COLUMBIA
BUSINESS
SCHOOL

Random Variables, Distributions, and Expected Value

Fall 2001 — Professor Paul Glasserman
B6014: Managerial Statistics — 403 Uris Hall

The Idea of a Random Variable

1. A **random variable** is a variable that takes specific values with specific probabilities. It can be thought of as a variable whose value depends on the outcome of an uncertain event.

2. We usually denote random variables by capital letters near the end of the alphabet; e.g., X, Y, Z.

3. Example: Let X be the outcome of the roll of a die. Then X is a random variable. Its possible values are 1, 2, 3, 4, 5, and 6; each of these possible values has probability 1/6.

4. The word "random" in the term "random variable" does *not* necessarily imply that the outcome is completely random in the sense that all values are equally likely. Some values may be more likely than others; "random" simply means that the value is uncertain.

5. When you think of a random variable, immediately ask yourself
   - What are the possible values?
   - What are their probabilities?

6. Example: Let Y be the *sum* of two dice rolls.
   - Possible values: $\{2, 3, 4, \ldots, 12\}$.
   - Their probabilities: 2 has probability 1/36, 3 has probability 2/36, 4 has probability 3/36, etc. (The important point here is not the probabilities themselves, but rather the fact that such a probability can be assigned to each possible value.)

7. The probabilities assigned to the possible values of a random variable are its **distribution**. A distribution completely describes a random variable.

8. A random variable is called **discrete** if it has countably many possible values; otherwise, it is called **continuous.** For example, if the possible values are any of these:

   - $\{1, 2, 3, \ldots, \}$
   - $\{\ldots, -2, -1, 0, 1, 2, \ldots\}$
   - $\{0, 2, 4, 6, \ldots\}$
   - $\{0, 0.5, 1.0, 1.5, 2.0, \ldots\}$
   - any finite set

   then the random variable is discrete. If the possible values are any of these:

   - all numbers between 0 and ∞
   - all numbers between $-\infty$ and ∞
   - all numbers between 0 and 1

   then the random variable is continuous. Sometimes, we approximate a discrete random variable with a continuous one if the possible values are very close together; e.g., stock prices are often treated as continuous random variables.

9. The following quantities would typically be modeled as discrete random variables:

   - The number of defects in a batch of 20 items.
   - The number of people preferring one brand over another in a market research study.
   - The credit rating of a debt issue at some date in the future.

   The following would typically be modeled as continuous random variables:

   - The yield on a 10-year Treasury bond three years from today.
   - The proportion of defects in a batch of 10,000 items.
   - The time between breakdowns of a machine.

Discrete Distributions

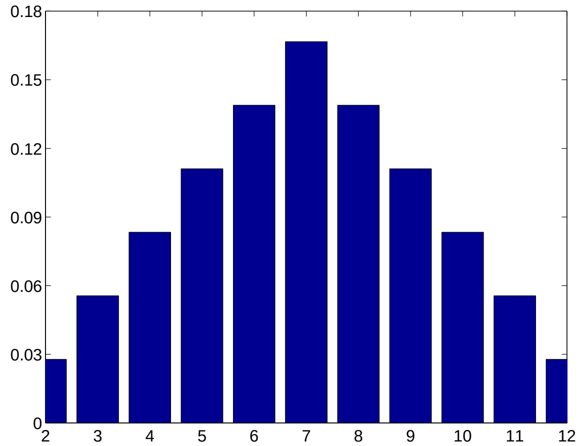
1. The rule that assigns specific probabilities to specific values for a discrete random variable is called its **probability mass function** or **pmf**. If X is a discrete random variable then we denote its pmf by P_X. For any value x, $P(X = x)$ is the probability of the event that $X = x$; i.e.,

   $$P(X = x) = \text{probability that the value of } X \text{ is } x.$$

2. Example: If X is the outcome of the roll of a die, then

   $$P(X = 1) = P(X = 2) = \cdots = P(X = 6) = 1/6,$$

   and $P(X = x) = 0$ for all other values of x.



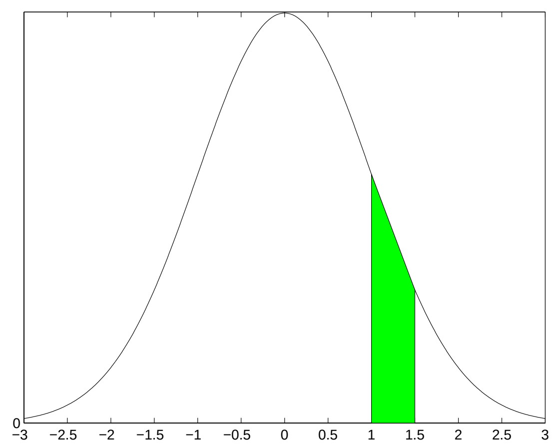


Figure 1: Left panel shows the probability mass function for the sum of two dice; the possible values are 2 through 12 and the heights of the bars give their probabilities. The bar heights sum to 1. Right panel shows a probability density for a continuous random variable. The probability $P(1 < X \leq 1.5)$ is given by the shaded area under the curve betwee 1 and 1.5. The total area under the curve is 1. The probability of any particular value, e.g., $P(X = 1)$ is zero because there is no area under a single point.

3. NOTE: We always use capital letters for random variables. Lower-case letters like x and y stand for possible values (i.e., numbers) and are not random.

4. A pmf is graphed by drawing a vertical line of height $P(X = x)$ at each possible value x. It is similar to a histogram, except that the height of the line (or bar) gives the *theoretical probability* rather than the *observed frequency*.

Continuous Distributions

1. The distribution of a continuous random variable cannot be specified through a probability mass function because if X is continuous, then $P(X = x) = 0$ for all x; i.e., the probability of any particular value is zero. Instead, we must look at probabilities of *ranges* of values.

2. The probabilities of ranges of values of a continuous random variable are determined by a **density** function. The density of X is denoted by f_X. The area under a density is always 1. The probability that X falls between two points a and b is the area under f_X between the points a and b. The familiar bell-shaped curve is an example of a density.

3. The **cumulative distribution function** or **cdf** gives the probability that a random variable X takes values less than or equal to a given value x. Specifically, the cdf of X, denoted by F_X, is given by
$$F_X(x) = P(X \leq x).$$
So, $F_X(x)$ is the area under the density f_X to the left of x.

4. For a continuous random variable, $P(X = x) = 0$; consequently, $P(X \le x) = P(X < x)$. For a discrete random variable, the two probabilities are not in general equal.

5. The probability that X falls between two points a and b is given by the difference between the cdf values at these points:

$$P(a < X \le b) = F_X(b) - F_X(a).$$

   Since $F_X(b)$ is the area under f_X to the left of b and since $F_X(a)$ is the area under f_X to the left of a, their difference is the area under f_X between the two points.

Expectations of Random Variables

1. The **expected value** of a random variable is denoted by $E[X]$. The expected value can be thought of as the "average" value attained by the random variable; in fact, the expected value of a random variable is also called its **mean**, in which case we use the notation μ_X. (μ is the Greek letter mu.)

2. The formula for the expected value of a discrete random variable is this:

$$E[X] = \sum_{\text{all possible } x} xP(X = x).$$

   In words, the expected value is the sum, over all possible values x, of x times its probability $P(X = x)$.

3. Example: The expected value of the roll of a die is

$$1(\frac{1}{6}) + 2(\frac{1}{6}) + 3(\frac{1}{6}) + \cdots + 6(\frac{1}{6}) = 21/6 = 3.5.$$

   Notice that the expected value is not one of the possible outcomes: you can't roll a 3.5. However, if you average the outcomes of a large number of rolls, the result approaches 3.5.

4. We also define the expected value for a function of a random variable. If g is a function (for example, $g(x) = x^2$), then the expected value of $g(X)$ is

$$E[g(X)] = \sum_{\text{all possible } x} g(x)P(X = x).$$

   For example,

$$E[X^2] = \sum_{\text{all possible } x} x^2P(X = x).$$

   In general, $E[g(X)]$ is not the same as $g(E[X])$. In particular, $E[X^2]$ is not the same as $(E[X])^2$.

5. The expected value of a continuous random variable cannot be expressed as a sum; instead it is an integral involving the density. (If you don't know what that means, don't worry; we won't be calculating any integrals.)

6. The **variance** of a random variable X is denoted by either $Var[X]$ or σ_X^2. (σ is the Greek letter sigma.) The variance is defined by

$$\sigma_X^2 = E[(X - \mu_X)^2];$$

this is the expected value of the squared difference between X and its mean. For a discrete distribution, we can write the variance as

$$\sigma_X^2 = \sum_x (x - \mu_X)^2 P(X = x).$$

7. An alternative expression for the variance (valid for both discrete and continuous random variables) is

$$\sigma_X^2 = E[(X^2)] - (\mu_X)^2.$$

This is the difference between the expected value of X^2 and the square of the mean of X.

8. The **standard deviation** of a random variable is the square-root of its variance and is denoted by σ_X. Generally speaking, the greater the standard deviation, the more spread-out the possible values of the random variable.

9. In fact, there is a Chebyshev rule for random variables: if $m > 1$, then the probability that X falls within m standard deviations of its mean is at least $1 - (1/m^2)$; that is,

$$P(\mu_x - m\sigma_X \leq X \leq \mu_X + m\sigma_X) \geq 1 - (1/m^2).$$

10. Find the variance and standard deviation for the roll of one die. Solution: We use the formula $Var[X] = E[X^2] - (E[X])^2$. We found previously that $E[X] = 3.5$, so now we need to find $E[X^2]$. This is given by

$$E[X^2] = \sum_{x=1}^{6} x^2 P_X(x) = 1^2(\frac{1}{6}) + 2^2(\frac{1}{6}) + \cdots + 6^2(\frac{1}{6}) = 15.167.$$

Thus,

$$\sigma_X^2 = Var[X] = E[X^2] - (E[X])^2 = 15.167 - (3.5)^2 = 2.917$$

and $\sigma = \sqrt{2.917} = 1.708$.

Linear Transformations of Random Variables

1. If X is a random variable and if a and b are any constants, then $a + bX$ is a **linear transformation** of X. It scales X by b and shifts it by a. A linear transformation of X is another random variable; we often denote it by Z.

2. Example: Suppose you have investments in Japan. The value of your investment (in yen) one month from today is a random variable X. Suppose you can convert yen to dollars at the rate of b dollars per yen after paying a commission of a dollars. What is the value of your invenstment, in dollars, one month from today? Answer: $-a + bX$.

3. Example: Your salary is a dollars per year. You earn a bonus of b dollars for every dollar of sales you bring in. If X is what you sell, how much do you make?

4. Example: It takes you exactly 16 minutes to walk to the train station. The train ride takes X hours, where X is a random variable. How long is your trip, in minutes?

5. If $Z = a + bX$, then

$$E[Z] = E[a + bX] = a + bE[X] = a + b\mu_X$$

and

$$\sigma_Z^2 = Var[a + bX] = b^2\sigma_X^2.$$

6. Thus, the expected value of a linear transformation of X is just the linear transformation of the expected value of X. Previously, we said that $E[g(X)]$ and $g(E[X])$ are generally different. The *only* case in which they are the same is when g is a linear transformation: $g(x) = a + bx$.

7. Notice that the variance of $a + bX$ does not depend on a. This is appropriate: the variance is a measure of spread; adding a does not change the spread, it merely shifts the distribution to the left or to the right.

Jointly Distributed Random Variables

1. So far, we have only considered individual random variables. Now we turn to properties of several random variables considered at the same time. The outcomes of these different random variables may be related.

2. Examples

   (a) Think of the price of each stock in the New York exchange as a random variable; the movements of these variables are related.

   (b) You may be interested in the probability that a randomly selected shopper buys prepared frozen meals. In designing a promotional campaign you might be even more interested in the probability that that same shopper also buys instant coffee and reads a certain magazine.

   (c) The number of defects produced by a machine in an hour is a random variable. The number of hours the machine operator has gone without a break is another random variable. You might well be interested in probabilities involving these two random variables together.

3. The probabilities associated with multiple random variables are determined by their **joint distribution.** As with individual random variables, we distinguish discrete and continuous cases.

4. In the discrete case, the distribution is determined by a **joint probability mass function**. For example, if X and Y are random variables, there joint pmf is

$$\begin{aligned} P_{X,Y}(x,y) &= P(X=x, Y=y) \\ &= \text{probability that } X=x \text{ and } Y=y. \end{aligned}$$

   For several random variables $X_1, \ldots, X_n$, we denote the joint pmf by $P_{X_1,\ldots,X_n}$.

5. It is often convenient to represent a joint pmf through a table. For example, consider a department with a high rate of turnover among employees. Suppose all employees are found to leave within 2-4 years and that all employees hired into this department have 1-3 years of previous work experience. The following table summarizes the joint probabilities of work experience (columns) and years stayed (rows):

	1	2	3
2	.03	.05	.22
3	.05	.06	.15
4	.14	.15	.15

   Thus, the proportion of employees that had 1 year prior experience and stayed for 2 years is 0.03. If we let Y = years stayed and X = years experience, we can express this as

$$P_{X,Y}(1,2) = P(X=1, Y=2) = 0.03.$$

   The table above determines all values of $P_{X,Y}(x,y)$.

6. What proportion of employees stay 4 years? What proportion are hired with just 1 year of experience? These are questions about **marginal probabilities**; i.e., probabilities involving just one of the random variables. A marginal probability for one random variable is found by adding up over all values of the other random variable; e.g.,

$$P(X=x) = \sum_y P(X=x, Y=y),$$

   where the sum ranges over all possible y values. In the table, the marginal probabilities correspond to the column-sums and row-sums. So, the answers to the two questions just posed are 0.44 and 0.22 (the last row-sum and the first column-sum).

7. From a joint distribution we also obtain **conditional distributions**. The conditional distribution of X given $Y=y$ is

$$P_{X|Y}(x|y) = P(X=x|Y=y) = \frac{P(X=x, Y=y)}{P(Y=y)}.$$

   To find a conditional distribution from a table, divide the corresponding row or column by the row-sum or column-sum.

8. Example: What is the distribution of years stayed among employees with 1 year of experience? Since we are conditioning on 1 year of experience, we only need to consider the first column. Its sum is 0.22. The conditional probabilities are the entries of that column divided by 0.22.

$$P_{Y|X}(2|1) = 3/22,\ P_{Y|X}(3|1) = 5/22,\ P_{Y|X}(4|1) = 14/22.$$

Notice that these conditional probabilities sum to one (as they should), though the original column entries do not. Find the conditional distribution of prior experience among employees that stayed 4 years.

9. A joint distribution determines marginal distributions but the marginal distributions do not determine the joint distribution! (The row-sums and column-sums do not determine the table entries.)

10. Two discrete random variables X and Y are **independent** if their joint distribution is the product of their marginal distributions: $P(X = x, Y = y) = P(X = x)P(Y = y)$ for all x, y. Another way to express this is to say that $P(X = x|Y = y) = P(X = x)$ for all x and y.

Covariance and Correlation

1. According to the table above, do employees hired with more years of experience tend to stay more years? This type of relationship between random variables is measured by **covariance** and **correlation**. The covariance between two random variables is

$$Cov[X, Y] = E[(X - \mu_X)(Y - \mu_Y)] = E[XY] - \mu_X \mu_Y.$$

If X tends to be large when Y is large, the covariance will be positive.

2. If two random variables are independent, their covariance is zero. However, the opposite is not (quite) true: two random variables can have zero covariance without being independent.

3. The **correlation coefficient** of X and Y is

$$\rho_{XY} = Corr[X, Y] = \frac{Cov[X, Y]}{\sigma_X \sigma_Y},$$

the ratio of the covariance to the product of the standard deviations. (ρ is the Greek letter rho.)

4. The correlation coefficient has the following properties:

   - It is always between -1 and 1.
   - A positive ρ_{XY} implies that X tends to be large when Y is large and vice-versa. A negative ρ_{XY} implies that X tends to be large when Y is small and vice-versa.

- Correlation measures the strength of **linear dependence** between two random variables. If $Y = a + bX$ and $b \neq 0$, then $|\rho_{XY}| = 1$; its sign positive or negative if b is positive or negative. Conversely, if $|\rho_{XY}| = 1$ then $Y = a + bX$ for some a, b.
- Independent random variables have zero correlation.

5. If $Y = X^2$, then the value of X completely determines the value of Y; however, the correlation is not 1 because the relationship is not linear.

6. Find the covariance and correlation between years of experience and years stayed in the table above.

7. For any random variables X and Y, we have

$$E[X + Y] = E[X] + E[Y],$$

regardless of whether or not X and Y are independent. More generally,

$$E[X_1 + X_2 + \cdots + X_n] = E[X_1] + E[X_2] + \cdots + E[X_n].$$

The variance is a bit more complicated:

$$Var[X + Y] = Var[X] + Var[Y] + 2Cov[X, Y].$$

More generally,

$$Var[aX + bY] = a^2 Var[X] + b^2 Var[Y] + 2abCov[X, Y].$$

In particular (with $a = 1$ and $b = -1$)

$$Var[X - Y] = Var[X] + Var[Y] - 2Cov[X, Y].$$

If X, Y are independent, then their covariance is zero and

$$Var[X + Y] = Var[X] + Var[Y].$$

For more than two random variables, we have

$$\begin{aligned} Var[X_1 + \cdots + X_n] &= Var[X_1] + \cdots + Var[X_n] \\ &\quad + 2Cov[X_1, X_2] + \ldots + 2Cov[X_1, X_n] \\ &\quad + \cdots + 2Cov[X_{n-1}, X_n]; \end{aligned}$$

there is a covariance term for each pair of variables. If the variables are independent, then this simplifies to

$$Var[X_1 + \cdots + X_n] = Var[X_1] + \cdots + Var[X_n].$$

If, in addition, $X_1, \ldots, X_n$ all have variance σ^2, then

$$Var[X_1 + \cdots + X_n] = (\sigma^2 + \cdots + \sigma^2) = n\sigma^2$$

and thus

$$\text{StdDev}[X_1 + \cdots + X_n] = \sqrt{n}\sigma.$$