PREFACE

A cursory search of Google Books reveals thousands of titles with the words *Probability and Statistics*, or *Mathematical Statistics*, thus prompting the inevitable question, why yet another? In view of this superabundance of choices we feel impelled to offer a few words of explanation in our defense for adding to this already crowded marketplace.

The idea for this textbook evolved out of a two-semester course in probability and statistics we have been offering for many years to sophomore and qualified freshman students, predominantly actuarial science majors. The goals of our course are twofold: to lay the foundations of calculus-based probability theory for our students, and to prepare them to pass the actuarial exam covering this material as early as possible in their collegiate careers.

The primary market segment we are targeting consists of students like ours - freshmen and sophomores - who are studying calculus-based probability while simultaneously learning the selfsame calculus it is based upon. We desire that our actuarial students be prepared to pass Exam P/1 (jointly offered by the Society of Actuaries (SOA) and the Casualty Actuarial Society (CAS)) no later than the end of their sophomore year. Consequently, the probability and statistics component of their education tends to overlap with the typical 3-semester calculus sequence they all take.

Our experience has been that the myriad existing textbooks in probability and statistics fall into two types. They are either designed to support a "calculus-free" environment suitable for general business school statistics courses, or they are intended for more advanced, calculus-based mathematical statistics courses for juniors and seniors with the requisite technical background. The first category does not provide the depth of understanding required for the actuarial exam, and the second type of book tends to be too formal and advanced for our students.

For these reasons we decided to produce an introductory text in calculus-based probability and statistics whose level is comparable to a modern-day calculus book, and that could reasonably be used by freshman or sophomore students studying the material concurrently with their calculus classes. We consciously strive to pace the material in a way that makes it accessible to a student whose background consists of just one semester of college-level calculus. This might be, for example, either entering freshmen with AP or high-school/college credit concurrently taking Calculus II, or sophomores with one or two semesters of calculus already under their belts.

Chapters 1-4 present the rudiments of probability theory for discrete distributions with little or no reference to calculus topics, save for the basic knowledge of infinite series required for understanding the geometric and the Poisson distributions. Chapter 5, entitled "Calculus, Probability, and Continuous Distributions" introduces continuous random variables and is the first chapter heavily dependent on derivatives and integrals. The material on continuous, jointly distributed random variables comes in Chapter 7, by which time our students will have been introduced to double integrals in their Calculus III class. The second part of the book, comprising Chapters 9-11, covers all of the syllabus topics for the statistics portion of CAS Exam 3L – Life Contingencies and Statistics Segment. Taken as a whole, this book provides ample content to serve as the text for the standard two-semester introductory sequence in mathematical statistics and probability.

The text contains nearly 800 exercises, many with multiple parts. Numerical answers are given in the back of the book and a supplementary manual with complete solutions is available separately. Many of the exercises, and some examples, are based on previous actuarial exam questions released by the SOA and the CAS. All of the SOA/CAS Exam P/1 Sample Exam Questions (142 in number at press time in late 2009) have been incorporated into the text. In addition, we have used many statistics questions from CAS Exam 3 and 3L, as well as questions from the earlier Exam 110 (Probability and Statistics), in Chapters 8-11. We are grateful to the Society of Actuaries[1] and the Casualty Actuarial Society[2] for their permission to use these materials.

While designed primarily with the actuarial audience in mind, we hope this text will have appeal for a broader audience of mathematics and statistics students. We have sought to engage students with a light, informal style, emphasizing detailed explanations and providing a multitude of examples. At the same time, we have sought at each stage to present a sufficient glimpse into the theoretical underpinnings to make the text suitable for more advanced students. The overarching theme however, is problem solving, and we emphasize the requisite skills throughout. We hope to have a struck a balance that will allow students at all levels to benefit from a close reading of the text.

We have benefited from the many helpful comments and valuable insights provided by our reviewers: Carolyn Cuff, Westminster College; Thomas Herzog, Department of Housing and Urban Development; Thomas Lonergan, CIGNA; Jeffrey Mitchell, Robert Morris University; Emiliano Valdez, University of Connecticut; Charles Vinsonhaler, University of Connecticut. We would also like to acknowledge the invaluable editorial assistance provided by Gail Hall of ACTEX, whose firm but gentle hand managed to guide this project to completion. We are also indebted to Marilyn Baleshiski, whose patience and skill rendered a lumpy manuscript into a finished text. Our thanks to you all. Needless to say, for all remaining errors and indiscretions, the authors have only each other to blame.

Finally, we note that personages (both named and unnamed) who appear in various exercises and examples are completely fictitious and any resemblance to real people (living or departed) is purely coincidental.

Len Asimow, Moon Township, PA
Mark Maxwell, Austin, TX

---

[1] Reprinted with permission of the Society of Actuaries
[2] Reprinted with permission of the Casualty Actuarial Society

$$
\begin{aligned}
Cov[U_1,U_2] &= \frac{B(B-1)}{N(N-1)} - \left(\frac{B}{N}\right)^2 \\
&= \left(\frac{B}{N}\right)\left(\frac{(B-1)}{(N-1)} - \frac{B}{N}\right) \\
&= -\left(\frac{B}{N}\right)\left(\frac{G}{N}\right)\left(\frac{1}{N-1}\right) = -\left(\frac{BG}{N^2}\right)\left(\frac{1}{N-1}\right).
\end{aligned}
$$

Now we look at the $n \times n$ box, which has n diagonal elements and $n(n-1)$ elements off the main diagonal to conclude,

$Var[W] =$

	U_1	U_2	$\cdots$	U_n
U_1	$\frac{BG}{N^2}$	$-\left(\frac{BG}{N^2}\right)\left(\frac{1}{N-1}\right)$	$\cdots$	$-\left(\frac{BG}{N^2}\right)\left(\frac{1}{N-1}\right)$
U_2	$-\left(\frac{BG}{N^2}\right)\left(\frac{1}{N-1}\right)$	$\frac{BG}{N^2}$	$\cdots$	$-\left(\frac{BG}{N^2}\right)\left(\frac{1}{N-1}\right)$
$\vdots$	$\cdots$	$\cdots$	$\ddots$	$\cdots$
U_n	$-\left(\frac{BG}{N^2}\right)\left(\frac{1}{N-1}\right)$	$-\left(\frac{BG}{N^2}\right)\left(\frac{1}{N-1}\right)$	$\cdots$	$\frac{BG}{N^2}$

$$
\begin{aligned}
&= n\frac{BG}{N^2} - n(n-1)\left(\frac{BG}{N^2}\right)\left(\frac{1}{N-1}\right) \\
&= \left(n\frac{BG}{N^2}\right)\left(1-\frac{n-1}{N-1}\right) = n\left(\frac{B}{N}\right)\left(\frac{G}{N}\right)\left(\frac{N-n}{N-1}\right).
\end{aligned}
$$

8.4 THE CONDITIONING FORMULAS

We now return to an old and familiar problem, that of calculating the mean and the variance of a given random variable Y. The novelty here is that we are given partial information about Y that depends on a second random variable X. More formally, Y is *conditioned* on X. In Chapter 7, we discussed how to derive the conditional distribution of Y given X from the joint distribution, $f(x,y)$, of X and Y. Here, we will work backwards, deriving the mean and variance for the distribution of Y from the conditional distribution $Y \mid X$.

Suppose, for example, that an auto insurance company insures two classes of customers – good drivers and bad drivers. The random variable Y represents the annual claim amount for a customer, and the company wishes to calculate the expected claim amount, $E[Y]$, and the variance, $\text{Var}[Y]$, of the claim amount. Reasonably enough, the insurance claim amount should depend on the class of driver insured. So, we use the second random variable, X to signify the class of driver (good or bad). Then, the conditional random variable, $Y \big| X = \text{good}$, will have one probability distribution, while the conditional random variable, $Y|X = \text{bad}$, will likely have a different distribution. That is, the annual insurance claim is *conditioned* by class X.

The perspective here is quite similar to the Bayesian approach to finding probabilities, discussed in Section 2.6. However, the goal now is to find the overall expectation and variance of the claim amount by using the conditional distributions. This will give us two levels of distributions, the conditional distributions $Y \mid X$ of claims by class, and then the distribution of X, representing the classes themselves.

Expectation and Variance by Conditioning

Let X and Y be jointly distributed random variables. Then,

(1) $E[Y] = E_X\left[\boldsymbol{E_Y[Y|X]}\right]$. This is called the ***double expectation formula***.

(2) $Var[Y] = E_X\left[\boldsymbol{Var_Y\left[Y|X\right]}\right] + Var_X\left[\boldsymbol{E_Y\left[Y|X\right]}\right]$.

These formulas, while quite useful, can be difficult to interpret and apply without practice. Before providing derivations we will illustrate their use with some examples. In reading through the examples, keep in mind that:

(1) $Y \mid X$ merely represents some random variable. Think of X as a known fixed value (like being a bad driver).

(2) Compute the inside quantities $\boldsymbol{E_Y\left[Y \mid X\right]}$ and $\mathbf{Var}_{\boldsymbol{Y}}\left[\boldsymbol{Y \mid X}\right]$ in the usual manner for computing mean and variance.

(3) Realize that these quantities $\boldsymbol{E_Y\left[Y \mid X\right]}$ and $\mathbf{Var}_{\boldsymbol{Y}}\left[\boldsymbol{Y \mid X}\right]$ are themselves random variables whose values depend on the given value of X. That is, these are both transformations of the random variable X.

(4) Therefore, we can compute the outside expectation and variance, which is calculated with respect to X.

(5) In many applications of these formulas, each value of X can be considered a group, such as the group of good drivers or the group of bad drivers. Formula (1) can be thought of as the weighted average of the group averages. The expression on the right in formula (2) is sometimes referred to as "the variance within groups" plus "the variance between groups." This uses the language of ***analysis of variance*** (ANOVA), an important statistical tool.

Example 8.4-1 Expectation by Conditioning

A certain professor of our acquaintance works in Moon Township, PA and lives in Pittsburgh, PA, about a 25 mile commute. The professor randomly chooses from 3 different routes home in a futile attempt to evade rush hour traffic. The routes are identified by the name of a major bridge along the way. The professor has accumulated data over a lengthy period of time on the mean drive times of the three routes. Using the data summary that follows, calculate the overall expected drive time.

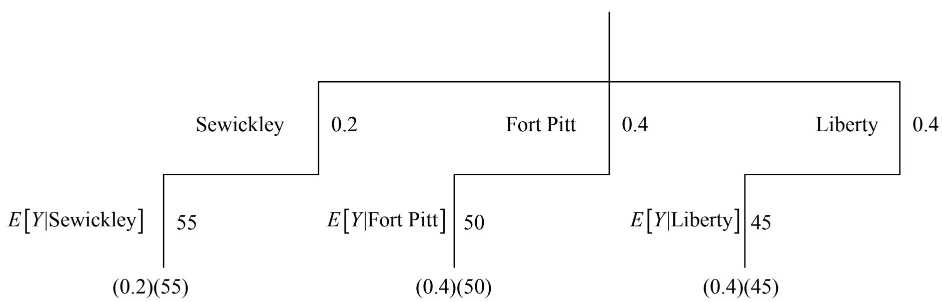
Route	**Probability of Route**	**Expected Time of Route**
Sewickley Bridge	0.2	55 minutes
Fort Pitt Bridge	0.4	50 minutes
Liberty Bridge	0.4	45 minutes

Solution

In the context of formula (1) we let X represent the route selected, and we let Y denote the drive time. Then X takes the three values: *Sewickley*, *Fort-Pitt*, and *Liberty*, with the probabilities shown. The drive times in the **3**rd column are conditioned upon the route chosen, and are the 3 values of $E_Y[Y \mid X]$. The calculation for the average drive time, $E[Y] = E_X\left[E_Y[Y \mid X]\right] = \sum_X \Pr[X] \cdot E_Y[Y \mid X]$, is easily found by summing the fourth column.

$X =$ **Route**	$\Pr[X]$	$E_Y[Y \mid X]$	$\Pr[X] \cdot E_Y[Y \mid X]$
Sewickley	0.2	55	$(0.2)(55) = 11$
Fort Pitt	0.4	50	$(0.4)(50) = 20$
Liberty	0.4	45	$(0.4)(45) = 18$
Total			49

This amounts to the same weighted average calculation used to determine the probability of an event by conditioning, as often displayed in a tree diagram. The only difference is that instead of conditional probabilities along the branches we have conditional expectations:



$$
\begin{aligned}
E[Y] = E\left[E[Y \mid X]\right] &= E[Y \mid X = \text{Sewickley}] \cdot \Pr[X = \text{Sewickley}] \\
&\quad + E[Y \mid X = \text{Fort Pitt}] \cdot \Pr[X = \text{Fort Pitt}] \\
&\quad + E\left[Y \mid X = \text{Liberty}\right] \cdot \Pr\left[X = \text{Liberty}\right] \\
&= (55)(0.2) + (50)(0.4) + (45)(0.4) = 49.
\end{aligned}
$$

❐

Note

Strictly speaking, X is not a random variable since it is not numerical-valued. Its values are labels rather than numbers, but everything works out correctly. If this were an issue we could assign the numbers 1, 2, 3 to the routes, but nothing in the calculations would be affected.

Example 8.4-2 Variance by Conditioning

The professor of Example 8.4-1 has also calculated variances in the times of the three routes. The information is displayed in a new summary table:

X = Route	Pr[X]	$E_Y[Y \mid X]$	$Var_Y[Y \mid X]$
Sewickley	0.2	55	10
Fort Pitt	0.4	50	25
Liberty	0.4	45	100

Using the conditional variances, calculate the overall variance in drive time.

Solution

We can play the same game as above, with the conditional variances, to calculate $E\left[Var_Y[Y \mid X]\right]$. But, as we see from formula (2) for the variance of Y, this is only half the story. We also need to calculate $\text{Var}_X\left[E_Y[Y \mid X]\right]$, which means the variance of the $E_Y[Y \mid X]$ column in the summary table. This will be done using the standard variance formula $\text{Var(whatever)} = E(\text{whatever}^2) - \left(E(\text{whatever})\right)^2$, which in this context means:

$$\begin{aligned}\text{Var}_X\left(E_Y[Y \mid X]\right) &= E_X\left(E_Y[Y \mid X]^2\right) - \left(E_X\left[E_Y[Y \mid X]\right]\right)^2 \\ &= E_X\left[E_Y[Y \mid X]^2\right] - \left(E[Y]\right)^2.\end{aligned}$$

The last equality follows from the double expectation formula for $E[Y] = E_X\left[\boldsymbol{E_Y}\left[\boldsymbol{Y \mid X}\right]\right]$.

Here is the complete calculation in tabular form:

X = Route	Pr[X]	$E_Y[Y \mid X]$	$Var_Y[Y \mid X]$	$E\left[E[Y \mid X]\right]$	$E\left[Var[Y \mid X]\right]$	$E\left[E[Y \mid X]^2\right]$
Sewickley	0.2	55	10	(0.2)(55) = 11	(0.2)(10) = 2	$(0.2)(55)^2 = 605$
Fort Pitt	0.4	50	25	(0.4)(50) = 20	(0.4)(25) = 10	$(0.4)(50)^2 = 1000$
Liberty	0.4	45	100	(0.4)(45) = 18	(0.4)(100) = 40	$(0.4)(45)^2 = 810$
				49	52	2415

$$Var\left[E[Y \mid X]\right] = E\left[E[Y \mid X]^2\right] - E[Y]^2 = 2415 - (49)^2 = 14,$$

$$E\left[Var[Y \mid X]\right] = 52,$$

and

$$Var[Y] = E\left[Var[Y \mid X]\right] + Var\left[E[Y \mid X]\right] = 52 + 14 = 66.$$

❒

Note

Using the statistical language mentioned previously, the $Var[Y \mid X]$ column can be considered as the variance within routes, with $E\left[\text{Var}[Y \mid X]\right] = 52$ being the (weighted) average variance within routes. Then, $\text{Var}\left[E[Y \mid X]\right] = 14$ is the variance between routes. In this example most of the variance is within routes (52 out of a total of 66), with an especially high variance on the Liberty route. Since the mean route times differ by only a few minutes, the variance between routes is only 14.

Example 8.4-3 Mean and Variance by Conditioning

Let X and Y be jointly distributed random variables such that X has mean two and variance three. Suppose the conditional mean of Y, given $X = x$, is x and the conditional variance of Y, given $X = x$, is x^2. Calculate the mean and the variance of Y.

Solution

We are asked to find $E[Y]$ and $Var[Y]$. We are given $E[Y \mid X] = X$, $Var[Y \mid X] = X^2$, $E_X[X] = 2$, and $\text{Var}_X[X] = 3$. From the double expectation theorem,

$$E[Y] = E_X\left[E_Y[Y|X]\right] = E_X[X] = 2.$$

$$E_X\left[\text{Var}[Y \mid X]\right] = E_X[X^2] = 7.$$

Remember that $E[X^2] = \text{Var}[X] + \left(E[X]\right)^2 = 3 + 2^2$.

$$Var_X\left[\boldsymbol{E}\left[\boldsymbol{Y}|\boldsymbol{X}\right]\right] = Var_X[\boldsymbol{X}] = 3.$$

Finally,

$$Var[Y] = E_X\left[Var_Y[Y \mid X]\right] + Var_X\left[E_Y[Y \mid X]\right] = 7 + 3 = 10.$$ ❒

Example 8.4-4 Random Parameter

Let X denote the amount of damage caused by a hurricane in millions. Suppose that X is exponentially distributed with mean Ω, where Ω, is itself a random variable that is uniformly distributed on the interval $[5, 45]$. Calculate the mean and the standard deviation of X.

Solution

X is exponentially distributed with mean Ω, so we know that $E_X[X \mid \Omega] = \Omega$ and $\text{Var}_X[X \mid \Omega] = \Omega^2$. Since Ω, is uniformly distributed on $[5, 45]$ we have $E[\Omega] = \frac{5+45}{2} = 25$ and $\text{Var}[\Omega] = \frac{(45-5)^2}{12} = \frac{400}{3}$. Thus, $E[X] = E_\Omega\left(E_X[X \mid \Omega]\right) = E_\Omega[\Omega] = 25$.

Next,

$$Var_{\Omega}\left(E[X\,|\,\Omega]\right) = Var_{\Omega}(\Omega) = \frac{400}{3}.$$

Then,

$$E_{\Omega}\left[\mathrm{Var}_X[X\,|\,\Omega]\right] = E_{\Omega}[\Omega^2] = \mathrm{Var}_{\Omega}[\Omega] + E_{\Omega}[\Omega]^2 = \frac{2275}{3}.$$

Then,

$$\mathrm{Var}[X] = E_{\Omega}\left(\mathrm{Var}_X[X\,|\,\Omega]\right) + \mathrm{Var}_{\Omega}\left(E_X[X\,|\,\Omega]\right) = \frac{2275}{3} + \frac{400}{3} = \frac{2675}{3} = 891.\overline{6}.$$

Finally, $\sigma_X = \sqrt{891.6} = 29.86.$ ❐

Another common application of the conditioning formulas involves random sums. That is, situations where the number of terms N in the sum of random variables is itself a random variable.

Random Sums

Let $X_1, X_2, \ldots, X_N$ be a random sample from a population X with mean and variance given by μ_X and σ_X^2. Let the number of terms N be a random variable independent of X, and let μ_N and σ_N^2 denote the mean and variance of N. Let

$$S = X_1 + X_2 + \cdots + X_N$$

denote the random sum. Then,

(1) $E[S] = \mu_S = \mu_N \cdot \mu_X$, and

(2) $Var[S] = \sigma_S^2 = \sigma_N^2 \cdot \mu_X^2 + \mu_N \cdot \sigma_X^2.$

Proof

If the number of terms is given as $N = n$, we know that the sum S has exactly n terms. Then,

$$E_S[S\,|\,N=n] = E[X_1 + \cdots + X_n] = E[X_1] + \cdots + E[X_n] = n \cdot \mu_X$$

and

$$Var\left[S\,|\,N=n\right] = Var[X_1 + \cdots + X_n] = Var[X_1] + \cdots + Var[X_n] = n \cdot \sigma_X^2.$$

Next, we apply the double expectation theorem to account for the random number of terms N.

$$E[S] = E_N\left[E_S[S\,|\,N]\right] = E_N[N \cdot \mu_X] = \mu_X \cdot E_N[N] = \mu_N \cdot \mu_X.$$

$$\begin{aligned} Var[S] &= Var_N\left[E_S[S\,|\,N]\right] + E_N\left[Var_S[S\,|\,N]\right] \\ &= Var_N[N \cdot \mu_X] + E_N[N \cdot \sigma_X^2] \\ &= {\mu_X}^2 \cdot Var_N[N] + \sigma_X^2 \cdot E_N[N] = \sigma_N^2 \cdot \mu_X^2 + \mu_N \sigma_X^2. \end{aligned}$$

❐

Example 8.4-5 Random Sums

An insurance portfolio has a random number of claims given by N, where N is a Poisson random variable with mean of 100. The claim amount is X with mean of 50 and variance of 101. All the claim amounts are independent and the claim amounts are also independent of the number of claims. Estimate the 95^{th} percentile of total claims.

Solution

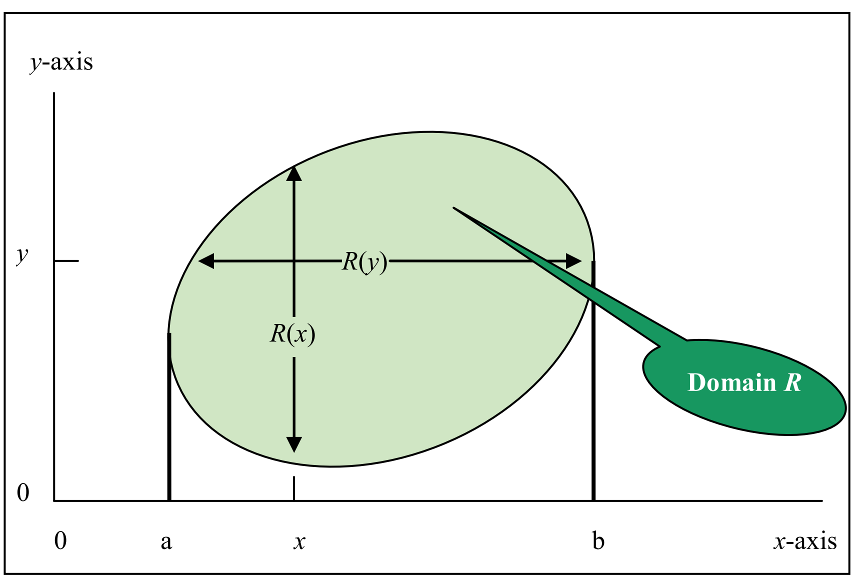
The total claims random variable is the random sum $S = X_1 + X_2 + \cdots + X_N$. Since N is a Poisson random variable, the mean and the variance of N are both equal to 100. It follows from the above discussion that $\mu_S = \mu_N \cdot \mu_X = (100)(50) = 5{,}000$ and $\sigma_S^2 = \sigma_N^2 \cdot \mu_X^2 + \mu_N \cdot \sigma_X^2$ $= (100)(50^2) + (100)(101) = 260{,}100$ Then $\sigma_S = 510$.

Let $S_{.95}$ denote the 95^{th} percentile of the sum. Making use of the Central Limit Theorem, we have,

$$.95 = \Pr\left[S \le S_{.95}\right] \approx \Pr\left[Z \le \frac{S_{.95} - \mu_S}{\sigma_S} = \frac{S_{.95} - 5000}{510} = 1.645\right].$$

Thus, $S_{.95} = \mu_S + z_{.95} \cdot \sigma_S = 5000 + (1.645)(510) = 5{,}838.95.$ ❒

Finally, we give derivations of the conditioning formulas. The exact details depend on whether the joint distribution of X and Y is discrete, continuous, or a mixture of the two. For example, in the driving route problems the route X is discrete and the drive time Y is continuous. For the proof below we will assume the joint distribution of X and Y is continuous on a domain R in the plane. This is the context of Section 7.5, and we will employ the notation and schematic diagram (reproduced below) from that section. The steps can be easily modified to relate to the other cases.



Proof of the conditioning formulas (joint continuous distribution case):

Let $f(x,y)$ be the joint density function for X and Y on the domain R. We make use of the fact that $f(y \mid X=x)=\frac{f(x,y)}{f_X(x)}$, or equivalently, $f(x,y)=f(y \mid X=x)\cdot f_X(x)$, for values of y along the vertical line segment $R(x)$. We describe the region R as consisting of the vertical line segments $R(x)$; $a \le x \le b$. Then,

$$\begin{aligned} E[Y] &= \iint_R y\cdot f(x,y)\,dydx = \int_a^b \int_{R(x)} y\cdot f(x,y)\,dydx \\ &= \int_a^b \int_{R(x)} y\cdot f(y \mid X=x)\cdot f_X(x)\,dy\,dx \\ &= \int_a^b f_X(x)\left[\int_{R(x)} y\cdot f(y \mid X=x)\,dy\right]dx \\ &= \int_a^b f_X(x)\,E_Y[Y \mid X=x]dx \\ &= E_X\left[E_Y[Y \mid X]\right]. \end{aligned}$$

This establishes the double expectation formula.

Now, applying the double expectation formula to Y^2 gives, $E[Y^2]=E_X\left[E_Y[Y^2 \mid X]\right]$.
Thus,

$$\begin{aligned} Var[Y] &= E[Y^2]-E[Y]^2 \\ &= E_X\left[E_Y[Y^2 \mid X]\right]-E_X\left[E_Y[Y \mid X]\right]^2 \\ &= E_X\left(Var_Y[Y \mid X]+[E_Y[Y \mid X]^2\right)-E_X\left[E_Y[Y \mid X]\right]^2 \\ &= E_X\left[Var_Y[Y \mid X]\right]+E_X\left[E_Y[Y \mid X]^2\right]-E_X\left[E_Y[Y \mid X]\right]^2 \\ &= E_X\left[Var_Y[Y \mid X]\right]+\left\{E_X\left[E_Y[Y \mid X]^2\right]-E_X\left[E_Y[Y \mid X]\right]^2\right\} \\ &= E_X\left[Var_Y[Y \mid X]\right]+Var_X\left[E_Y[Y \mid X]\right]. \end{aligned}$$

Exercise 8-46 Let the random variable B denote the amount of beer (in ounces) that you drink at a sporting event. This random variable depends on the sporting event that you attend, G. Consider the following table:

G	**Pr(*G*)**	**Conditional *Beer* distribution, $B \mid G$**
Cubs at Wrigley	.4	$\sim N(\mu=50, \sigma^2=16)$
Steelers at Heinz	.3	$\sim Uniform[20,120]$
Red Sox at Fenway	.2	$\sim Exponential(\mu=65)$
Curling at Olympics	.1	24 ounces w/prob=1

Compute $E[B]$ and σ_B.

Exercise 8-47 Let X denote the amount of damage caused by a tornado. Suppose that X is uniformly distributed on the interval $[0,\theta]$, where θ is itself a random variable with exponential distribution and mean 5 million.

(a) Compute the expected damage caused by the tornado.
(b) Compute the standard deviation of the amount of damage caused by the tornado.
(c) What is the probability that at least 1 million dollars of damage is caused by the tornado?

Exercise 8-48 ***Compound Poisson*:** Suppose that $X_1, X_2, \ldots$ is a random sample of independent losses. Let N denote the total (random) number of losses. Assume that N is independent of the X_i's and has Poisson distribution with mean μ_N. Show that for the total (aggregate) losses, $S = X_1 + X_2 + \cdots + X_N$ we have:

(a) $E[S] = E[X] \cdot E[N] = \mu_X \cdot \mu_N$, and
(b) $Var[S] = \mu_N \cdot E[X^2]$.

Exercise 8-49 Suppose that X is a random variable that denotes the amount of damage to your automobile. Suppose that $X \sim Uniform[5000, 20000]$. Let N denote the number of accidents that you have. Suppose that N is a Bernoulli random variable with probability of success p. That is, you have one accident with probability p and zero accidents with probability $q = 1 - p$. Find the expected value and variance for the total losses.

Exercise 8-50 SOA/CAS P Sample Exam Questions #54
An auto insurance company insures an automobile worth 15,000 for one year under a policy with a 1,000 deductible. During the policy year there is a 0.04 chance of partial damage to the car and a 0.02 chance of a total loss of the car. If there is partial damage to the car, the amount X of damage (in thousands) follows a distribution with density function

$$f(x) = \begin{cases} 0.5003\, e^{-x/2} & \text{for } 0 < x < 15 \\ 0 & \text{otherwise} \end{cases}.$$

What is the expected claim payment?

(A) 320 (B) 328 (C) 352 (D) 380 (E) 540

Exercise 8-51 Suppose that loss random variables X, Y, and Z are mutually independent and have the following probability distributions:

X	$\Pr(X)$
0	.4
1	.3
2	.3

Y	$\Pr(Y)$
0	.25
1	.35
3	.40

Z	$\Pr(Z)$
0	.10
1	.15
2	.20
4	.55

Let $S = X + Y + Z$ represent the total loss amount.

(a) Complete the adjacent table for the probability distribution for $S = X+Y+Z$. *Hint*: This problem is quite tractable and we would draw a three stage tree to determine all the probabilities.

(b) Compute $E[S]$.

(c) Compute σ_S.

$S = X+Y+Z$	$\Pr[S]$
0	.01
1	.0365
2	
3	
4	
5	
6	
7	.16975
8	.066
9	.066

Exercise 8-52 SOA/CAS Course 1 2000 Sample Examination #16

Micro Insurance Company issued insurance policies to 32 independent risks. For each policy, the probability of a claim is $\frac{1}{6}$. The benefit amount given that there is a claim has probability density function

$$f(y) = \begin{matrix} 2(1-y) & 0 < y < 1 \\ 0 & \text{otherwise} \end{matrix}.$$

Calculate the expected value of total benefits paid.

(A) $\frac{16}{9}$ (B) $\frac{8}{3}$ (C) $\frac{32}{9}$ (D) $\frac{16}{3}$ (E) $\frac{32}{3}$

Exercise 8-53 Compute the variance of the total benefits paid in Exercise 8-52.

Hint: The number of claims $N \sim Binomial\left(n = 32, p = \frac{1}{6}\right)$.

Exercise 8-54 SOA/CAS P Sample Exam Questions #120

An insurance policy is written to cover a loss *X*, where *X* has density function

$$f(x) = \begin{cases} \frac{3}{8}x^2 & \text{for } 0 \le x \le 2 \\ 0 & \text{otherwise} \end{cases}$$

The time (in hours) to process a claim of size *X*, where $0 \le x \le 2,$ is uniformly distributed on the interval from x to $2x.$ Calculate the probability that a randomly chosen claim on this policy is processed in three hours or more.

(A) 0.17 (B) 0.25 (C) 0.32 (D) 0.58 (E) 0.83

Exercise 8-55 SOA/CAS P Sample Exam Questions #139

A driver and a passenger are in a car accident. Each of them independently has probability 0.3 of being hospitalized. When hospitalization occurs, the loss is uniformly distributed on [0,1]. When two hospitalizations occur, the losses are independent. Calculate the expected number of people in the car who were hospitalized, given that the total loss due to hospitalizations from the accident is less than 1.

(A) 0.510 (B) 0.534 (C) 0.600 (D) 0.628 (E) 0.800

8.5 CHAPTER 8 SAMPLE EXAMINATION

1. Let X be the random variable whose density function is given by $f_X(x) = 3x^2;$ $0 \le x \le 1.$ Find the density function $f_Y(y)$ of $Y = e^{2X}.$

2. Let *X* be exponentially distributed with mean 5 and let *Y* be exponentially distributed with mean 3. Let $Z = Min(X,Y).$
   (a) Find $F_Z(z) = \Pr(Z \le z).$
   (b) How is $Z = Min(X,Y)$ distributed?
   (c) Find $E[Z].$
   (d) Find $Var[Z].$

3. Suppose that $X \sim N(\mu_X = 2, \sigma_X = 3)$ and $Y \sim N(\mu_Y = -1, \sigma_X = 5)$ are independent. Let $S = X + Y$ denote the sum.
   (a) Find the moment generating function for the sum, $M_S(t).$
   (b) How is the sum distributed?
   (c) Determine $\Pr(S < 4).$