \text{for} \quad 0 \leq x \leq 1.
\]

We can verify that this is a proper density function by checking that \( \int_{0}^{1} f_X(x) \, dx = 1 \).

In a similar way, \( f_Y(y) = \frac{3}{2}y^2 + \frac{1}{2} \) for \( 0 \leq y \leq 1 \).

**LM-4.3 Independence of random variables \( X \) and \( Y \)**

Random variables \( X \) and \( Y \) with density functions \( f_X(x) \) and \( f_Y(y) \) are said to be independent (or stochastically independent) if the probability space is rectangular (\( a \leq x \leq b, \ c \leq y \leq d \), where the endpoints can be infinite) and if the joint density function is of the form

\[
    f(x, y) = f_X(x) \cdot f_Y(y). \tag{4.5}
\]

If \( X \) and \( Y \) are independent, then for any functions \( g \) and \( h \),

\[
    E[g(X) \cdot h(Y)] = E[g(X)] \cdot E[h(Y)] , \quad \text{and in particular,} \quad E[X \cdot Y] = E[X] \cdot E[Y]. \tag{4.6}
\]

For the discrete joint distribution in Example LM4-1 we can see that \( X \) and \( Y \) are not independent, because, for instance, \( f( -1, -1) = \frac{1}{6} \neq \frac{1}{3} \cdot \frac{7}{18} = f_X(-1) \cdot f_Y(-1) \). For the continuous joint distribution of Example LM4-2, we see that

\[
    f(x, y) = \frac{3}{2}(x^2 + y^2) \neq \left( \frac{3}{2}x^2 + \frac{1}{2}\right) \left( \frac{3}{2}y^2 + \frac{1}{2}\right) = f_X(x) \cdot f_Y(y), \quad \text{so} \ X \text{ and } Y \text{ are not independent.}
\]

**Example LM4-7:**

Suppose that \( X \) and \( Y \) are independent continuous random variables with the following density functions:

\[
    f_X(x) = 1 \quad \text{for} \quad 0 < x < 1 \quad \text{and} \quad f_Y(y) = 2y \quad \text{for} \quad 0 < y < 1.
\]

Find \( P[Y < X] \).
Solution:
Since $X$ and $Y$ are independent, the density function of the joint distribution of $X$ and $Y$ is
$$f(x, y) = f_X(x) \cdot f_Y(y) = 2y,$$
and is defined on the rectangle created by the intervals for $X$ and $Y$, which, in this case, is the unit square. $P[Y < X] = \int_0^1 \int_0^x 2y \, dy \, dx = \frac{1}{3}$.

LM-4.4 Conditional distribution of $Y$ given $X = x$
Conditional distributions are very important in the loss models and credibility material and must be understood well.

The way in which a conditional distribution is defined follows the basic definition of conditional probability, $P[A|B] = \frac{P[A \cap B]}{P[B]}$. In fact, given discrete joint distribution, this is exactly how a conditional distribution is defined. Example LM4-1 described a discrete joint distribution of $X$ and $Y$, and then Example LM4-5 showed how to formulate the marginal distributions of $X$ and $Y$. We now wish to formulate a conditional distribution. For instance, for the joint distribution of Example LM4-1, suppose we wish to describe the conditional distribution of $X$ given $Y$. What we are trying to describe are conditional probabilities of the form $P[X = x|Y = 1]$.

We find these conditional probabilities in the usual way that conditional probability is defined.
$$P[X = -1|Y = 1] = \frac{P[(X=-1) \cap (Y=1)]}{P[Y=1]}.$$

The denominator is the marginal probability that $Y = 1$, $f_Y(1) = \frac{1}{3}$. The numerator is the joint probability $f(-1, 1) = \frac{1}{18}$, which is found in the joint probability table. Then,
$$P[X = -1|Y = 1] = \frac{f(-1, 1)}{f_Y(1)} = \frac{1/18}{1/3} = \frac{1}{6}.\text{ We would denote this conditional probability}$$
$$f_{X|Y}(-1|Y = 1).\text{ In a similar way, we can get } f_{X|Y}(0|Y = 1) = \frac{f(0, 1)}{f_Y(1)} = \frac{1/9}{1/3} = \frac{1}{3}, \text{ and}$$
$$f_{X|Y}(1|Y = 1) = \frac{f(1, 1)}{f_Y(1)} = \frac{1/6}{1/3} = \frac{1}{2}.\text{ This completely describes the conditional distribution of } X \text{ given } Y = 1.\text{ As with any discrete distribution, probabilities must add to 1, and this is the case for this conditional distribution, since}$$
$$f_{X|Y}(-1|Y = 1) + f_{X|Y}(0|Y = 1) + f_{X|Y}(1|Y = 1) = \frac{1}{6} + \frac{1}{3} + \frac{1}{2} = 1.$$
We also apply this same algebraic form to define the conditional density in the continuous case, with
\( f(x, y) \) being the joint density and \( f_X(x) \) being the marginal density. In the continuous case, the conditional mean of \( Y \) given \( X = x \) would be

\[
E[Y|X = x] = \int y \cdot f_{Y|X}(y|X = x) \, dy ,
\]

where the integral is taken over the appropriate interval for the conditional distribution of \( Y \) given \( X = x \). The conditional density/probability is also written as \( f_{Y|X}(y|x) \) or \( f(y|x) \).

If \( X \) and \( Y \) are independent random variables, then \( f_{Y|X}(y|X = x) = f_Y(y) \) and \( f_{X|Y}(x|Y = y) = f_X(x) \), which indicates that the density of \( Y \) does not depend on \( X \) and vice-versa. The conditional density function must satisfy the usual requirement of a density function, \( \int_{-\infty}^{\infty} f_{Y|X}(y|x) \, dy = 1 \). Note also that if the marginal density of \( X \) is known, \( f_X(x) \), and the conditional density of \( Y \) given \( X = x \) is also known, \( f_{Y|X}(y|X = x) \), then the joint density of \( X \) and \( Y \) can be formulated as

\[
f(x, y) = f_{Y|X}(y|X = x) \cdot f_X(x) .
\]

**Example LM4-8:**
Find the conditional distribution of \( Y \) given \( X = -1 \) for the joint distribution of Example LM4-1. Find the conditional expectation of \( Y \) given \( X = -1 \).

**Solution:**
The marginal probability function for \( X \) was found in Example LM4-5, where it was found that \( f_X(-1) = \frac{1}{3} \). The conditional probability function of \( Y \) given \( X = -1 \) is

\[
f_{Y|X}(y|X = -1) = \frac{f_Y(-1,y)}{f_X(-1)} = \frac{f_Y(-1,y)}{\frac{1}{3}} .
\]
Then,

\[
f_{Y|X}(-1|X = -1) = \frac{f_Y(-1,-1)}{\frac{1}{3}} = \frac{1}{6} ; \quad f_{Y|X}(0|X = -1) = \frac{f_Y(-1,0)}{\frac{1}{3}} = \frac{1}{9} = \frac{1}{3} ,
\]

and \( f_{Y|X}(1|X = -1) = \frac{f_Y(-1,1)}{\frac{1}{3}} = \frac{1}{18} = \frac{1}{6} .
\]

\[
\begin{align*}
E[Y|X = -1] &= \sum_{y} y \cdot f_{Y|X}(y|X = -1) = (-1)\left(\frac{1}{2}\right) + (0)\left(\frac{1}{3}\right) + (1)\left(\frac{1}{6}\right) = -\frac{1}{3} .
\end{align*}
\]

**Example LM4-9:**
Find the conditional density and conditional expectation and conditional variance of \( X \) given \( Y = .3 \) for the joint distribution of Example LM4-2.

**Solution:**
\[
f_{X|Y}(x|Y = .3) = \frac{f(x,.3)}{f_Y(.3)} = \frac{\frac{3}{2}(x^2+.3^2)}{\frac{2}{3}3^2+\frac{1}{2}} = \frac{\frac{3}{2}(x^2+.09)}{.635}.
\]
The conditional expectation is

\[
E[X|Y = .3] = \int_{0}^{1} x \cdot f_{X|Y}(x|Y = .3) \, dx = \int_{0}^{1} x \cdot \frac{\frac{3}{2}(x^2+.09)}{.635} \, dx = .697.
\]

The conditional second moment of \( X \) given \( Y = .3 \) is

\[
E[X^2|Y = .3] = \int_{0}^{1} x^2 \cdot f_{X|Y}(x|Y = .3) \, dx = \int_{0}^{1} x^2 \cdot \frac{\frac{3}{2}(x^2+.09)}{.635} \, dx = .543 .
\]

The conditional variance is

\[
Var[X|Y = .3] = E[X^2|Y = .3] - (E[X|Y = .3])^2 = .543 - (.697)^2 = .057 .
\]

We can construct the joint density \( f(x, y) \) from knowing the conditional density \( f_{Y|X}(y|x) \) and the marginal density \( f_X(x) \) using the relationship \( f(x, y) = f(y|x) \cdot f_X(x) \). When doing this, care must be taken to ensure that proper two-dimensional region is being formulated for the joint distribution.
Example LM4-10:
Suppose that $X$ has a continuous distribution with pdf $f_X(x) = 2x$ on the interval $(0, 1)$, and $f_X(x) = 0$ elsewhere. Suppose that $Y$ is a continuous random variable such that the conditional distribution of $Y$ given $X = x$ is uniform on the interval $(0, x)$. Find the mean and variance of (the marginal distribution of) $Y$.

**Solution:**
We find the unconditional (marginal) distribution of $Y$. We are given $f_X(x) = 2x$ for $0 < x < 1$, and $f_Y(x|X=x) = \frac{1}{x}$ for $0 < y < x$. Then, $f(x, y) = f(y|x) \cdot f_X(x) = \frac{1}{x} \cdot 2x = 2$ for $0 < y < x < 1$.

The unconditional (marginal) distribution of $Y$ has pdf.

\[
f_Y(y) = \int_{-\infty}^{\infty} f(x, y) \, dx = \int_{y}^{1} 2 \, dx = 2(1 - y) \quad \text{for} \quad 0 < y < 1 \quad \text{and} \quad f_Y(y) = 0 \quad \text{elsewhere}.
\]

Then

\[
E[Y] = \int_{0}^{1} y \cdot 2(1 - y) \, dy = \frac{1}{3}, \quad E[Y^2] = \int_{0}^{1} y^2 \cdot 2(1 - y) \, dy = \frac{1}{6}, \quad \text{and}
\]

\[
Var[Y] = E[Y^2] - (E[Y])^2 = \frac{1}{6} - \left( \frac{1}{3} \right)^2 = \frac{1}{18}.
\]

The idea applied in Example LM4-10 is quite important in the continuous Bayesian credibility approach that will be covered later in the study guide.

**LM-4.5 Covariance Between Random Variables X and Y**
If random variables $X$ and $Y$ are jointly distributed with joint density/probability function $f(x, y)$, the covariance between $X$ and $Y$ is

\[
Cov[X, Y] = E[(X - E[X])(Y - E[Y])] = E[(X - \mu_X)(Y - \mu_Y)] = E[XY] - E[X] \cdot E[Y]. \tag{4.10}
\]

Note that $Cov[X, X] = Var[X]$. If $X$ and $Y$ are independent, then

\[
E[X \cdot Y] = E[X] \cdot E[Y] \quad \text{and} \quad Cov[X, Y] = 0. \tag{4.11}
\]

For constants $a$, $b$, $c$, $d$, $e$, $f$ and random variables $X$, $Y$, $Z$ and $W$,

\[
Cov[aX + bY + cZ + dW + e + f] = adCov[X, Z] + aeCov[X, W] + bdCov[Y, Z] + beCov[Y, W] \tag{4.12}
\]

An important application of the covariance is in finding the variance of the sum of $X$ and $Y$. Suppose that $a$, $b$ and $c$ are constants. Then

\[
Var[aX + bY + c] = a^2Var[X] + b^2Var[Y] + 2abCov[X, Y]. \tag{4.13}
\]

If $X$ and $Y$ are independent, then $Var[X + Y] = Var[X] + Var[Y]$.

If $X_1, X_2, \ldots, X_n$ are independent, then $Var[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} Var[X_i]$. \tag{4.14}

If $X_1, X_2, \ldots, X_n$ are independent, and if $S = \sum_{i=1}^{n} X_i$, then the moment generating function of $S$ is

\[
M_S(t) = \prod_{i=1}^{n} M_{X_i}(t). \tag{4.15}
\]
LM-4.6 Coefficient of correlation between random variables $X$ and $Y$

The coefficient of correlation between random variables $X$ and $Y$ is defined to be

$$\rho(X, Y) = \frac{\text{Cov}[X,Y]}{\sigma_X \sigma_Y}, \quad (4.16)$$

where $\sigma_X$ and $\sigma_Y$ are the standard deviations of $X$ and $Y$ respectively. Note that $-1 \leq \rho_{X,Y} \leq 1$ always.

Example LM4-11:

Find $\text{Cov}[X,Y]$ for the jointly distributed discrete random variables in Example LM4-1 above.

Solution:

$\text{Cov}[X,Y] = E[XY] - E[X] \cdot E[Y]$. In Example LM4-3 it was found that $E[XY] = \frac{1}{6}$.

The marginal probability function for $X$ is $P[X = 1] = \frac{1}{6} + \frac{1}{6} + \frac{1}{9} = \frac{4}{9}$,

$P[X = 0] = \frac{2}{9}$ and $P[X = -1] = \frac{1}{3}$, and the mean of $X$ is

$E[X] = (1)(\frac{4}{9}) + (0)(\frac{2}{9}) + (-1)(\frac{1}{3}) = \frac{1}{9}$.

In a similar way, the probability function of $Y$ is found to be $P[Y = 1] = \frac{1}{3}$, $P[Y = 0] = \frac{5}{18}$, and $P[Y = -1] = \frac{7}{18}$, with a mean of $E[Y] = -\frac{1}{18}$. Then, $\text{Cov}[X,Y] = \frac{1}{6} - (\frac{1}{9})(-\frac{1}{18}) = \frac{14}{81}$. □
1. Let $X$ and $Y$ be discrete random variables with joint probability function

$$ f(x, y) = \begin{cases} \frac{2^{x+y}}{x+y} & \text{for } x, y = 1, 2 \\ 0, & \text{otherwise} \end{cases} $$

Calculate $E\left[ \frac{X}{Y} \right]$. 

A) $\frac{8}{5}$  
B) $\frac{5}{4}$  
C) $\frac{4}{3}$  
D) $\frac{25}{18}$  
E) $\frac{5}{3}$

2. Let $X$ and $Y$ be continuous random variables with joint cumulative distribution function

$$ F(x, y) = \frac{1}{250}(20xy - x^2y - xy^2) \quad \text{for } 0 \leq x \leq 5 \quad \text{and} \quad 0 \leq y \leq 5. $$

Determine $P[X > 2]$. 

A) $\frac{3}{125}$  
B) $\frac{11}{50}$  
C) $\frac{12}{25}$  
D) $\frac{1}{250}(39y - 3y^2)$  
E) $\frac{1}{250}(36y - 2y^2)$

3. Let $X$ and $Y$ be discrete random variables with joint probabilities given by

<table>
<thead>
<tr>
<th>$X$</th>
<th>$Y$</th>
<th>$P(X, Y)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1</td>
<td>$\frac{1}{2}$</td>
</tr>
<tr>
<td>4</td>
<td>$\theta_1 + \theta_2$</td>
<td>$\theta_1 + 2\theta_2$</td>
</tr>
<tr>
<td></td>
<td>$\theta_1 + \theta_2$</td>
<td>$\theta_1 + \theta_2$</td>
</tr>
</tbody>
</table>

Let the parameters $\theta_1$ and $\theta_2$ satisfy the usual assumption associated with a joint probability distribution and the additional constraints $-.25 \leq \theta_1 \leq .25$ and $0 \leq \theta_2 \leq .35$. If $X$ and $Y$ are independent, then $(\theta_1, \theta_2) =$

A) $(0, \frac{1}{6})$  
B) $(\frac{1}{4}, 0)$  
C) $(-\frac{1}{4}, \frac{1}{3})$  
D) $(-\frac{1}{8}, \frac{1}{4})$  
E) $(\frac{1}{16}, \frac{1}{8})$

4. Let $X$ and $Y$ be continuous random variables with joint density function $f(x, y) = \begin{cases} \frac{1}{1-x} & \text{for } 0 < x < y < 1 \\ 0, & \text{otherwise} \end{cases}$. 

Determine the density function of the conditional distribution of $Y$ given $X = x$, where $0 < x < 1$. 

A) $\frac{1}{1-x}$  
B) $2(1-x)$  
C) $2$  
D) $\frac{1}{y}$  
E) $\frac{1}{1-y}$

5. A wheel is spun with the numbers 1, 2 and 3 appearing with equal probability of $\frac{1}{3}$ each. If the number 1 appears, the player gets a score of 1.0; if the number 2 appears, the player gets a score of 2.0; if the number 3 appears, the player gets a score of $X$, where $X$ is a normal random variable with mean 3 and standard deviation 1. If $W$ represents the player's score on 1 spin of the wheel, then what is $P[W \leq 1.5]$? 

A) .13  
B) .33  
C) .36  
D) .40  
E) .64

6. Let $X$ and $Y$ be continuous random variables with joint density function

$$ f(x, y) = \begin{cases} \frac{x+y}{x+y} & \text{for } 0 < x < 1, \ 0 < y < 1 \\ 0, & \text{otherwise} \end{cases} $$

What is the marginal density function for $X$, where nonzero? 

A) $y + \frac{1}{2}$  
B) $2x$  
C) $x$  
D) $\frac{x+y^2}{2}$  
E) $x + \frac{1}{2}$
7. Let $X$ and $Y$ be discrete random variables with joint probability function
\[ f(x, y) = \begin{cases} \frac{(x+1)(y+2)}{54} & \text{for } x=0,1,2; \ y=0,1,2, \\ 0, & \text{otherwise} \end{cases} \]
What is $E[Y|X = 1]$?

A) $\frac{11}{27}$  B) 1  C) $\frac{11}{9}$  D) $\frac{y+2}{9}$  E) $\frac{y^2+2y}{9}$

8. Let $X$ and $Y$ be continuous random variables with joint density function
\[ f(x, y) = \begin{cases} 6x & \text{for } 0<x<y<1, \\ 0, & \text{otherwise} \end{cases} \]
Note that $E[X] = \frac{1}{2}$ and $E[Y] = \frac{3}{4}$. What is $Cov[X, Y]$?

A) $\frac{1}{40}$  B) $\frac{1}{20}$  C) $\frac{1}{10}$  D) $\frac{1}{5}$  E) 1

9. The distribution of Smith's future lifetime is $X$, an exponential random variable with mean $\alpha$, and the distribution of Brown's future lifetime is $Y$, an exponential random variable with mean $\beta$. Smith and Brown have future lifetimes that are independent of one another. Find the probability that Smith outlives Brown.

A) $\frac{\alpha}{\alpha+\beta}$  B) $\frac{\beta}{\alpha+\beta}$  C) $\frac{\alpha-\beta}{\alpha}$  D) $\frac{\beta-\alpha}{\beta}$  E) $\frac{\alpha}{\beta}$

10. In reviewing some data on smoking ($X$, number of packages of cigarettes smoked per year), income ($Y$, in thousands per year) and health ($Z$, number of visits to the family physician per year) for a sample of males, it is found that
\[ E[X] = 10, \ Var[X] = 25, \ E[Y] = 50, \ Var[Y] = 100, \ E[Z] = 6, \ Var[Z] = 4, \] and $Cov(X, Y) = -10$, $Cov(X, Z) = 2.5$ (covariances).

Dr. N.A. Ively, a young statistician, attempts to describe the variable $Z$ in terms of $X$ and $Y$ by the relation $Z = X + cY$, where $c$ is a constant to be determined. Dr. Ively's methodology for determining $c$ is to find the value of $c$ for which $Cov(X, Z)$ remains equal to 2.5 when $Z$ is replaced by $X + cY$. What value of $c$ does Dr. Ively find?

A) 2.00  B) 2.25  C) 2.50  D) $-2.00$  E) $-2.25$

11. In order to simplify an actuarial analysis Actuary A uses an aggregate distribution $S = X_1 + \cdots + X_N$, where $N$ has a Poisson distribution with mean 10 and $X_i = 1.5$ for all $i$.

Actuary A’s work is criticized because the actual severity distribution is given by
\[ Pr(Y_i = 1) = (Y_i = 2) = 0.5, \] for all $i$, where the $Y_i$’s are independent.

Actuary A counters this criticism by claiming that the correlation coefficient between $S$ and $S^* = Y_1 + \cdots + Y_N$ is high.

Calculate the correlation coefficient between $S$ and $S^*$.

A) 0.75  B) 0.80  C) 0.85  D) 0.90  E) 0.95
1. \( E\left[\frac{X}{Y}\right] = \sum_{x=1}^{2} \sum_{y=1}^{2} \frac{X_i}{Y_i} \cdot p(X_i, Y_i) = 1 \cdot \frac{2}{9} + \frac{1}{2} \cdot \frac{1}{5} + 2 \cdot \frac{4}{9} + 1 \cdot \frac{2}{9} = \frac{25}{18} \)
   Answer: D

2. \( F_X(2) = P[X \leq 2] = \lim_{y \to \infty} F(2, y) = F(2, 5) = \frac{130}{250} = \frac{13}{25} \), so that
   \[ P[X > 2] = 1 - P[X \leq 2] = \frac{12}{25}. \]
   Answer: C

3. Since the total probability must be 1, we have \( 4\theta_1 + 6\theta_2 = 1 \). The marginal distributions of \( X \) and \( Y \) have
   \[ P[X = 1] = P[X = 5] = P[Y = 2] = P[Y = 4] = 2\theta_1 + 3\theta_2 = \frac{1}{2}. \]
   Then, because of independence, \( P[X = 1, Y = 2] = P[X = 1] \cdot P[Y = 2] = \frac{1}{4} = \theta_1 + \theta_2 \).
   Solving the two equations in \( \theta_1 \) and \( \theta_2 \) \((4\theta_1 + 6\theta_2 = 1 \) and \( \theta_1 + \theta_2 = \frac{1}{4} \)) results in \( \theta_1 = \frac{1}{4}, \theta_2 = 0 \).
   Answer: B

4. The region of joint density is the triangular region above the line \( y = x \) and below the horizontal line \( y = 1 \) for \( 0 < x < 1 \). The conditional density of \( y \) given \( X = x \) is \( f(y \mid X = x) = \frac{f(x, y)}{f_X(x)} \),
   where \( f_X(x) \) is the marginal density function of \( x \).
   \[ f_X(x) = \int_{-\infty}^{\infty} f(x, y) \, dy = \int_{x}^{1} 2 \, dy = 2(1 - x), \] so that \( f(y \mid X = x) = \frac{2}{2(1-x)} = \frac{1}{1-x} \)
   and the region of density for the conditional distribution of \( Y \) given \( X = x \) is \( x < y < 1 \). It is true
   in general that if a joint distribution is uniform (has constant density in a region) then any
   conditional (though not necessarily marginal) distribution will be uniform on it restricted region of
   probability - the conditional distribution of \( Y \) given \( X = x \) is uniform on the interval \( x < y < 1 \),
   with constant density \( \frac{1}{1-x} \).
   Answer: A

5. Let \( N \) denote the number that appears on the wheel, so that
   \[ P[N = 1] = P[N = 2] = P[N = 3] = \frac{1}{3}. \]
   Then, conditioning over \( N \),
   \[ P[W \leq 1.5] = P[W \leq 1.5, N = 1] \cdot P[N = 1] + P[W \leq 1.5, N = 2] \cdot P[N = 2] + P[W \leq 1.5, N = 3] \cdot P[N = 3]. \]
   If \( N = 1 \) then \( W = 1 \), so that \( P[W \leq 1.5 \mid N = 1] = 1 \), and
   if \( N = 2 \) then \( W = 2 \), so that \( P[W \leq 1.5 \mid N = 2] = 0 \).
   If \( N = 3 \) then \( W \sim N(3, 1) \) so that
   \[ P[W \leq 1.5 \mid N = 3] = P[W - 3 \leq \frac{1.5 - 3}{1} | N = 3] = P[Z \leq -1.5] = 0.07 \]
   \((Z \) has a standard normal distribution - the probability is found from the table).
   Then, \( P[W \leq 1.5] = 1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + (.07) \cdot \frac{1}{3} = .357 \).
   Answer: C

6. \( f_X(x) = f_0^1 (x + y) \, dy = x + \frac{1}{2} \).
   Answer: E
7. \( f_X(1) = P[X = 1] = \sum_{y=-\infty}^{\infty} f(1, y) = f(1, 0) + f(1, 1) + f(1, 2) = \frac{1}{3} \).

Then we have conditional probabilities
\[ P[Y = 0|X = 1] = \frac{f(1,0)}{P[X=1]} = \frac{4/54}{1/3} = \frac{2}{9} \text{, and similarly,} \]
\[ P[Y = 1|X = 1] = \frac{1}{3} \text{ and } P[Y = 2|X = 1] = \frac{4}{9} \text{.} \]

Then, \( E[Y|X = 1] = 0 \cdot \frac{2}{9} + 1 \cdot \frac{1}{3} + 2 \cdot \frac{4}{9} = \frac{11}{9} \). Answer: C

8. \( \text{Cov}[X, Y] = E[XY] - E[X] \cdot E[Y] \)

The region of probability is the triangle above the line \( y = x \) in the unit square \( 0 \leq x \leq 1, 0 \leq y \leq 1 \).

\[ E[XY] = \int_0^1 \int_0^y xy \cdot 6x \, dx \, dy = \frac{2}{5} \]

\[ \rightarrow \text{Cov}[X, Y] = \frac{2}{5} - \frac{1}{2} \cdot \frac{3}{4} = \frac{1}{40} \text{.} \]

Alternatively,
\[ E[XY] = \int_0^1 \int_x^1 xy \cdot 6x \, dy \, dx = \frac{2}{5} \text{.} \]

Answer: A

9. \( P[Y < X] = \int_0^\infty \int_y^\infty f_Y(y) f_X(x) \, dx \, dy \) (since \( X \) and \( Y \) are independent, the joint density function of \( X \) and \( Y \) is the product of the two separate density functions).

The density function of \( X \) is \( \frac{1}{\alpha} e^{-x/\alpha} \), and of \( Y \) is \( \frac{1}{\beta} e^{-y/\beta} \), so that
\[ P[Y < X] = \int_0^\infty \int_y^\infty \frac{1}{\alpha} e^{-x/\alpha} \frac{1}{\beta} e^{-y/\beta} \, dx \, dy = \int_0^\infty \frac{1}{\beta} e^{-y/\beta} e^{-y/\alpha} \, dy = \frac{\frac{1}{\beta}}{\frac{1}{\alpha} + \frac{1}{\beta}} = \frac{\alpha}{\alpha + \beta} \text{.} \] Answer: A

10. \( \text{Cov}(X, X + cY) = \text{Cov}(X, X) + c\text{Cov}(X, Y) = \text{Var}[X] + c\text{Cov}(X, Y) = 25 - 10c \text{.} \)

This is set equal to \( \text{Cov}(X, Z) = 2.5 \text{, so that } 25 - 10c = 2.5 \rightarrow c = 2.25 \text{.} \) Answer: B

11. The covariance between \( S \) and \( S^* \) is \( E[SS^*] - E[S] \cdot E[S^*] \).

\[ E[S] = E[N] \cdot E[X] = (10)(1.5) = 15 \text{,} \]
\[ E[S^*] = E[N] \cdot E[Y] = (10)(1.5) = 15 \text{.} \]

We use the double expectation rule
\[ E[S \cdot S^*] = E[E[S \cdot S^*|N]] \text{.} \]
\[ E[S \cdot S^*|N] = E[(X_1 + \cdots + X_N)(Y_1 + \cdots + Y_N)|N] \]
\[ = E[1.5N(Y_1 + \cdots + Y_N)|N] = (1.5N)(N \cdot E[Y]) = 1.5N^2 \cdot (1.5) = 2.25N^2 \text{.} \]

Then
\[ E[S \cdot S^*] = E[2.25N^2] = 2.25E[N^2] = 2.25(\text{Var}[N] + (E[N])^2) = 2.25(10 + 100) = 247.5 \text{.} \]
\[ \text{Cov}[S, S^*] = E[S \cdot S^*] - E[S] \cdot E[S^*] = 247.5 - (15)(15) = 22.5 \text{.} \]
\[ \text{Var}[S] = \text{Var}[1.5N] = 2.25\text{Var}[N] = 22.5 \, \text{,} \]
\[ \text{Var}[S^*] = \text{Var}[N](E[Y])^2 + E[N]\text{Var}[Y] = 10E[Y^2] \text{ (since } N \text{ is Poisson)} \]
\[ = (10)((1^2)(.5) + (2^2)(.5)) = 25 \text{.} \]

Then the correlation coefficient between \( S \) and \( S^* \) is \( \frac{\text{Cov}[S, S^*]}{\sqrt{\text{Var}[S] \cdot \text{Var}[S^*]}} = \frac{22.5}{\sqrt{(22.5)(25)}} = .949 \text{.} \)

Answer: E
MODELING SECTION 5
PARAMETRIC DISTRIBUTIONS AND TRANSFORMATIONS

The material in this section relates to Sections 4.2.1-4.2.2, 5.2.1-5.2.3, 5.3 and 5.4 of "Loss Models". The suggested time frame for this section is 2 hours. There has been very infrequent reference to this topic on the exam. Later topics do not depend on this material, and it can be postponed and covered at a later time.

LM-5.1 Parametric Distributions

For any random variable, the mean and variance (skewness, etc.) are "parameters" of the distribution (some of these might be infinite) that can be calculated if the form of the distribution is known, or they can be estimated when a sample of data is available. There is also the notion of a parametric distribution, which means that the random variable $X$ has a pdf (or cdf) which is formulated in terms of parameters. In this definition, "parametric distribution" refers to a collection of distributions based on the set of all possible values of the parameters. Some examples are as follows.

(i) The uniform distribution on the interval $[0, \theta]$ has pdf

$$f(x) = \begin{cases} 1/\theta & \text{if } 0 \leq x \leq \theta \\ 0 & \text{otherwise} \end{cases}.$$  

This is a parametric distribution with parameter $\theta$ (the mean is $\theta/2$, and other distribution quantities such as variance, skewness, etc. are formulated in terms of the parameter $\theta$).

(ii) The normal distribution has parameters $\mu$ (mean) and $\sigma$ (standard deviation) has pdf

$$f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \quad \text{for } -\infty < x < \infty.$$  

(iii) The Poisson distribution with parameter $\lambda > 0$ is a discrete, non-negative integer-valued random variable with pf

$$f(x) = P[X = x] = \frac{e^{-\lambda} \lambda^x}{x!} \quad \text{for } x = 0, 1, 2, \ldots$$  

(iv) The exponential distribution with parameter $\theta > 0$ has pdf

$$f(x) = \frac{1}{\theta} e^{-x/\theta} \quad \text{for } x > 0$$  

and cdf

$$F(x) = 1 - e^{-x/\theta}.$$  

There are a large number of parametric distributions in the Tables for Exam C (taken from an Appendix of the "Loss Models" book). Information on the pdf, cdf, moments, etc. is given there.

Scale distribution and scale parameter

A continuous parametric distribution is a scale distribution if $cX$ is a member of the set of distributions whenever $X$ is a member and $c > 0$ is a constant. $\theta$ is a scale parameter if the corresponding parameter for $cX$ is $c\theta$. It is possible for a distribution to be a scale distribution and yet not have a scale parameter (the lognormal distribution is an example).

Example LM5-1:

The exponential distribution has a pdf of the form

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

or equivalently, cdf of the form

$$F(x; \theta) = 1 - e^{-x/\theta}, \quad \theta > 0.$$  

Show that this is a scale distribution and that $\theta$ is the scale parameter.

Solution:

If $Y = cX$ and $c > 0$, then $P[Y \leq y] = P[X \leq y/c] = 1 - e^{-(y/c)/\theta} = 1 - e^{-y/c\theta}$. Therefore the exponential family is a scale family and the exponential parameter is a scale parameter.
The Exam C table of distributions has a number of distributions which involve the parameter $\theta$. The table has been arranged so that $\theta$ always is a scale parameter. Any distribution that has the parameter $\theta$ in the table is a scale distribution. The only continuous distributions in the table that do not have a parameter $\theta$ are the lognormal and the log-$t$ distributions.

The Pareto distribution is another example of a scale distribution. Suppose that $X$ has a Pareto distribution with parameters $\theta$ and $\alpha$, and suppose that $c > 0$ is a constant. Let $Y = cX$. Then $Y$ also has a Pareto distribution with parameters $c\theta$ and $\alpha$ ($\theta$ is the scale parameter, $\alpha$ is unchanged).

**LM-5.2 Families of Distributions**

Section 5.3 of "Loss Models" outlines how families of probability distributions can be organized. In addition, it is shown that there is a very general 4-parameter distribution, the Transformed Beta that has most of the other distributions in the Exam C Table as limits or special cases. There has not been any reference to this on released exams, so it may not be necessary to devote a lot of time to it.

A couple of simple observations that can be made are that the exponential distribution is a special case of the gamma distribution and it is also a special case of the Weibull distribution.

Suppose $X$ has a gamma distribution with parameters $\theta$ and $\alpha$. The pdf of $X$ is $f(x) = \frac{x^{\alpha-1}e^{-x/\theta}}{\theta^\alpha \Gamma(\alpha)}$.

If $\alpha = 1$, the pdf becomes $f(x) = \frac{e^{-x/\theta}}{\theta}$, which is the pdf of the exponential distribution with parameter $\theta$.

Suppose $X$ has a Weibull distribution with parameters $\theta$ and $\tau$. The pdf of $X$ is $f(x) = \frac{\tau(x/\theta)^{\tau-1}e^{-(x/\theta)^\tau}}{\theta^\tau \Gamma(\tau)}$.

If $\tau = 1$, the pdf becomes $f(x) = \frac{e^{-x/\theta}}{\theta}$, which is the pdf of the exponential distribution with parameter $\theta$.

There are some other examples in Section 5.3 of "Loss Models" in which limits of pdf's are taken as a parameter goes to 0 or $\infty$, and the limiting distribution is of a recognizable form.

**LM-5.2.1 The Linear Exponential Family**

A random variable with probability function or a density function that can be written in the form $f(x; \theta) = \frac{p(x) e^{r(x)\theta}}{q(\theta)}$, where $p(x)$ does not depend on $\theta$ is said to be from the exponential family. Also, the region on which the random variable is defined cannot depend on $\theta$. Some examples of random variables that are exponential family members are:

(i) Poisson with parameter $\lambda$:
   
   $f(x; \lambda) = \frac{e^{-\lambda} \lambda^x}{x!} = \frac{e^{-\lambda} e^{(\ln \lambda)x}}{x!}, \quad p(x) = \frac{1}{x!}, \quad r(\lambda) = \ln \lambda$ and $q(\lambda) = e^\lambda$.

(ii) Binomial with $m$ trials and parameter $p$:
   
   $f(x; p) = \binom{m}{x} p^x (1-p)^{m-x} = \binom{m}{x} \left( \frac{p}{1-p} \right)^x (1-p)^{m} = \binom{m}{x} e^{\ln \left( \frac{p}{1-p} \right) x}, \quad p(x) = \binom{m}{x}, \quad r(p) = \ln \left( \frac{p}{1-p} \right), \quad q(p) = (1-p)^{-m}$

(iii) Exponential with parameter $\theta$:
   
   $f(x; \theta) = \frac{e^{-x/\theta}}{\theta}, \quad p(x) = 1, \quad r(\theta) = -\frac{1}{\theta}, \quad q(\theta) = \theta$
The textbook shows that the normal distribution is a member of the linear exponential family if the mean is the parameter, and the gamma distribution is also a member of the linear exponential family.

We have the following general rules for a member of the linear exponential family:

\[ E(X) = \mu(\theta) = \frac{q'(\theta)}{r'(\theta)q(\theta)} \quad \text{and} \quad \text{Var}(X) = \frac{\mu'(\theta)}{r'(\theta)}. \tag{5.1} \]

**Example LM5-2:**

Use Equation 5.1 to find the mean and variance of a binomial random variable.

**Solution:**

For a binomial with \( m \) trials and probability \( p \) of success, the parameter for the linear exponential family representation is \( \theta = p \).

\[ f(x; p) = \frac{\binom{m}{x} p^x (1-p)^{m-x}}{\binom{m}{x}}, \]

where

\[ p(x) = \binom{m}{x}, \quad r(p) = \ln\left(\frac{p}{1-p}\right), \quad q(p) = (1-p)^{-m}. \]

and

\[ q'(p) = m(1-p)^{-m-1} \]

so that

\[ \frac{q'(\theta)}{r(\theta)q(\theta)} = \frac{m(1-p)^{-m-1}}{\left(\frac{1}{p} + \frac{1}{1-p}\right)(1-p)^{-m}} = mp = \mu(p) \quad \text{(mean of the binomial)}. \]

Then \( \mu'(p) = m \)

so that

\[ \frac{\mu'(p)}{r'(p)} = \frac{m}{\frac{1}{p} + \frac{1}{1-p}} = mp(1-p) \quad \text{(variance of the binomial)}. \]

---

**LM-5.3 A Comment on the Pareto (two-parameter) and the Single Parameter Pareto**

The Exam C table gives information on the Pareto distribution, which has parameters \( \theta \) and \( \alpha \), and pdf \( f(x) = \frac{\alpha \theta^\alpha}{x^{\alpha+1}} \) for \( x > 0 \). The Single Parameter Pareto distribution is also defined in the table, with pdf \( g(y) = \frac{\alpha y^{\alpha}}{y^{\alpha+1}} \) for \( y > \theta \).

These two distributions are essentially the same. They are related through the relation \( Y = X + \theta \). The Single Parameter Pareto is just the Pareto shifted to the right by a distance of \( \theta \); if \( y = x + \theta \), then \( Y > \theta \) is equivalent to \( x > 0 \).

Note that the mean of the Pareto is \( E[X] = \frac{\theta}{\alpha-1} \), and the mean of the Single Parameter Pareto is \( E[Y] = \frac{\alpha \theta}{\alpha-1} \), which is equal to \( E[X] + \theta \) (as should be the case if \( Y = X + \theta \)).
If a Pareto distribution is mentioned in the Exam C context, it refers to the two-parameter version; that is the default meaning of Pareto. The Single Parameter Pareto distribution will be explicitly named as such if that is the version being intended.

Both versions of the Pareto distribution arise in the estimation topics to be covered later in the Exam C material. In the case of the Single Parameter Pareto, the value of $\theta$ would be given, so only the one parameter $\alpha$ is estimated. In the case of the default (two parameter) Pareto distribution, both $\theta$ and $\alpha$ would be estimated.

LM-5.4 The Distribution of a Transformed Continuous Random Variable

The Loss Models book presents a few ways of constructing new continuous distributions from existing ones. If $X$ is a continuous random variable with cdf $F_X(x)$ and pdf $f_X(x)$ and if $Y$ is constructed as a function or transformation of $X$, then for some transformations it may be possible to formulate $F_Y(y)$ and $f_Y(y)$ in terms of $F_X(x)$ and $f_X(x)$.

There is a general rule that can be applied if $Y = g(X)$ and $g$ is an invertible function. An invertible function is one for which we can reformulate the original function to write $X$ in terms of $Y$, say $X = k(Y)$.

An example of an invertible function is $y = g(x) = 2x + 3$, which can be reformulated as $x = \frac{y-3}{2}$. Another example is $y = g(x) = x^2$ for $x > 0$, which has inverse $x = y^{1/2}$ (positive square root).

In general, $g$ will be an invertible function if it is strictly increasing or strictly decreasing on the region for which it is being used. That is why $g(x) = x^2$ for $x > 0$ was invertible; it is strictly increasing for $x > 0$ (but not for all real $x$).

Suppose that $k(y)$ is the inverse function of $g(x)$, so that $x = k(y) = k(g(x)) = x$ and $y = g(x) = g(k(y))$. For instance, if $g(x) = e^x$, then $k(y) = \ln(y)$ is the inverse of $g$, since $k(g(x)) = \ln(e^x) = x$ and $g(k(y)) = e^{\ln y} = y$.

Under these assumptions, we have the following relationship: if the pdf of $X$ is $f_X(x)$, and if $Y = g(X)$, then the pdf of $Y$ is

$$f_Y(y) = f_X(k(y)) \cdot |k'(y)|$$

(5.2)

(the absolute value ensures that the pdf of $Y$ is non-negative).

Also $F_Y(y) = F_X(k(y))$ if $g$ is increasing, and $F_Y(y) = 1 - F_X(k(y))$ if $g$ is decreasing.

It is also important to determine the region of probability for the transformed variable $Y$.

If the region of probability for $X$ is the interval $(a, b)$, then the region of probability for $Y$ will be the interval $(g(a), g(b))$ if $g$ is an increasing function, and $(g(b), g(a))$ if $g$ is a decreasing function.

**Example LM5-3:**

$X$ has the pdf $f_X(x) = 2x$ for $0 < x < 1$.

$Y$ is defined by the transformation $Y = -\ln(X)$. Find the pdf of $Y$.  

Actex Learning

SOA Exam C - Construction and Evaluation of Actuarial Models
Solution:
The inverse of the function \( g(x) = -\ln(x) \) is \( k(y) = e^{-y} \).

\[
f_Y(y) = f_X(k(y)) \cdot |k'(y)| = 2e^{-y} \cdot |e^{-y}| = 2e^{-2y}
\]
defined on the region \( 0 < y < \infty \).

\( X \) is defined on the interval \((0, 1)\) and \( g(x) = -\ln(x) \) is a decreasing function, so \( Y \) is defined on the interval \((g(1), g(0))\).

We see that \( g(1) = -\ln(1) = 0 \), and to find \( g(0) = -\ln(0) \) we take the limit of \( g(t) = -\ln(t) \) as \( t \to 0 \). This limit is \( +\infty \).

The region of probability for \( Y \) is \( 0 < y < \infty \).

We now summarize some of the more typical transformations that can arise. The pdf and cdf of \( X \) are denoted \( f_X(x) \) and \( F_X(x) \), with similar notation for the transformed variable \( Y \).

**Constant multiple transformation**

If \( Y = cX = g(X) \), where \( c > 0 \) is a constant, then \( X = \frac{Y}{c} = k(Y) \) is the inverse transformation.

Then

\[
f_Y(y) = f_X(k(y)) \cdot |k'(y)| = f_X\left(\frac{y}{c}\right) \cdot \frac{1}{c}.
\]

(5.3)

A scale family can be created using the constant multiple transformation. If \( X \) is a continuous random variable with \( X > 0 \), then the family of random variables \( \{\theta X : \theta > 0\} \) is a parametric family with scale parameter \( \theta \).

**Power Transformation**

The general form of the power transformation is

\[
Y = g(X) = X^{1/\tau}.
\]

(5.4)

Then \( x = k(y) = y^\tau \) is the inverse function of \( g \).

If \( \tau > 0 \), then

\[
f_Y(y) = \tau y^{\tau-1} f_X(y^\tau) \quad \text{and} \quad F_Y(y) = F_X(y^\tau).
\]

(5.5)

\( Y \) is called a "transformed distribution" of \( X \).

If \( \tau < 0 \), then

\[
f_Y(y) = -\tau y^{\tau-1} f_X(y^\tau) = \tau^* y^{-\tau-1} f_X(y^{-\tau}) \quad \text{and} \quad F_Y(y) = 1 - F_X(y^\tau),
\]

where

\[
\tau^* = -\tau.
\]

(5.6)

If \( \tau = -1 \) then \( Y \) is called an inverse distribution of \( X \), and if \( \tau < 0 \) but is not \( -1 \) then \( Y \) is called an inverse transformed distribution of \( X \).

For example, using the exponential distribution with mean \( \theta = 2 \) as the base distribution \( X \) for this transformation, with \( \tau > 0 \), we have

\[
f_Y(y) = f_X(k(y)) \cdot |k'(y)| = \frac{1}{2} e^{-k(y)/2} \cdot \tau y^{\tau-1} = \frac{1}{2} e^{-y/2} \cdot \tau y^{\tau-1} = \frac{\tau (y/2^{1/\tau})^{\tau} e^{-(y/2^{1/\tau})^\tau}}{y}.
\]

This is the pdf of a Weibull random variable with parameters \( \tau \) and \( \theta = 2^{1/\tau} \).

Note that the "\( \theta \)" in the Weibull distribution is not the same numerical value as the \( \theta = 2 \) in the original exponential distribution.
As another example, suppose that we use the Pareto with parameters $\alpha$ and $\theta$ as the base distribution. If $Y = X^{-1} = g(X)$ (so $\tau = -1$ in the power transformation), we have $X = k(Y) = \frac{1}{Y}$. This time we will use the relationship $F_Y(y) = F_X(k(y))$ to determine the distribution of $Y$. Since $g(x) = \frac{1}{x}$ is a decreasing function, we get

$$F_Y(y) = 1 - F_X\left(\frac{1}{y}\right) = \left(\frac{\theta}{y + \theta}\right)^\alpha = \left(\frac{y}{y + \frac{1}{\theta}}\right)^\alpha.$$  

This is the cdf of the inverse Pareto with parameters $\alpha$ and $\frac{1}{\theta}$.

Note that the "$\theta$" parameter from the original Pareto distribution has been inverted in the transformation to the inverse Pareto, but the parameter $\alpha$ has been maintained.

Using the transformation $Y = X^{-1}$ is how we get inverse transformations in general. We must be careful to identify the effect on the original distribution parameters and how they relate to the parameters in the transformed distribution. Each of the following distributions and their inverses involve a parameter labeled $\theta$ (in the Exam C table) and another parameter (except for the exponential distribution): exponential, Pareto, loglogistic, paralogistic, gamma, and Weibull.

For each of these, when the transformation $Y = X^{-1}$ is applied, the distribution of $Y$ is the inverse of the original distribution, and the "$\theta$" in the inverse distribution is numerically equal to $\frac{1}{\theta}$, where $\theta$ is the value from the original distribution, and the other parameters are unchanged.

For instance, if $X$ has a gamma distribution with parameters $\alpha = 3$ and $\theta = 5$, then $Y = X^{-1}$ has an inverse gamma distribution with $\alpha = 3$ and $\theta = \frac{1}{5}$. A couple of additional points to note are:

- in the case of the Pareto which has parameters $\alpha$ and $\theta$, and the inverse Pareto which has $\tau$ and "new" $\theta$, and "new" $\theta$ is $\frac{1}{\text{old} \theta}$, $\tau = \alpha$ when constructing the inverse distribution, and

- if $X$ has a loglogistic distribution with parameters $\gamma$ and $\theta$, then $Y = X^{-1}$ has a loglogistic distribution with parameters $\gamma$ and $\frac{1}{\theta}$ (the inverse of loglogistic is also loglogistic).

**Example LM5-4:**

For the random variable $X$ with pdf $f(x) = 2x$ for $0 < x < 1$ (and 0 elsewhere), find the pdf of (i) $cX$ ($c > 0$), (ii) $X^{-1}$, and (iii) $X^{1/\tau}$ ($\tau > 0$).

**Solution:**

(i) $f_Y(y) = \frac{1}{c} \cdot f_X\left(\frac{y}{c}\right) = \frac{1}{c} \cdot \frac{2y}{c^2} = \frac{2y}{c^2}$ for $0 < \frac{y}{c} < 1$,

or equivalently, $f_Y(y) = \frac{2y}{c^2}$ for $0 < y < c$.

(ii) $f_Y(y) = y^{-2} \cdot f_X(y^{-1}) = \frac{1}{y^2} \cdot \frac{2}{y} = \frac{2}{y^3}$ for $0 < \frac{1}{y} < 1$,

or equivalently, for $1 < y < \infty$. Note that $Y$ has a single parameter Pareto distribution with $\alpha = 2$ and $\theta = 1$.

(iii) $f_Y(y) = \tau y^{\tau - 1} \cdot f_X(y^\tau) = \tau y^{\tau - 1} \cdot 2y^\tau = 2\tau y^{2\tau - 1}$ for $0 < y < 1$. □
Example LM5-5:
Find expressions for the pdf of the transformed gamma and the inverse gamma by applying the definition of transforming and inverting a distribution. Apply the transformations to the base gamma distribution with \( \theta = 1 \).

**Solution:**

- **X** is the gamma random variable, with pdf \( f(x) = \frac{(x)^{\alpha}e^{-x}}{\Gamma(\alpha)} \).
- **Transformed gamma:** \( Y = X^{1/\tau} \); \( f_Y(y) = \tau y^{\tau-1}f_X(y^\tau) = \tau y^{\tau-1} \cdot \frac{(y^{1/\tau})^{\alpha}e^{-y^{1/\tau}}}{y^{\frac{\tau}{\alpha}}\Gamma(\alpha)} = \frac{\tau y^{\alpha-1}e^{-y}}{\Gamma(\alpha)} \).
- **Inverse gamma:** \( Y = X^{-1} \), \( f_Y(y) = y^{-2}f_X(y^{-1}) = y^{-2} \cdot \frac{(y^{-1})^{\alpha}e^{-y^{-1}}}{y^{-\alpha-1}\Gamma(\alpha)} = \frac{y^{\alpha-1}e^{-1/y}}{\Gamma(\alpha)} \).

This is the inverse gamma with \( \theta = 1 \).

**Exponential transformation**

The exponential transformation is

\[ Y = e^X = g(X) \, \text{so that} \, X = \ln(Y) = k(Y) \, . \]

Then

\[ F_Y(y) = F_X(\log y) \, , \, f_Y(y) = \frac{1}{y} \cdot f_X(\log y) \, . \quad (5.7) \]

For this transformation, if \( X \) is normal with mean \( \mu \) and variance \( \sigma^2 \), then \( Y = e^X \) has a lognormal distribution.

**Sums of certain random variables**

Suppose that \( X_1, X_2, ..., X_k \) are independent random variables and \( Y = \sum_{i=1}^{k} X_i \)

<table>
<thead>
<tr>
<th>distribution of ( X_i )</th>
<th>distribution of ( Y )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bernoulli ( B(1, q) )</td>
<td>binomial ( B(k, q) )</td>
</tr>
<tr>
<td>binomial ( B(n_i, q) )</td>
<td>binomial ( B(\Sigma n_i, q) )</td>
</tr>
<tr>
<td>Poisson ( \lambda_i )</td>
<td>Poisson ( \Sigma \lambda_i )</td>
</tr>
<tr>
<td>geometric ( \beta )</td>
<td>negative binomial ( r = k , , , \beta )</td>
</tr>
<tr>
<td>negative binomial ( r_i , , , \beta )</td>
<td>negative binomial ( \Sigma r_i , , , \beta )</td>
</tr>
<tr>
<td>normal ( N(\mu_i, \sigma_i^2) )</td>
<td>( N(\Sigma \mu_i, \Sigma \sigma_i^2) )</td>
</tr>
<tr>
<td>exponential with mean ( \theta )</td>
<td>gamma with ( \alpha = k , , , \theta )</td>
</tr>
<tr>
<td>gamma with ( \alpha_i , , , \theta )</td>
<td>gamma with ( \Sigma \alpha_i , , , \theta )</td>
</tr>
<tr>
<td>Chi-square with ( k_i ) df</td>
<td>Chi-square with ( \Sigma k_i ) df</td>
</tr>
</tbody>
</table>
MODELING - PROBLEM SET 5

Parametric Distributions and Transformations - Section 5

1. \( f_X(x) = \frac{e^{-\frac{1}{x}}}{x^2}, \ x > 0 \). \( Y = \theta X \). Find \( f_Y(y) \).
   A) \( \frac{e^{-\frac{\theta}{x}}}{x^2} \)
   B) \( \frac{\theta e^{-\frac{\theta}{x}}}{x^2} \)
   C) \( \frac{\theta^2 e^{-\frac{\theta}{x}}}{x^2} \)
   D) \( \frac{e^{-\frac{\theta}{x}}}{\theta x^2} \)
   E) \( \frac{\theta^2 e^{-\frac{\theta}{x}}}{\theta^2 x^2} \)

2. \( X \) has a uniform distribution on the interval \((0, c)\). \( Y = 2X \). Find the distribution of \( Y \).
   A) uniform on \((0, \frac{c}{2})\)
   B) uniform on \((0, 2c)\)
   C) uniform on \((c, 2c)\)
   D) uniform on \((c, 3c)\)
   E) None of A, B, C or D is correct

3. \( X \) has a uniform distribution on the interval \((0, c)\). \( Y = X^{1/2} \). Find the distribution of \( Y \).
   A) \( f_Y(y) = \frac{2y}{c} \) for \( 0 < y < \sqrt{c} \)
   B) \( f_Y(y) = \frac{2y}{c} \) for \( 0 < y < c^2 \)
   C) \( f_Y(y) = \frac{y^2}{c} \) for \( 0 < y < \sqrt{c} \)
   D) \( f_Y(y) = \frac{y^2}{c} \) for \( 0 < y < c^2 \)
   E) \( f_Y(y) = \frac{\sqrt{y}}{c} \) for \( 0 < y < \sqrt{c} \)

4. \( X \) has a uniform distribution on the interval \((0, c)\). \( Y = e^X \). Find the distribution of \( Y \).
   A) \( f_Y(y) = \frac{\ln y}{c}, \ 0 < \ln y < c, \ 1 < y < e^c \)
   B) \( f_Y(y) = \frac{1}{c \ln y}, \ 0 < \ln y < c, \ 1 < y < e^c \)
   C) \( f_Y(y) = \frac{1}{cy}, \ 0 < \ln y < c, \ 1 < y < e^c \)
   D) \( f_Y(y) = \frac{e^y}{c}, \ 0 < \ln y < c, \ 1 < y < \ln c \)
   E) \( f_Y(y) = \frac{1}{c e^y}, \ 0 < \ln y < c, \ 1 < y < \ln c \)

5. \( X \) has a Weibull distribution with parameters \( \tau \) and \( \theta \).
   \( Y = g(X) \) has an exponential distribution with mean \( 0^\tau \).
   Find the transformation \( g(X) \).

6. \( X \) has a Pareto distribution with parameters \( \alpha \) and \( \theta \). \( Y = \ln \left( \frac{X+\theta}{\theta} \right) \).
   Find the distribution of \( Y \).
7. \(X\) has an exponential distribution with mean \(\theta\). \(Y = e^X\). Find the distribution of \(Y\).

A) Weibull
B) Inverse Weibull
C) Exponential
D) Inverse Exponential
E) Single Parameter Pareto

8. Claim severities are modeled using a continuous distribution and inflation impacts claims uniformly at an annual rate of \(i\). Which of the following are true statements regarding the distribution of claim severities after the effect of inflation?
1. An Exponential distribution will have scale parameter \((1 + i)\theta\)
2. A 2-parameter Pareto distribution will have scale parameters \((1 + i)\alpha\) and \((1 + i)\theta\)
3. A Paralogistic distribution will have a scale parameter \(\theta/(1 + i)\)

A) 1 only        B) 3 only        C) 1 and 2 only        D) 2 and 3 only        E) 1, 2 and 3

9. Claim size, \(X\), follows a Pareto distribution with parameters \(\alpha\) and \(\theta\).
A transformed distribution, \(Y\), is created such that \(Y = X^{1/\tau}\).
Which of the following is the probability density function of \(Y\)?

A) \(\frac{\tau \theta y^{\tau - 1}}{(y + \theta)^{\tau + 1}}\)
B) \(\frac{\alpha \theta \tau y^{\tau - 1}}{(y + \theta)^{\tau + 1}}\)
C) \(\frac{\theta \alpha y^{\tau - 1}}{(y + \theta)^{\tau + 1}}\)
D) \(\frac{\alpha \tau (y/\theta)^{\tau}}{y[1+(y/\theta)^{\tau}]^{\tau+1}}\)
E) \(\frac{\alpha \theta^n}{(y + \alpha)^{n+\tau}}\)

10. The aggregate losses of Eiffel Auto Insurance are denoted in euro currency and follow a Lognormal distribution with \(\mu = 8\) and \(\sigma = 2\). Given that 1 euro = 1.3 dollars, which set of lognormal parameters describes the distribution of Eiffel's losses in dollars?

A) \(\mu = 6.15, \sigma = 2.26\)
B) \(\mu = 7.74, \sigma = 2.00\)
C) \(\mu = 8.00, \sigma = 2.60\)
D) \(\mu = 8.26, \sigma = 2.00\)
E) \(\mu = 10.40, \sigma = 2.60\)

11. The following information is available regarding the random variables \(X\) and \(Y\):
- \(X\) follows a Pareto distribution with \(\alpha = 2\) and \(\theta = 100\)
- \(Y = \ln[1 + (X/\theta)]\)
Calculate the variance of \(Y\).

A) Less than 0.1
B) At least 0.1, but less than 0.2
C) At least 0.2, but less than 0.3
D) At least 0.3, but less than 0.4
E) At least 0.4

12. Calculate the skewness of a Pareto distribution with \(\alpha = 4\) and \(\theta = 1,000\).

A) Less than 2
B) At least 2, but less than 4
C) At least 4, but less than 6
D) At least 6, but less than 8
E) At least 8
1. \( f_Y(y) = \frac{1}{\theta} \cdot f_X \left( \frac{y}{\theta} \right) = \frac{1}{\theta} \cdot e^{-\left(\frac{y}{\theta}\right)} \cdot \left(\frac{y}{\theta}\right)^{x-1} = \frac{\theta e^{-\theta/x}}{x^2} \). Answer: B

2. \( f_X(x) = \frac{1}{c} \) for \( 0 < x < c \). \( f_Y(y) = \frac{1}{2c} \cdot f_X \left( \frac{y}{2} \right) = \frac{1}{2c} \) for \( 0 < y < c \), or equivalently, \( f_Y(y) = \frac{1}{2c} \) for \( 0 < y < 2c \). \( Y \) is uniform on \( (0, 2c) \). Answer: B

3. \( f_Y(y) = 2y \cdot f_X(y^2) = \frac{2y}{c} \) for \( 0 < y^2 < c \), or equivalently, \( 0 < y < \sqrt{c} \). Answer: A

4. \( f_Y(y) = \frac{1}{e} \cdot f_X(e^y), \) or equivalently, \( 0 < e^y < c \). Answer: C

5. \( f_Y(y) = f_X(k(y)) \cdot k'(y), \) \( f_X(x) = \frac{x^{\alpha-1} e^{-(x/\theta)^\alpha}}{\theta^\alpha} \).
We are given that \( f_Y(y) = \frac{1}{\theta^r} e^{-y/\theta^r} \).
By inspection, it appears that \( y = x^r = g(x) \) may be the proper transformation.
With this transformation \( x = k(y) = y^{1/r} \), \( k'(y) = \frac{1}{r} y^{1-\frac{1}{r}} \),
and \( f_Y(y) = \frac{(y^{1/r})^{\alpha-1} e^{-(y^{1/r}/\theta)^\alpha}}{\theta^\alpha} \cdot \frac{1}{r} y^{1-1} = \frac{1}{\theta^r} \cdot e^{-y/\theta^r} \).
This is the pdf of an exponential random variable with mean \( \theta^r \).

6. \( f_Y(y) = f_X(k(y)) \cdot k'(y), \)
In this case, \( y = g(x) = \ln \left( \frac{x+\theta}{\theta} \right) \rightarrow x = \theta(e^y - 1) = k(y), \) and \( k'(y) = \theta e^y \).
For the Pareto distribution, \( f_X(x) = \frac{\alpha \theta^\alpha}{(x+\theta)^{\alpha+1}} \).
Then \( f_Y(y) = \frac{\alpha \theta^\alpha}{[\theta(e^y-1)+\theta]^{\alpha+1}} \cdot \theta e^y = \alpha e^{-\alpha y} \).
This is the pdf of an exponential distribution with mean \( \frac{1}{\alpha} \).

7. \( f_Y(y) = \frac{1}{y} \cdot f_X(ln y) = \frac{1}{y} \cdot \frac{1}{e} \cdot e^{-\frac{(ln y)/\theta}{\theta}} = \frac{1}{y} \cdot \frac{1}{\theta} \cdot \frac{1}{y^{\alpha}} = \frac{\alpha}{y^{\alpha+1}} \) for \( y = e^x > 1 \), where \( \alpha = \frac{1}{\theta} \).
\( Y \) has the pdf of a single parameter Pareto distribution with \( \theta = 1 \). Answer: E

8. To say that random variable \( X \) has scale parameter \( \theta \) means that if \( c \) is a constant and \( Y = cX \), then \( Y \) has the same distributional form as \( X \) with \( \theta \) replaced by \( c\theta \). In this case, after inflation the loss is \( Y = (1+i)X \), and so will have the same distributional type with scale parameter \( (1+i)\theta \).

The distributions in the Exam C Table have been formulated so that the parameter \( \theta \) is a scale parameter (the only continuous distribution in the table that do not use the parameter \( \theta \) is the lognormal, all others have scale parameter \( \theta \)).
8. continued
1. The exponential distribution has scale parameter $\theta$, so after inflation $Y = (1 + i)X$ will be exponential with parameter $\theta$ \textbf{True}.
2. The Pareto has scale parameter $\theta$, so after inflation \( Y = (1 + i)X \) will have a Pareto distribution with scale parameter $(1 + i)\theta$ (the Pareto distribution also has parameter $\alpha$, which is not a scale parameter). \textbf{False} Answer: A
3. The Paralogistic distribution has scale parameter $\theta$, $Y = (1 + i)X$ will have a Paralogistic distribution with scale parameter $(1 + i)\theta$ (the Paralogistic distribution also has parameter $\alpha$, which is not a scale parameter). \textbf{False} Answer: A

9. We use the following transformation rule. If $X$ has pdf $f_X(x)$ and if $Y = h(X)$, then we find the inverse transformation $X = k(Y)$. The pdf of $Y$ is $f_Y(y) = f_X(k(y)) \cdot |k'(y)|$. In this example, $f_X(x) = \frac{\alpha \theta^\alpha}{(x + \theta)^{\alpha+1}}$ (Pareto).
The transformation is $Y = h(X) = X^{1/\tau}$. The inverse transformation is $X = Y^\tau = k(Y)$.
The pdf of $Y$ is $f_Y(y) = f_X(y^\tau) \cdot \tau y^{\tau - 1} = \frac{\alpha \theta^\alpha}{(y^\tau + \theta)^{\alpha+1}} \cdot \tau y^{\tau - 1} = \frac{\alpha \theta^\alpha \tau y^{\tau - 1}}{(y^\tau + \theta)^{\alpha+1}}$.

10. The mean of a lognormal distribution with parameters $\mu$ and $\sigma$ is $e^{\mu + \frac{1}{2} \sigma^2}$ and the second moment is $e^{2\mu + 2\sigma^2}$.
If the loss measured in euros is $X$, then $E[X] = e^{10}$ and $E[X^2] = e^{24}$.
The loss measured in dollars is $Y = 1.3X$, with mean $E[Y] = E[1.3X] = 1.3e^{10}$ and with second moment $E[Y^2] = E[(1.3X)^2] = 1.69e^{24}$.
If $Y$ is lognormal with parameters $\mu'$ and $\sigma'$, then $E[Y] = e^{\mu' + \frac{1}{2} \sigma'^2}$ and $E[Y^2] = e^{2\mu' + 2\sigma'^2}$.
Therefore, $\mu' + \frac{1}{2} \sigma'^2 = \ln(1.3e^{10}) = \ln 1.3 + 10$, and $2\mu' + 2\sigma'^2 = \ln 1.69 + 24$.
Solving this system of two equations in $\mu'$ and $\sigma'$ results in $\mu' = 8.26$ and $\sigma' = 2.00$.
Answer: D

11. $Y = \ln(1 + \frac{X}{\theta}) = g(X)$, We can find the pdf of $Y$ from the relationship $f_Y(y) = f_X(k(y)) \cdot |k'(y)|$, where $k(y)$ is the inverse function to $g(x)$.
From $Y = \ln(1 + \frac{X}{\theta})$, we get $e^Y = 1 + \frac{X}{\theta}$, and then $X = \theta (e^Y - 1) = k(Y)$ is the inverse function. Therefore, $f_Y(y) = f_X(\theta (e^y - 1)) \cdot (\theta e^y)$.
The pdf of the Pareto distribution is $f_X(x) = \frac{\alpha \theta^\alpha}{(x + \theta)^{\alpha+1}}$.
Then, $f_Y(y) = \frac{\alpha \theta^\alpha}{(\theta (e^y - 1) + \theta)^{\alpha+1}} \cdot (\theta e^y) = \alpha e^{-\alpha y}$.
This is the pdf of an exponential distribution with mean $\frac{1}{\alpha}$, so the variance is $\frac{1}{\alpha^2}$.
We are given $\alpha = 2$, so the variance of $Y$ is .25. Answer: C

12. The skewness is $\frac{E[(X - E[X])^3]}{(Var[X])^{3/2}} = \frac{E[X^3] - 3E[X^2]E[X] + 2(E[X])^3}{[E[X^2] - (E[X])^2]^{3/2}}$.
For the Pareto distribution, we have $E[X] = \frac{\theta}{\alpha - 1}$, $E[X^2] = \frac{2\theta^2}{(\alpha - 2)(\alpha - 1)}$, and $E[X^3] = \frac{1000}{6}$.
The skewness is $\frac{1000^3 - 3(1000^2/3)(1000/3) + 2(1000/3)^3}{[1000^2/3 - (1000/3)^2]^{3/2}} = \frac{1 - \frac{1}{4} + \frac{2}{7}}{\left(\frac{1}{4} - \frac{2}{7}\right)^{3/2}} = 7.07$.
Answer: D
MODELING SECTION 6 - DISTRIBUTION TAIL BEHAVIOR

The material in this section relates to Section 3.4 of "Loss Models". The suggested time frame for this section is 1 hour. There is been very infrequent reference to this topic on the exam. Later topics do not depend on this material, and it can be postponed and covered at a later time.

LM-6.1 Measuring Tail Weight Using Existence of Moments

A right tail of the distribution for random variable $X$ is an interval of the form $(x, \infty)$, with probability

$$P[X > x] = S_X(x) = 1 - F_X(x) = \int_x^\infty f_X(t) \, dt$$

($S_X$ is the survival function of $X$). A random variable with a lot of probability in the right tails is said to have "heavy right tails", or just heavy tails. Heavy tails are characteristic of a random variable that has relatively high probability for large numerical outcomes. The opposite would be true for a light-tailed distribution. There are various ways of classifying tails as heavy or light.

One classification considers the moments of $X$. Under this classification, distributions for which $E[X^k]$ is finite for all $k > 0$ indicate a light right tail, and distributions for which $E[X^k]$ is infinite for $k$ above a certain value indicate a heavy right tail. Any distribution on $(0, \infty)$ whose pdf is proportional (or asymptotically proportional) to $\frac{1}{x^n}$ will have heavy right tails because $E[X^k]$ will be infinite for $k \geq n - 1$. Any distribution $(0, \infty)$ whose pdf is proportional to $x^b e^{-bx}$, with $b > 0$, will have a light right tail since

$$\int_0^\infty x^m e^{-bx} \, dx < \infty \text{ if } m > 0 \text{ and } b > 0.$$  

Heavy right-tailed distributions from the Exam C table of distributions based on this existence of moments classification are: Pareto, inverse Pareto, loglogistic, paralogistic, inverse paralogistic, inverse gamma (and inverse exponential) and inverse Weibull.

Light right-tailed distributions from the Exam C table of distributions based on this existence of moments classification are: gamma (and exponential), Weibull, normal, lognormal, and inverse Gaussian.

LM-6.2 Comparing the Tail Weights of Two Distributions

The tail weights of two distributions can be compared by taking the limit of the ratio of their survival functions. Suppose that $X$ and $Y$ are two continuous random variables. Then

$$\lim_{x \to \infty} \frac{S_X(x)}{S_Y(x)} = \lim_{x \to \infty} \frac{f_X(x)}{f_Y(x)} \quad (\text{this follows from l'Hospital's rule}).$$  

Suppose that the limit is

$$\lim_{x \to \infty} \frac{S_X(x)}{S_Y(x)} = \lim_{x \to \infty} \frac{f_X(x)}{f_Y(x)} = c$$

We define the relative tail weights of $X$ and $Y$ as follows:

- if $c = 0$, we say that $X$ has a lighter right tail than $Y$
- if $0 < c < \infty$, we say that $X$ and $Y$ have similar (proportional) right tails
- if $c = \infty$, we say that $X$ has a heavier right tail than $Y$
Example LM6-1:

Compare the tail weights of the inverse Pareto distribution and the inverse gamma distribution (assume the inverse gamma distribution has $\alpha > 1$).

Solution:

\[
\frac{f_{\text{inv-Pareto}}(x)}{f_{\text{inv-gamma}}(x)} = \frac{x^\alpha e^{-x/(x+\theta)}}{(\theta e^{-x/(x+\theta)})^\alpha} = x^{\alpha-1} e^{\theta/x}. \quad \text{for } x > 0.
\]

\[
\lim_{x \to \infty} e^{\theta/x} = e^0 = 1; \quad \frac{x^\alpha}{(x+\theta)^{\alpha+1}} \cdot x^{\alpha-1} \to \infty \text{ as } x \to \infty \text{ (since } \alpha > 1\).
\]

The inverse Pareto has a heavier tail than the inverse gamma with $\alpha > 1$.

---

LM-6.3 Measuring Tail Weight Using Hazard Rate and Mean Residual Lifetime

The hazard rate (the hazard rate is the force of mortality in survival analysis terminology, and it is also called failure rate) is

\[
h(x) = \frac{f(x)}{S(x)} = \frac{F'(x)}{1-F(x)} = -\frac{d}{dx} \log S(x) = -\frac{S'(x)}{S(x)}.
\]  \tag{6.2}

If $X \geq 0$ (a non-negative random variable), then $S(x) = e^{-\int_0^x h(t) \, dt}$.

In Exam MLC, the notation $T(x)$ is the continuous random variable representing time until death of someone now alive at age $x$. The expected value of $T(y)$ is the expected time until death for someone at age ($y$) which is $E[T(y)] = \bar{\theta}_y = \int_0^\infty t \, \mu_x \, dt$, called the complete expectation of life.

If $X$ represents the random variable of age at death, then $E[T(x)] = E[X - X|X > x]$. What this indicates is that a newborn must survive to age $x$, and then we measure the time until death from age $x$ for someone still alive. $\bar{\theta}_x$ is also referred to as the mean residual lifetime (given survival to $x$), because it measures the average additional number of years until death from age $x$ given that an individual has survived to age $x$.

You may recall from Exam MLC that

\[
\bar{\theta}_x = \int_0^\infty t \, P_x \, dt = \int_0^\infty S_X(x+t) \, dt = \frac{\int_0^\infty S_X(u) \, du}{S_X(x)},
\]

so that

\[
\bar{\theta}_x = \int_0^{\infty} \frac{S_X(x+t)}{S_X(x)} \, dt = \frac{\int_0^{\infty} S_X(u) \, du}{S_X(x)}.
\]

It is natural to describe mean residual lifetime in terms of an age at death random variable $X$ as we just have done. Algebraically, we can define the mean residual lifetime for any non-negative random variable in the same way. We might see the notation $\bar{\theta}(x)$ instead of $\bar{\theta}_x$. Mean residual lifetime will be an important concept that arises again when we consider policy deductibles a little later on in the study guide.

Tail weight of a continuous distribution can be classified by the behavior of the hazard rate and also by the behavior of the mean residual lifetime. Distributions with increasing hazard rate functions have a light tail and those with decreasing hazard rate functions have a heavy tail. The following is a summary of some relationships involving tail weight, hazard rate, survival function and mean residual lifetime of a random variable.
Light right tail corresponds to the following conditions
\[ \frac{f(x+y)}{f(x)} \text{ is a decreasing function of } x \text{ for all values of } y \geq 0 \]
\[ \Rightarrow \text{ the hazard rate } h(x) \text{ is an increasing function of } x \]
\[ \Rightarrow e(x) \text{ (mean residual lifetime) is a decreasing function of } x \]
\[ \Rightarrow \frac{\sqrt{\text{Var}[X]}}{E[X]} \leq 1 \text{ (coefficient of variation of } X \text{ is } \leq 1) \]

(reverse implications are not true, in general).

\[ X \text{ has an increasing hazard rate } \iff \frac{S(x+y)}{S(x)} \text{ is a decreasing function of } x. \]  \hspace{1cm} (6.3)

Examples of such a distribution are gamma with \( \alpha > 1 \) and Weibull with \( \tau > 1 \). Note that the exponential distribution has a constant hazard rate, but all moments exist, so that it is considered a light right-tailed distribution using the existence of moments criterion.

Heavy right tail corresponds to the following conditions
\[ \frac{f(x+y)}{f(x)} \text{ is an increasing function of } x \text{ for all values of } y \geq 0 \]
\[ \Rightarrow \text{ the hazard rate } h(x) \text{ is a decreasing function of } x \]
\[ \Rightarrow e(x) \text{ is an increasing function} \]
\[ \Rightarrow \frac{\sqrt{\text{Var}[X]}}{E[X]} \geq 1 \text{ (coefficient of variation of } X \text{ is } \geq 1) \]

(reverse implications are not true, in general).

\[ X \text{ has a decreasing hazard rate } \iff \frac{S(x+y)}{S(x)} \text{ is an increasing function of } x. \]  \hspace{1cm} (6.4)

Examples of such a distribution are Pareto, inverse Pareto, inverse gamma, inverse Weibull, gamma with \( \alpha < 1 \) and Weibull with \( \tau < 1 \).

Note that all moments of the gamma distribution exist even if \( 0 < \alpha < 1 \), so that when the existence of moments is used as the measure of the right tail weight, the gamma always has light right tails. On the other hand, if the behavior of the hazard rate is used as the measure of right-tail behavior then the gamma has a light right-tail if \( \alpha > 1 \) and a heavy right tail if \( \alpha < 1 \) (the concept of heavy/light right-tail becomes somewhat vague in this case, and it may be more meaningful when comparing the relative tail weights of two distributions). A similar comment applies to the Weibull distribution. All moments of the Weibull distribution exist, but the hazard rate is increasing when \( \tau > 1 \), the hazard rate is constant when \( \tau = 1 \), and the hazard rate is decreasing when \( \tau < 1 \).

The following are some additional relationships involving hazard rate and mean residual lifetime.

- The mean residual lifetime is \( e(x) = \frac{\int_0^\infty f(t) \, dt}{S(x)} = \int_0^\infty \frac{S(t) \, dt}{S(x)} = \int_0^\infty \frac{S(x+y)}{S(x)} \, dy \), and
  \[ S(x) = \frac{e(0)}{e(x)} \exp\left[-\int_0^x \frac{1}{e(t)} \, dt\right] \text{ for } x \geq 0. \]  \hspace{1cm} (6.5)

- \( \lim_{x \to \infty} e(x) = \lim_{x \to \infty} \frac{1}{h(x)} \).  \hspace{1cm} (6.6)

- \[ S(x) = \frac{e(0)}{e(x)} \cdot \exp\left[-\int_0^x \frac{1}{e(t)} \, dt\right] \]  \hspace{1cm} (6.7)

- The equilibrium distribution of the random variable \( X \) has pdf \( f_\epsilon(x) = \frac{S(x)}{E[X]} \).  \hspace{1cm} (6.8)

The hazard rate of the equilibrium distribution is \( h_\epsilon(x) = \frac{1}{e(x)} \), where \( e(x) \) is the mean residual lifetime of \( X \).
Example LM6-2:

X has mean residual lifetime \( e(x) = x + 1 \) for \( x \geq 0 \).

Find \( S(x) \), \( f(x) \) and \( h(x) \). Determine the tail behavior of \( X \) by considering the moments of \( X \), the behavior of the hazard rate, and the behavior of the mean residual lifetime. Find the pdf of the equilibrium distribution.

Solution:

\[
S(x) = \frac{e(0)}{e(x)} \cdot \exp\left[ - \int_0^x \frac{1}{e(t)} \, dt \right] = \frac{1}{x+1} \cdot \exp\left[ - \int_0^x \frac{1}{x+1} \, dx \right] = \frac{1}{x+1} \cdot \exp\left[ - \ln(x+1) \right] = \frac{1}{(x+1)^\gamma}.
\]

Then \( f(x) = - S'(x) = \frac{2}{(x+1)^{\gamma+1}} \), and \( h(x) = \frac{f(x)}{S(x)} = \frac{2}{x+1} \).

The first moment of \( X \) is \( E[X] = \int_0^\infty x \cdot f(x) \, dx = \int_0^\infty \frac{2x}{(x+1)^\gamma} \, dx = 1 \)

(The integral can be found by using the substitution \( u = x + 1 \)).

The second moment of \( X \) is \( E[X^2] = \int_0^\infty x^2 \cdot f(x) \, dx = \int_0^\infty \frac{2x^2}{(x+1)^\gamma} \, dx = \infty \).

Using the moment criterion for tail weight indicates that \( X \) has a heavy tail.

The hazard rate is decreasing and the mean residual lifetime is increasing, which also is an indication of a heavy tail.

The equilibrium distribution has pdf \( f_\epsilon(x) = \frac{S(x)}{E[X]} = \frac{1}{(x+1)^\gamma} \) for \( x \geq 0 \). \( \square \)

Example LM6-3:

Suppose that \( X \) has a Weibull distribution with parameters \( \theta \) and \( \tau \), and pdf \( f(x) = \frac{\tau(x/\theta)^{\tau-1}e^{-(x/\theta)^\tau}}{\theta^\tau} \).

Find the hazard rate, and determine the behavior of the mean residual lifetime of \( X \).

Solution:

The survival function for this distribution is \( S(x) = e^{-(x/\theta)^\tau} \).

The hazard rate is \( h(x) = \frac{f(x)}{S(x)} = \frac{\tau x^{\tau-1}}{\theta^{\tau}} \).

If \( \tau > 1 \) then \( \lambda(x) \) is an increasing function of \( x \), and therefore \( X \) has a decreasing mean residual lifetime. If \( \tau < 1 \), the distribution has a decreasing hazard rate and an increasing mean residual lifetime. \( \square \)
1. Using the criterion of existence of moments, determine which of the following distributions have heavy tails.

I. Normal distribution with mean $\mu$, variance $\sigma^2$.

II. Lognormal distribution with parameters $\mu$ and $\sigma^2$.

III. Single parameter Pareto.

A) I only  
B) II only  
C) III only  
D) All but I  
E) All but II

2. You are given that $X$ has pdf $f(x) = \frac{4/\pi}{1+x^2}$ for $0 < x < \infty$.

How many of the following distributions have a lighter right tail than $X$?

I. Pareto with $\alpha = 1$

II. Pareto with $\alpha > 1$

III. Paralogistic with $\alpha = 1$

IV. Inverse paralogistic with $\tau > 1$

A) 0  
B) 1  
C) 2  
D) 3  
E) 4

3. $X$ has pdf $f(x) = 2xe^{-x^2}$, $x > 0$.

(a) Find $S(x)$ and determine whether or not $\frac{S(x+y)}{S(x)}$ is an increasing function of $x$.

(b) Find the hazard rate and determine whether it is increasing or decreasing.

(c) Use the fact that the standard normal distribution is symmetric around the origin and that $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = 1$ to show that $E[X] = \int_{0}^{\infty} e^{-x^2} dx = \frac{\sqrt{\pi}}{2}$.

(d) Find the pdf of the equilibrium distribution for $X$. 

Actex Learning

SOA Exam C - Construction and Evaluation of Actuarial Models
MODELING - PROBLEM SET 6 SOLUTIONS

1.   I.  The moment generating function of the normal is \( M_X(t) = e^{\mu t + \frac{1}{2} \sigma^2 t^2} \). Each successive derivative exists and is finite when \( t = 0 \). All moments exist, so the tail is not heavy.

II.  From the Exam C table, the \( k \)th moment of the lognormal is \( E[X^k] = e^{\mu k + \frac{1}{2} \sigma^2 k^2} \), which is finite for every \( k \). All moments exist, so the tail is not heavy.

III. From the Exam C table, the \( k \)th moment of the single parameter Pareto with parameters \( \alpha \) and \( \theta \) is \( E[X^k] = \frac{\alpha^k}{\alpha - k} \), which exists only for \( k < \alpha \). The tail is heavy.  Answer: C

2.   I.  \( \frac{f(x)}{f_1(x)} = \frac{4/\pi}{1+x^2} \cdot \frac{(x+\theta)^2}{\theta} \to \frac{4}{\pi \theta} \) as \( x \to \infty \). Same right tail weight.

II.  \( \frac{f(x)}{f_{11}(x)} = \frac{4/\pi}{1+x^2} \cdot \frac{(x+\theta)^{n+1}}{\theta^{n+1}} = \frac{4}{\pi \theta^{n+1}} \cdot \frac{(x+\theta)^2}{1+x^2} \cdot (x + \theta)^{n-1} \to \infty \) as \( x \to \infty \).  \( X \) has heavier right tail weight.

III. \( \frac{f(x)}{f_{111}(x)} = \frac{4/\pi}{1+x^2} \cdot \theta[1 + (x/\theta)]^2 = \frac{4 \theta^2}{\pi} \cdot \frac{(1+(x/\theta)^2)}{1+x^2} \to \frac{4 \theta}{\pi} \) as \( x \to \infty \). Same right tail weight.

IV.  \( \frac{f(x)}{f_{1111}(x)} = \frac{4/\pi}{1+x^2} \cdot \frac{x[1+(x/\theta)^2]}{\tau^2(x/\theta)^2} = \frac{4}{\pi \tau^2} \cdot \frac{x[1+(x/\theta)^2]}{1+x^2} \to \infty \) as \( x \to \infty \) since \( \tau > 1 \).  \( X \) has heavier right tail weight.  Answer: C

3.   (a) \( S(x) = \int_x^\infty f(t) \, dt = \int_x^\infty 2te^{-t^2} \, dt = e^{-x^2} \) (the antiderivative of \( 2te^{-t^2} \) is \( -e^{-t^2} \)). \[ \frac{S(x+y)}{S(x)} = e^{-(x+y)^2} / e^{-x^2} = e^{-2xy-y^2} , \] which is a decreasing function of \( x \) for any \( y > 0 \).

(b) The hazard rate is \( h(x) = \frac{f(x)}{S(x)} = \frac{2xe^{-x^2}}{e^{-x^2}} = 2x \), which is an increasing function of \( x \). This is also implied by (a), since \( \frac{S(x+y)}{S(x)} \) decreasing is equivalent to \( h(x) \) increasing.

(c) Since the standard normal density is symmetric around the origin, it follows that
\[ \int_0^\infty \frac{1}{\sqrt{2\pi}} \cdot e^{-x^2/2} \, dx = \frac{1}{2} \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} \cdot e^{-x^2/2} \, dx = \frac{1}{2} , \]
and then \( \int_0^\infty e^{-x^2/2} \, dx = \frac{\sqrt{\pi}}{2} \).

Then, with the change of variable \( x = t \sqrt{2} \), the integral becomes
\[ \int_0^\infty e^{-x^2/2} \, dx = \int_0^\infty \sqrt{2} \cdot e^{-t^2} \, dt = \sqrt{\frac{\pi}{2}} , \]
so that \( \int_0^\infty e^{-t^2} \, dt = \sqrt{\frac{\pi}{2}} \) and \( E[X] = \int_0^\infty S(x) \, dx = \int_0^\infty e^{-x^2} \, dx = \sqrt{\frac{\pi}{2}} \).

(d) The pdf of the equilibrium distribution is \( f_e(x) = \frac{S(x)}{E[X]} = \frac{e^{-x^2}}{\sqrt{\pi/2}} \).