Random Variables

Chapter 2

Roulette and Random Variables

A Roulette wheel has 38 pockets. 18 of them are red and 18 are black; these are numbered from 1 to 36. The two remaining pockets are green and are numbered 0 and 00. The wheel is spun and people bet on where a small ball placed in the wheel will land.

We can conduct an experiment in which we bet on a number once in Roulette with possible outcomes ‘we win’ or ‘we lose’.

Roulette and Random Variables

Most people would probably be less interested in *whether* they win or lose than in *how much* they win or lose.

We might also like to know how much we would expect to win if the experiment is repeated 5 or 6 times and how much our winnings might vary over multiple repetitions of the experiment.

Roulette and Random Variables

The questions

1. How much can we win or lose on a particular play?
2. How much would we expect to win if we played 5 times?
3. How much will the amount we win vary?

and others can be addressed mathematically if we convert the experimental outcomes ‘win’ and ‘lose’ into numbers.

Roulette and Random Variables

A **random variable** assigns a numerical value to each outcome in an experiment.

For instance, we could define a random variable X=‘the number on a pocket that the ball lands on in the roulette wheel’. (We’d let 0 and 00 both be counted as 0). This isn’t very interesting, but we could do it.

A better ideas would be to define a random variable X = ‘the amount we win (or lose) in one $1 bet on a number in Roulette’.

If we lose, X=-1 (remember we bet $1) and if we win, X=35.

Roulette and Random Variables

Sometimes the experimental outcomes map naturally to a set of numbers. If we conduct an experiment to measure the height of a person in this class, the possible values for the random variable X = ‘height of individual’ will all be between about 60 and 80 inches.

At other times, the mapping is less natural, but still useful. For instance, we may conduct an experiment to determine how many members of our class watch “Grey’s Anatomy”. The possible outcomes for each person are ‘watches’ and ‘does not watch’. In this case, we can assign a 1 to the ‘yeses’ and a 0 to the ‘nos’.

Roulette and Random Variables

Define a random variable for each of the following experimental outcomes – what are the values the random variable can take on?

- Number of children
- Area of yard
- Favorite grocery store
- Temperature
- Sex (M, F)

There are two types of random variables that are of particular interest to us. What is a natural way to classify random variables?

Roulette and Random Variables

The random variable associated with the roulette experiment is called **discrete** random variable because *the number of outcomes it can assume can be counted*.

Some random variables correspond to outcomes that cannot be counted (e.g. things that are measured) – these are called **continuous** random variables.

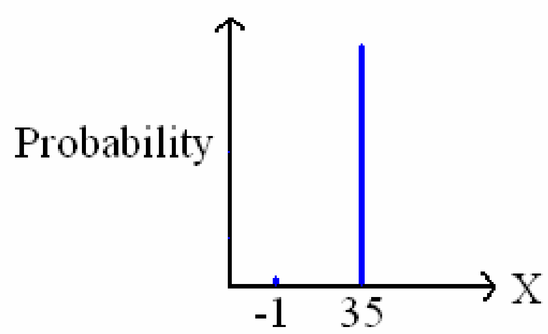
- Is the random variable that represents earnings in roulette discrete or continuous?
  - What is the probability that X=-1 (the probability that we lose)?
  - What is the probability that X=35 (the probability that we win)?

The Distribution of A Discrete R.V.

A random variable is usually denoted by an uppercase letter from near the end of the alphabet (e.g. X, Y, Z), while the specific values the random variable takes on are denoted by the corresponding lowercase letters.

The values of a discrete random variable along with their associated probabilities constitute a '**probability mass function**' (pmf). These functions are usually represented in tables or graphs as shown below.

x	P(X=x)
-1	37/38
35	1/38



The Distribution of A Discrete R.V.

In a probability mass function, we denote the probability associated with the *ith* outcome x_i as p_i, that is $P(X = x_i) = p_i$. These values must satisfy $0 \leq p_i \leq 1$ and $\Sigma_i p_i = 1$.

Verify that the probability mass function for the amount we win in roulette is a legitimate pmf, i.e. it satisfies the above conditions.

A probability mass function is sometimes referred to as the **distribution** of a discrete random variable.

The Distribution of A Discrete R.V.

Example: Suppose we randomly select a USU undergraduate student. Let X="*year in college*", then P(X=1)=0.17, P(X=2)=0.21 , P(X=3)=0.22.

- What is P(X=4)? P(X=5)?.

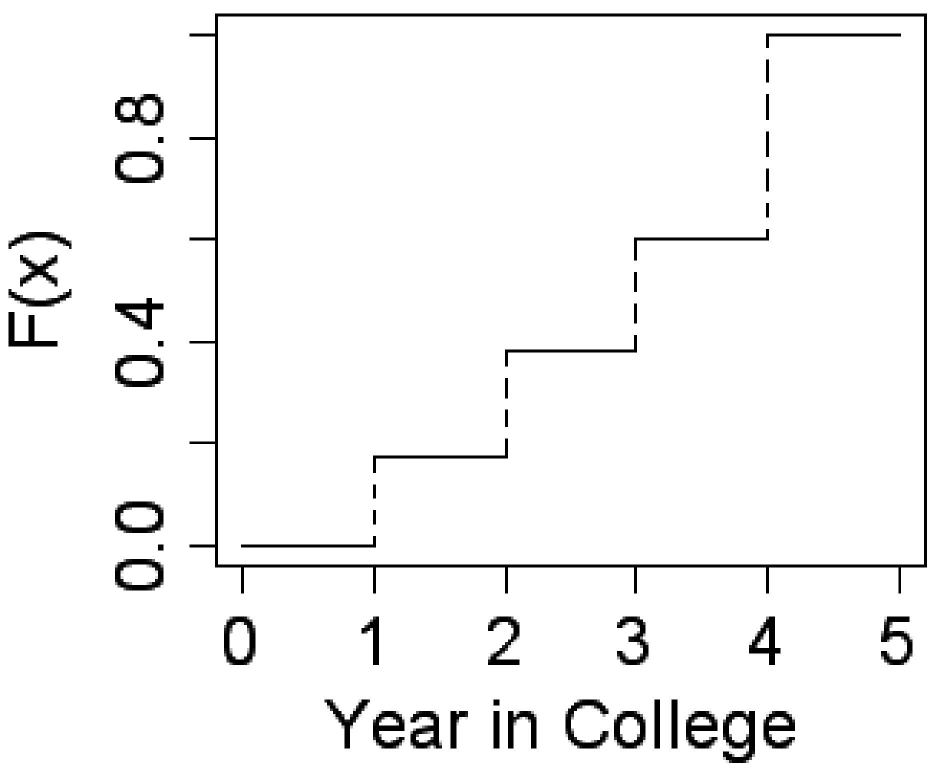
- Construct the pmf for X=*year in college* in tabular and graph forms.

Suppose we randomly select a USU student and want to know the probability he/she is a freshman or a sophomore. If X=*year in college*, we can express the probability as $P(X \le 2)$.

A function of the form $F(x)= P(X \le x)$, is called the **cumulative distribution function** (cdf) of the random variable X.

The Distribution of A Discrete R.V.

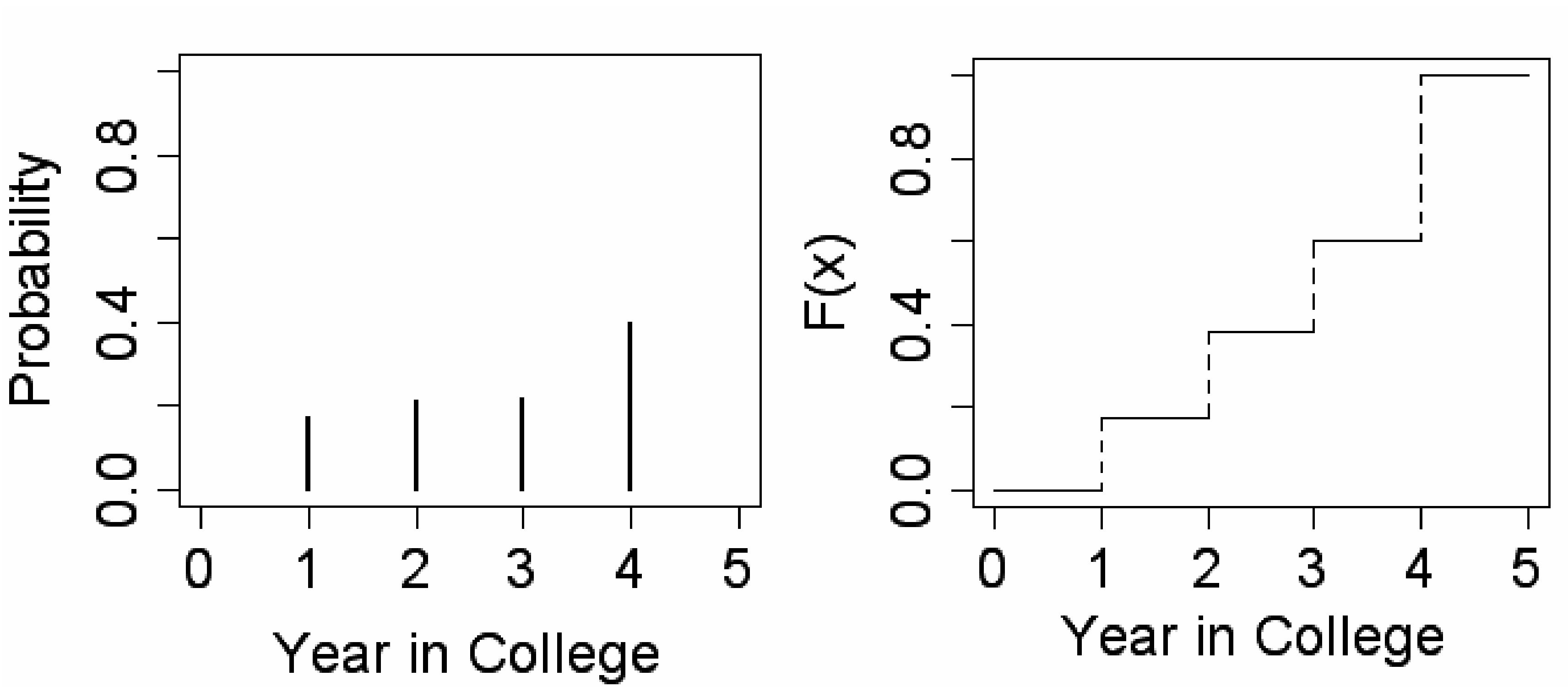
The cdf for X=*year in college*.



x	F(x)
$-\infty < x < 1$	0
$1 \le x < 2$	.17
$2 \le x < 3$	.38
$3 \le x < 4$	.6
$4 \le x < \infty$	1

The Distribution of A Discrete R.V.

The pmf and cdf for X=*year in college*.



The Distribution of A Discrete R.V.

- Find the CDF (in tabular and graphical forms) for X = 'the amount we win (or lose) in one $1 bet on a number in Roulette'.

- A game is played in which a fair die is tossed. If the roll is even, the player earns a dollar amount corresponding to the die roll, otherwise, he pays that amount. Calculate the probability mass function and the cumulative distribution function of the player's net earnings.

Expectation and Variance of a Discrete R.V.

The **expected value** (also expectation or mean), μ, of a random variable can be thought of as a weighted average and is defined as follows:

- For a discrete random variable, with pmf $P(X=x_i)=p_i$, the expected value $E(X)=\Sigma_i p_i x_i$.
- The expected value for X = 'Net winnings in one $1 bet on a number in Roulette' is

$$-1(37/38)+35(1/38) = -1/19 \approx -0.05.$$

This means that if we bet $1 on a number in roulette many times, we can expect to lose about $0.05 per play.

Expectation and Variance of a Discrete R.V.

Obviously, we won't "win" -\$0.05 on every play, rather we will win -\$1 or \$35, the variance tells us about the spread of the values of the random variable.

The **variance** (denoted Var(X) or σ^2) of a random variable X measures the variability in the values taken on by the random variable. It is defined to be $Var(X)=E((X-E(X))^2)$. This can also be expressed as $Var(X)=E(X^2)-(E(X))^2$.

The **standard deviation** of a random variable X is the positive square root of the variance and is therefore denoted by σ.

Example 1: Find the variance of X = 'Net winnings in one \$1 bet on a number in Roulette'

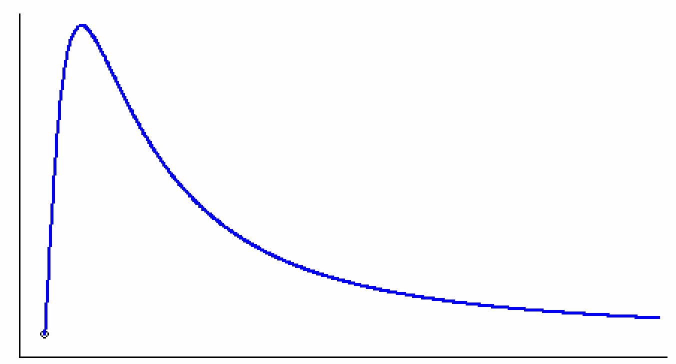
Example 2: Find the expected value and standard deviation of X=*year in college*.

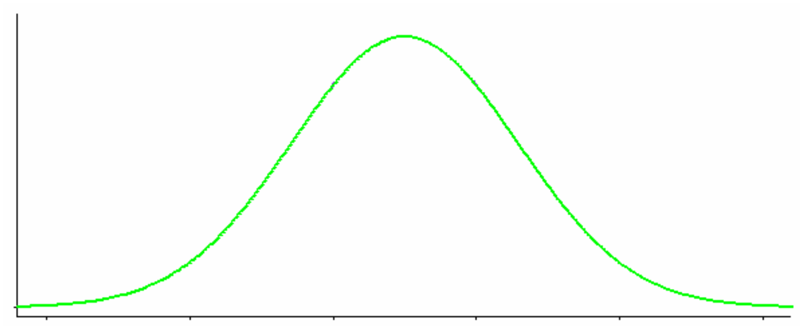
The Distribution of A Continuous R.V.

What problems might we encounter in attempting to identify a probability mass function for a continuous random variable?

A continuous random variable can be described by a **probability density function** $f(x)$. $f(x) \geq 0$ and $\int_S f(x)dx = 1$.

The graphs below are of probability density functions of continuous random variables.



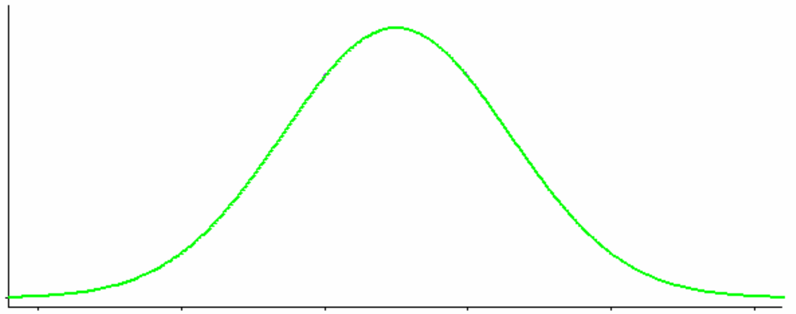


The Distribution of A Continuous R.V.

We can use the probability density function to find the probability that X lies in an interval, i.e. $P(a<x<b)$. The probability density function of a continuous random variable is also called its distribution.

The area under the curve over a given interval gives the probability that the random variable takes on a value in that region.

Note that $P(X=x)=f(x)=0$ for all values of x.



The Distribution of A Continuous R.V.

For pdf f(x), we can compute the probability over an interval as $P(a \le x \le b) = \int_a^b f(x)dx$

This has two possible interpretations:

1. The probability that an individual randomly selected from the underlying population will have a value of X that falls between a and b.

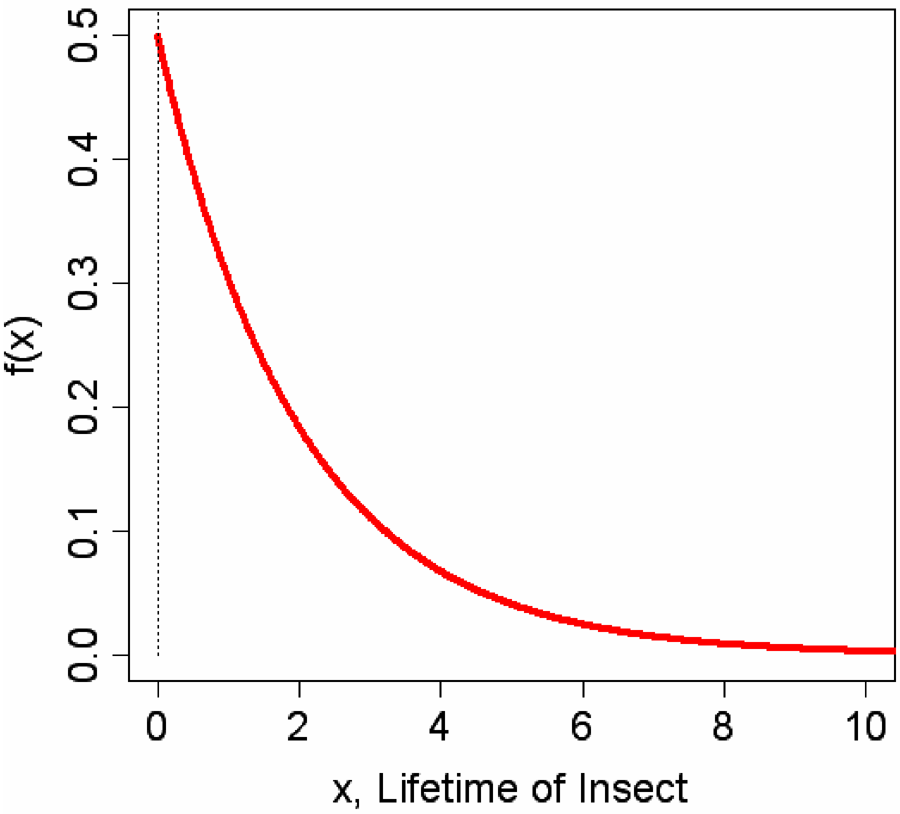
2. The proportion of individuals in the underlying population that have a value of X between a and b.

The Distribution of A Continuous R.V.

Example (Insect 1): The lifetime in months, X, of a particular insect has the probability distribution shown at the right and given by

$$f(x) = \frac{1}{2} e^{-x/2}, x \geq 0$$

Verify that this is a legitimate probability density function.

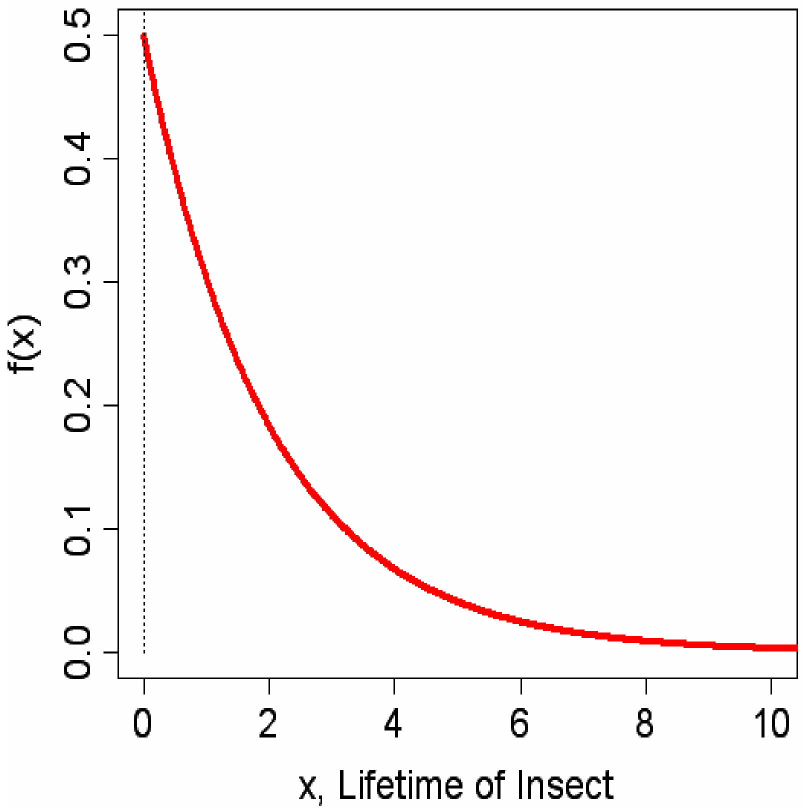


The Distribution of A Continuous R.V.

1. On the plot, diagram the area that represents the probability that a given insect lives no longer than 6 months.

2. Diagram the area that represents the proportion who live between 2 and 8 months.

3. Compute the probabilities that you diagrammed



The Distribution of A Continuous R.V.

Example (Insect 2): Suppose a second type of insect is know to live no longer than 4 months. The probability density function associated with the lifetime Y of this insect is given below:

$$f(y) = \frac{A}{2} e^{-y/2}, 0 \le y \le 4$$

~ What must A be in order for this to be a valid pdf?

The definition of the **cumulative distribution function** for a continuous random variable is the same as for a discrete random variable: $F(x) = P(X \le x)$.

The Distribution of A Continuous R.V.

Percentiles, quantiles, and quartiles provide information about the spread of a random variable.

For and random variable with cdf F(x), the pth **quantile** is the value of x for which F(x)=p. This is also referred to as the px100th **percentile**. The lower and upper **quartiles** refer to the .25, and .75 quantiles or the 25^{th} and 75^{th} percentiles.

For a random variable X with CDF F(X), the median is the value of X for which F(x)=0.5.

The interquartile range (IQR) is the distance between the lower and upper quartiles. That is if x_1 is the lower quartile and x_2 is the upper quartile, $IQR=x_2-x_1$.

The Distribution of A Continuous R.V.

1. Find and plot the cumulative distribution function for lifetime of insect 1.

2. Use the cdf to compute the probabilities that you diagrammed on the plot of the pdf.

3. What is the median lifetime?

4. What is the 75th percentile of lifetime?

Expectation and Variance

Suppose we'd like to know the average lifetime for an insect of the first type (insect 1).

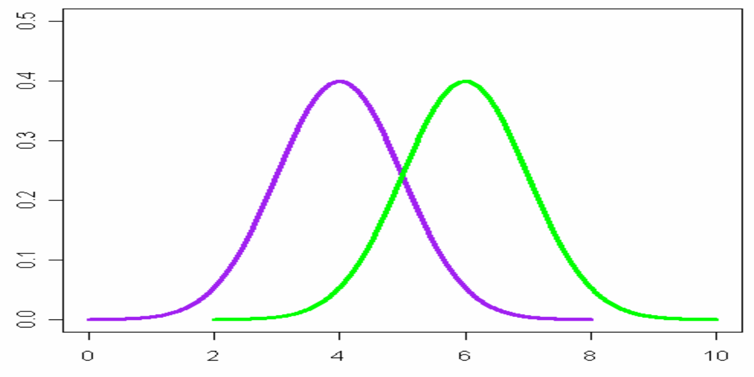
As with discrete random variable, the **expected value,** μ, of a continuous random variable can be thought of as a weighted average. For a continuous random variable with probability density function f(x), the expected value $E(X)=\int_S xf(x)dx$.

The **variance** of a continous random variable X is defined in the same way as for a discrete random variable, that is, $Var(X)=E((X-E(X))^2)$ or $Var(X)=E(X^2)-(E(X))^2$.

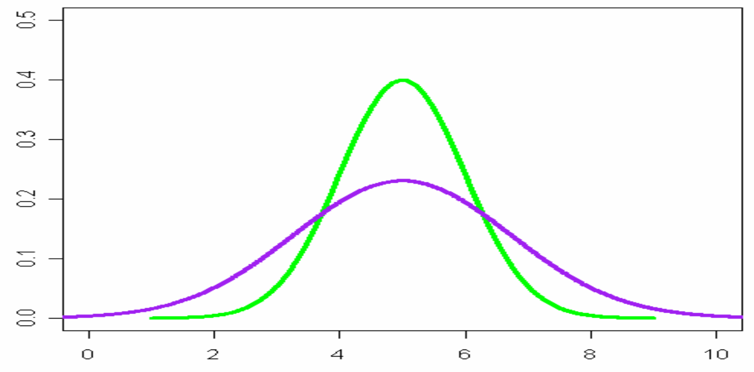
~ Find the expected value and variance of X as defined for insect 1.

Expectation and Variance

The top figure at the right shows the distributions of two random variables with equal variances but different means (expected values).



The bottom figure shows the distributions of two random variables with equal means, but different variances.



Combinations and Functions of R.V.s

Suppose we want to play roulette for bigger stakes, specifically we want to bet $100 on a number instead of $1. We can define a new random variable Y="the amount we win in one $100 bet on a number in roulette" or Y=100X. Then we can use what we know about X to find a new expectation and variance.

Suppose, instead, we want to keep the stakes the same, but we want to place three $1 bets instead of just one. Again we can define a new random variable Z= "the amount we win in three $1 bets on a number in roulette" or Z=X+X+X.

Combinations and Functions of R.V.s

Let X be a random variable such that $E(X)=\mu$ and $Var(X)=\sigma^2$. Suppose we're interested in a random variable Y defined as $Y = a + bX$, where a and b are constants. In other words, Y is a linear function of X. Then

$$E(Y)=E(a+bX)=E(a)+bE(X)=a+b\mu,$$

and

$$Var(Y)=Var(a+bX)=Var(a)+b^2Var(X)=0+b^2\sigma^2=b^2\sigma^2.$$

Combinations and Functions of R.V.s

Example: The random variables X, Y, and Z are independent with E(X)=1, Var(X)=3, E(Y)=2, Var(Y)=6, E(Z)=3, Var(Z)=4. Calculate the expected values and variances of the following random variables:

1. 2Z
2. Y-6
3. 3X + Y - 5Z - 7

Example: The length X of rainbow trout in a particular fishing hole, has expected value 35 cm with a variance of 10 cm. To translate to inches, let Y=.39(X). What are the expected value and variance of rainbow trout length in inches?

Combinations and Functions of R.V.s

Let X_1 and X_2 be independent random variables such that $E(X_1)=\mu_1$ and $Var(X_1)=\sigma_1^2$, and $E(X_2)=\mu_2$ and $Var(X_2)=\sigma_2^2$.

Let Y be a linear combination of X_1 and X_2, $Y=c_1X_1+c_2X_2$, where c_1 and c_2 are constants. Then

$$E(Y)=E(c_1X_1 + c_2X_2)=c_1E(X_1)+c_2E(X_2)=c_1\mu_1+c_2\mu_2,$$

and

$$Var(Y)=Var(c_1X_1 + c_2X_2)=$$
$$c_1^2Var(X_1)+c_2^2Var(X_2)=c_1^2\sigma_1^2+c_2^2\,\sigma_2^2.$$

The expectation of Y given above holds regardless of the dependence of X_1 and X_2. **However, the variance formula holds only if X_1 and X_2 are independent.**

Combinations and Functions of R.V.s

Example 1: The random variables X, Y, and Z are independent with E(X)=1, Var(X)=3, E(Y)=2, Var(Y)=6, E(Z)=3, Var(Z)=4. Calculate the expected values and variances of the following random variables:

1. 3X-2Y
2. 5X+2Y+Z-12

Example 2: Recall the insect 1 lifetime example. Suppose that we randomly sample 10 insects of this type and record the length of time they live.

- What is the expected combined lifetime of these 10 insects in months? in weeks (assume 4 weeks/month)? in years?
- What is the standard deviation of their combined lifetime in months? in weeks? in years?

Combinations and Functions of R.V.s

If $X_1,\ldots X_n$ is a sequence of independent random variables each with expectation μ and variance σ^2. The average of these random variables is

$$\overline{X} = \frac{X_1 + \ldots + X_n}{n}$$

What are the expected value and variance of $\overline{X}$?

Combinations and Functions of R.V.s

Example: Consider again the insect1 example. Suppose that we randomly select 20 insects and compute the average lifetime.

- What is the expected value of the average?

- What is the variance of the average?

- By what factor do we need to increase our sample size to divide the standard deviation of the average in half?

Combinations and Functions of R.V.s

There are not general rules to relate expectations and variances of a random variable X to a nonlinear function Y of that random e.g. $Y=X^2$, $Y=e^x$. In this case, we can construct the cumulative distribution function of Y from that of X. Consider the random variables X and Y such that $f_x(x)=x$ for $0 \le x \le 1$ and $Y=X^{1/2}$.

$F_x(x)=(1/2)X^2$

$F_y(Y)=P(Y \le y)=P(X^{1/2} \le y)=P(X \le y^2)= F_x(y^2)=(1/2)(y^2)^2=(1/2)(y^4)$

Therefore $f_y(y)=2y^3$.

Combinations and Functions of R.V.s

Example: Suppose that the random variable X has probability density $f_x(x)=1$ for $2 \le x \le 4$. Find the probability density function of the random variable Y in the following cases:

1. $Y=1/X$
2. $Y=3^x$
3. $Y=X^2$