Random Variables, Probability Distributions, and Expected Values

James H. Steiger

October 27, 2003

1 Goals for this Module

In this module, we will present the following topics

1. Random variables
2. Probability distribution
3. The expected value of a random variable
   (a) The discrete case
   (b) The continuous case
4. Functions of a random variable
5. The algebra of expected values
6. Variance of a random variable
7. Bivariate distributions
8. Covariance and correlation for two random variables

2 Random Variables

In many situations, it is cumbersome to deal with outcomes in their original form, so instead we assign numbers to outcomes. This allows us to deal directly with the numbers. This usually is more convenient than dealing with the outcomes themselves, because, in our culture, we have so many refined mechanisms for dealing with numbers. Perhaps the simplest example is when we are talking about the outcome of a coin toss. Instead of dealing with "Heads" and "Tails," we instead deal with a numerical coding like "1" and "0." The coding rule that uniquely assigns numbers to outcomes is called a *random variable*, and is defined as follows

Definition 2.1 *A* random variable *is a function from a sample space* Ω *into the real numbers.*

Interestingly, a random variable does not, in itself, have any randomness. It is simply a coding rule. When a probabilistic process generates an outcome, it is immediately coded into a number by the random variable coding rule. The coding rule is fixed. The randomness observed in the numbers "flows through" the outcomes into the numbers via the coding rule. An example should make this clear.

Example 2.1 ***(Summarizing the Results of Coin Tossing)*** *Suppose you toss a fair coin 2 times and observe the outcomes. You then define the random variable X to be the number of heads observed in the 2 coin tosses. This is a valid random variable, because it is a function assigning real numbers to outcomes, as follows*

Table 1: A simple random variable

Outcome (in Ω)	HH	HT	TH	TT
Value of X	2	1	1	0

Like all functions, a random variable has a *range*, which is the set of all possible values (or *realizations*) it may take on. In the above example, the range of the random variable X is $R(X) = \{0, 1, 2\}$. Notice that, although the 4 outcomes in Ω are equally likely (each having probability 1/4), the values of X are not equally likely to occur.

3 Discrete Probability Distributions

A *probability distribution* for a discrete random variable X is defined formally as follows

Definition 3.1 *The* probability distribution function P_X *for a discrete random variable X is a function assigning probabilities to the elements of its range $R(X)$.*

Remark 3.1 *If we adopt the notation that large letters (like X) are used to stand for random variables, and corresponding small letters (like x) are used to stand for realized values (i.e., elements of the range of) these random variables, we see that $P_X(x) = \Pr(X = x)$.*

Example 3.1 ***(A Simple Probability Distribution)*** *Consider the random variable* X *discussed in Table 1 in the preceding example. The probability distribution of* X *is obtained by collating the probabilities for the 3 elements in* $R(X)$*, as follows*

Table 2: Probability distribution for the random variable of Table 1

x	$P_X(x)$
2	1/4
1	1/2
0	1/4

Example 3.2 ***(Simulating a Fair Coin with a Fair Die)*** *Suppose you throw a fair die, and code the outcomes as in the table below*

Outcome (in Ω)	*1*	*2*	*3*	*4*	*5*	*6*
Value of X	*1*	*0*	*1*	*0*	*1*	*0*

The random variable X would then have the probability distribution shown in the following table

x	$P_X(x)$
1	1/2
0	1/2

4 Expected Value of a Random Variable

The *expected value*, or *mean* of a random variable X, denoted $E(X)$ (or, alternatively, μ_X), is the long run average of the values taken on by the random variable. Technically, this quantity is defined differently depending on whether a random variable is discrete or continuous. For some random variables, $E(|X|) = \infty$, and we say that the expected value *does not exist.*

4.1 The Discrete Case

Recall that, in the case of a frequency distribution where the observed variable takes on k distinct values X_i with frequencies f_i, the sample mean can be computed directly by

$$\overline{X}_{\bullet} = \frac{1}{N}\sum_{i=1}^{k} X_i f_i$$

This can also be written

$$\begin{aligned}\overline{X}_\bullet &= \frac{1}{N}\sum_{i=1}^{k} X_i f_i \\ &= \sum_{i=1}^{k} X_i \frac{f_i}{N} \\ &= \sum_{i=1}^{k} X_i r_i\end{aligned}$$

where the r_i represent relative frequencies.

Thus the average of the discrete values in a sample frequency distribution can be computed by taking the sum of cross products of the values and their relative frequencies. The expected value of a discrete random variable X is defined in an analogous manner, simply replacing relative frequencies with probabilities.

Definition 4.1 ***(Expected Value of a Discrete Random Variable)*** *The expected value of a discrete random variable X whose range $R(X)$ has k possible values x_i is*

$$\begin{aligned}E(X) &= \sum_{i=1}^{k} x_i \Pr(X = x_i) \\ &= \sum_{i=1}^{k} x_i P_X(x_i)\end{aligned}$$

An alternative notation seen in mathematical statistics texts is

$$E(X) = \sum_{x_i \in R(X)} x_i P_X(x_i)$$

Example 4.1 ***(Expected Value of a Fair Die Throw)*** *When you throw a fair die, the probability distribution for the outcomes assigns uniform probability $1/6$ to the outcomes in $R(X) = \{1, 2, 3, 4, 5, 6\}$. To expected value* *can be calculated as in the following table.*

x	$P_X(x)$	$xP_X(x)$
6	1/6	6/6
5	1/6	5/6
4	1/6	4/6
3	1/6	3/6
2	1/6	2/6
1	1/6	1/6
		$\mathbf{21/6 = 7/2}$

A fair die has an expected value of 3.5*, or* 7/2*.*

4.2 The Continuous Case

For continuous random variables, the probability of an individual outcome in $R(X)$ is not defined, and $R(X)$ is uncountably infinite. The expected value is defined as the continuous analog of the discrete case, with the probability density function $f(x)$ replacing probability, and integration replacing summation.

Definition 4.2 ***(The Expected Value of a Continuous Random Variable)*** *The expected value of a continous random variable X having probability density function $f(x)$ is*

$$E(X) = \int_{-\infty}^{\infty} x\ f(x)\ dx$$

5 Functions of a random variable

Recall that the random variable X is a function, a coding rule. Consequently, functions $g(X)$ of X will also be random variables, and have a distribution of their own. However, the range of $g(X)$ will frequently be different from that of X. Consider the following example:

Example 5.1 ***(A Function of a Random Variable)*** *Let X be a discrete uniform random variable assigning uniform probability* 1/5 *to the numbers*

-1, 0, 1, 2, 3. *Then* $Y = X^2$ *is a random variable with the following probability distribution*

y	$P_Y(y)$
9	1/5
4	1/5
1	2/5
0	1/5

Note that the probability distribution of Y is obtained by simply collating the probabilities for each value in $R(X)$ linked to a value in $R(Y)$. However, computing the expected value of Y does not, strictly speaking, require this collation effort. That is, the expected value of $Y = g(X)$ may be computed directly from the probability distribution of X, without extracting the probability distribution of Y. Formally, the expected value of $g(X)$ is defined as follows

Definition 5.1 ***(The Expected Value of a Function of a Random Variable)*** *The expected value of a function* $g(X)$ *of a random variable* X *is computed, in the discrete and continuous cases, respectively, as*

$$E\left(g\left(X\right)\right) = \sum_{x_i \in R(X)} g(x_i) P_X(x_i) \tag{1}$$

and

$$E\left(g\left(X\right)\right) = \int_{-\infty}^{\infty} g(x)\ f(x)\ dx$$

Example 5.2 ***(Expected Value of the Square of a Die Throw)*** *Consider once again the random variable* X *representing the outcomes of a throw of a fair die, with* $R(X) = \{1, 2, 3, 4, 5, 6\}$*. In the table below, we compute the expected value of the random variable* X^2*.*

x	x^2	$P_X(x)$	$x^2 P_X(x)$
6	36	1/6	36/6
5	25	1/6	25/6
4	16	1/6	16/6
3	9	1/6	9/6
2	4	1/6	4/6
1	1	1/6	1/6
			91/6

5.1 The Algebra of Expected Values

Theorem 5.1 ***(Expectation of a Linear Transform)*** *Consider linear functions $aX + b$ of a discrete random variable X. The expected value of the linear transform follows the rule*

$$E\left(aX + b\right) = aE(X) + b$$

Proof. Eschewing calculus, we will prove only the discrete case. From Equation 1, the basic rules of summation algebra, and the fact that the sum of $P_X(x_i)$ over all values of x_i is 1, we have

$$\begin{aligned} E(aX + b) &= \sum_{x_i \in R(X)} (ax_i + b)\, P_X(x_i) \\ &= \sum_{x_i \in R(X)} ax_i\ P_X(x_i) + \sum_{x_i \in R(X)} b\ P_X(x_i) \\ &= a \sum_{x_i \in R(X)} x_i\ P_X(x_i) + b \sum_{x_i \in R(X)} P_X(x_i) \\ &= aE(X) + b(1) \\ &= aE(X) + b \end{aligned}$$

■

The result of Theorem 5.1 is precisely analogous to the earlier result we established for lists of numbers, and summarized in the "Vulnerability Box." That is, multiplicative constants and additive constants come straight through in the expected value, or mean, of a random variable. This result includes several other results as special cases, and these derivative rules are sometimes called "The Algebra of Expected Values."

Corollary 5.1 ***The Algebra of Expected Values*** *For any random variable X, and constants a and b, the following results hold*

$$E(a) = a \tag{2}$$

$$E(aX) = aE(X) \tag{3}$$

$$E(X + Y) = E(X) + E(Y) \tag{4}$$

Note that these results are analogous to the two constant rules and the distributive rule of summation algebra.

5.2 Expected Value of a Linear Combination

The expected value of a linear combination of random variables behaves the same as the mean of a linear combination of lists of numbers.

Proposition 5.1 ***(Mean of a Linear Combination of Random Variables)*** *Given J random variables X_j, $j = 1, \ldots J$, with expected values μ_j. The linear combination $\kappa = \sum_{j=1}^{J} c_j X_j$ has expected value given by*

$$\mu_\kappa = E(\kappa) = \sum_{j=1}^{J} c_j \mu_j$$

6 Variance of a Random Variable

Deviation scores for a random variable X are defined via the deviation score random variable $dX = X - E(X)$.A random variable is said to be *in deviation score form* if it has an expected value of zero. The *variance of a random variable X* is the expected value of its squared deviation scores. Formally, we say

Definition 6.1 ***(Variance of a Random Variable)*** *The variance of a random variable X is defined as*

$$Var(X) = \sigma_X^2 = E(X - E(X))^2 = E\left(X - \mu_X\right)^2 = E\left(dX^2\right)$$

Just as we usually prefer a *computational formula* when computing a sample variance, we also often prefer the following alternative formula when computing the variance of a random variable. This formula can be proven easily using the algebra of expected values.

Proposition 6.1 ***(Variance of a Random Variable)*** *The variance of a random variable X is equal to*

$$Var(X) = \sigma_X^2 = E\left(X^2\right) - \left(E\left(X\right)\right)^2 \tag{5}$$

In the following example, we demonstrate both methods for computing the variance of a discrete random variable.

Example 6.1 ***(The Variance of a Fair Die Throw)*** *Consider again the random variable X representing the 6 outcomes of a fair die throw. We have already established in Example 4.1 that $E(X) = 7/2$ and in Example 5.2 that $E(X^2) = 91/6$.Employing Equation 5, we have*

$$\begin{aligned} Var(X) &= E\left(X^2\right) - \left(E\left(X\right)\right)^2 \\ &= 91/6 - (7/2)^2 \\ &= 364/24 - 294/24 \\ &= 70/24 \\ &= 35/12 \end{aligned}$$

Alternatively, we may calculate the variance directly

x	$P_X(x)$	$x - E(X)$	$(x - E(X))^2$	$(x - E(X))^2 P_X(x)$
6	1/6	5/2	25/4	25/24
5	1/6	3/2	9/4	9/24
4	1/6	1/2	1/4	1/24
3	1/6	−1/2	1/4	1/24
2	1/6	−3/2	9/4	9/24
1	1/6	−5/2	25/4	25/24
				70/24 = 35/12

The sum of the numbers in the far right column is σ_X^2.

6.1 Z-Score Random Variables

Random variables are said to be *in Z-score form* if and only if they have an expected value (mean) of zero and a variance of 1. A random variable may be converted into Z-score form by subtracting its mean then dividing by its standard deviation, i.e.,

$$Z_X = \frac{X - E(X)}{\sqrt{Var\left(X\right)}} = \frac{X - \mu_X}{\sigma_X}$$

7 Random Vectors and Multivariate Probability Distributions

The *random vector* is a generalization of the concept of a random variable. Whereas a random variable codes outcomes as single numbers by assigning a unique number to each outcome, a random vector assigns a unique ordered list of numbers to each outcome. Formally, we say

Definition 7.1 ***(Random Vector)*** *An n-dimensional random vector is a function from a sample space Ω into $\Re^n$, n-dimensional Euclidean space.*

The bivariate random vector (X, Y) assigns pairs of numbers to outcomes. For many practical purposes, we can simply consider such a random vector as the outcome of observations on two variables. Earlier we discussed the *bivariate frequency distribution of X and Y*, which assigned probability to ordered pairs of observations on X and Y. There is a similar notion for discrete random variables. We use the notation $P_{X,Y}(x, y)$ to indicate $\Pr(X = x \ \cap \ Y = y)$. The *range* $R(X, Y)$ of the random vector is the set of all possible realizations of (X, Y) Terms that we defined previously, the *conditional distribution*, and the *marginal distribution*, are also relevant here.

Example 7.1 ***(A Bivariate Probability Distribution)*** *Suppose you toss 2 coins, X and Y, and code the outcomes $1 = H$, $2 = T$. Suppose the 4 possible outcomes on the random vector (X, Y) and their probabilities are shown in the bivariate probability distribution table below. (Note that the coins behave in a peculiar manner. They are perfectly correlated!)*

Table 3: A Bivariate Discrete Probability Distribution

(x, y)	$P_{X,Y}(x, y)$
$(1, 1)$	$1/2$
$(1, 2)$	0
$(2, 1)$	0
$(2, 2)$	$1/2$

7.1 Functions of a Random Vector

You can define new random variables that are functions of the random variables in a random vector. For example, we can define the random variable XY, the product of X and Y. This new random variable will have a probability distribution that can be obtained from the bivariate distribution by collating and summing probabilities. For example, to obtain the probability that $XY = 1$, one must sum the probabilities for all realizations (x, y) where $xy = 1$.

Example 7.2 ***(The Product of Two Random Variables)*** *Consider again the bivariate distribution of X and Y given in Table 3. The distribution of the random variable $W = XY$ is given in the table below. Note in passing that $E(XY) = 5/2$*

w	$P_W(w)$
4	1/2
1	1/2

7.2 Marginal Distributions

We may extract marginal distributions from the multivariate distribution. For example, the marginal distributions $P_X(x)$ and $P_Y(y)$ of X and Y can be extracted from $P_{X,Y}(x, y)$ by summation. For example, to calculate $P_X(1) = \Pr(X = 1)$, we simply observe all realizations (x, y) where $x = 1$, and sum their probabilities. Formally, we say, for a particular value x^*,

$$P_X(x^*) = \sum_{(x^*,y)\in R(X,Y)} P_{X,Y}(x^*, y)$$

If we do this for all possible values that X may take on, we obtain the marginal distribution. For example, the marginal distribution of X for the bivariate distribution in Table 3 is

Table 4: Marginal Distribution of X

x	$P_X(x)$
2	1/2
1	1/2

7.3 Conditional Distributions

The *conditional distribution* of Y given $X = a$ is denoted $P_{Y|X=a}(y)$. It is computed by collating only those observations for which $X = a$, and restandardizing the probabilitities so that they add to 1. Thus, we have

$$P_{Y|X=a}(y) = \frac{\Pr(Y = y \cap X = a)}{\Pr(X = a)} = \frac{P_{X,Y}(a, y)}{P_X(a)}$$

7.4 Independence of Random Variables

Solution 7.1 *Two random variables are* independent *if their conditional and marginal distributions are the same, so that knowledge of the status of one variable does not change the probability structure for the other.*

Definition 7.2 ***(Independence of Random Variables)*** *Two random variables X and Y are independent if, for all realized values of X,*

$$P_{Y|X=a}(y) = P_Y(y)$$

or, equivalently

$$P_{X,Y}(x,y) = P_X\,(x)\,P_Y\,(y)$$

Remark 7.1 *If X and Y are independent, then*

$$E(XY) = E(X)E(Y)$$

However it is not the case that if X and Y are independent, $E(X/Y) = E(X)/E(Y)$. This incorrect supposition is at the heart of a number of erroneous results from some surprisingly authoritative source. For example, in the analysis of variance, we calculate an F statistic as

$$F = \frac{MS_{\text{between}}}{MS_{\text{within}}}$$

Although MS_{between} and MS_{within} are independent, it is not the case that $E(F) = E(MS_{\text{between}})/E(MS_{\text{within}})$, despite the fact that several textbooks on analysis of variance (including the classic by Winer) state this.

8 Covariance and Correlation for Two Random Variables

The *covariance* of two random variables X and Y is defined as the average cross-product of deviation scores. Specifically,

Definition 8.1 ***(Covariance of Two Random Variables)*** *The covariance between two random variables X and Y is defined as*

$$Cov(X,Y) = \sigma_{XY} = E[\ (X - E\,(X))\ (\ Y - E\,(Y))\ ]$$

The covariance may also be computed as

$$Cov(X,Y) = \sigma_{XY} = E(XY) - E(X)E(Y)$$

The correlation between two random variables X *and* Y *is*

$$Corr(X,Y) = \rho_{X,Y} = E(Z_X Z_Y) = \frac{\sigma_{X,Y}}{\sigma_X \sigma_Y}$$

Example 8.1 ***(Covariance and Correlation of Two Random Variables)*** *Consider again the two strange coins in Table 3.* X *and* Y *each has the same marginal distribution, taking on the values* 1 *and* 2 *with probability* 1/2*. So* $E(X) = E(Y) = 3/2$*. We also have* $E(X^2) = E(Y^2) = 5/2$*, so*

$$\sigma_X^2 = E\left(X^2\right) - [E(X)]^2 = 5/2 - (3/2)^2 = 10/4 - 9/4 = 1/4 = \sigma_Y^2$$

and

$$\sigma_X = \sigma_Y = 1/2$$

To calculate the covariance, we need $E(XY)$*,which, as we noted in Example 7.2, is* 5/2*. The covariance may be calculated as*

$$Cov(X,Y) = \sigma_{X,Y} = E(XY) - E(X)E(Y) = 5/2 - (3/2)(3/2) = 1/4$$

The correlation between X *and* Y*, not surprisingly, turns out to be* 1*, i.e.,*

$$Corr(X,Y) = \frac{\sigma_{X,Y}}{\sigma_X \sigma_Y} = \frac{1/4}{(1/2)(1/2)} = 1$$

8.1 Variance and Covariance of Linear Combinations

The variance and covariance of linear combination(s) of random variables show the same behavior as a the corresponding quantities for lists of numbers. Simply apply the same heuristic rules as before.

Example 8.2 *Suppose* $A = X + Y$ *and* $B = X - Y$*. Then*

$$\sigma_A^2 = \sigma_X^2 + \sigma_Y^2 + 2\sigma_{X,Y}$$

$$\sigma_B^2 = \sigma_X^2 + \sigma_Y^2 - 2\sigma_{X,Y}$$

and

$$\sigma_{A,B} = \sigma_X^2 - \sigma_Y^2$$