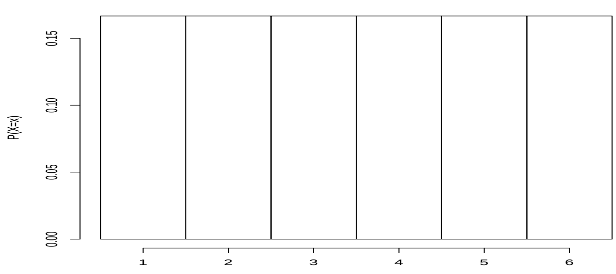
University of California, Los Angeles
Department of Statistics

Statistics 100A **Instructor: Nicolas Christou**

Random variables

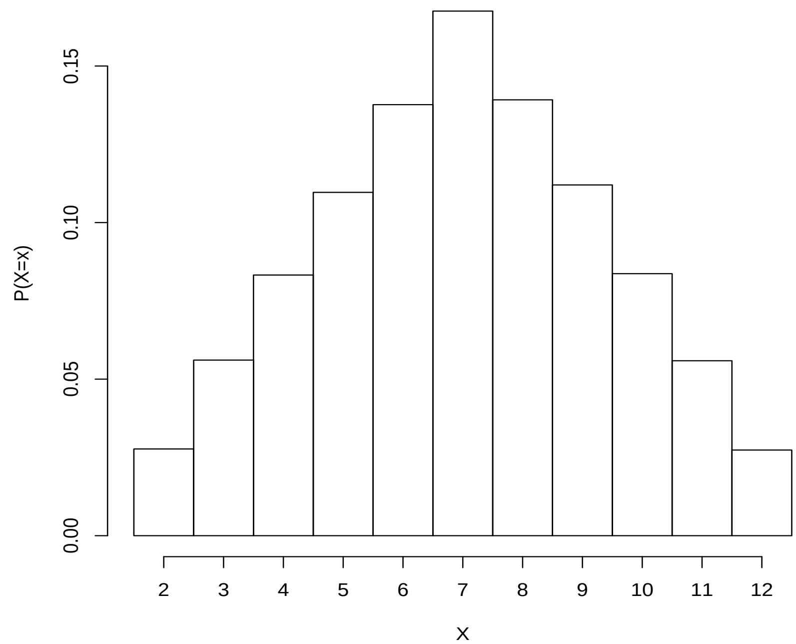
- Discrete random variables.
- Continuous random variables.
- **Discrete random variables**. Denote a discrete random variable with X:
  It is a variable that takes values with some probability.
  Examples:
  a. Roll a die. Let X be the number observed.
  b. Draw 2 cards with replacement. Let X be the number of aces among the 2 cards.
  c. Roll 2 dice. Let X be the sum of the 2 numbers observed.
  d. Toss a coin 5 times. Let X be the number of tails among the 5 tosses.
  e. Randomly select a US household. Let X be the number of people live in this household.
- **Probability distribution of a discrete random variable *X***
  It is the list of all possible values of X with the corresponding probabilities. It can be represented by a table, a graph, or a function.
  Examples:
  a. Roll a die. Let X be the number observed. The probability distribution of X is:

X	$P(X=x)$
1	$\frac{1}{6}$
2	$\frac{1}{6}$
3	$\frac{1}{6}$
4	$\frac{1}{6}$
5	$\frac{1}{6}$
6	$\frac{1}{6}$



b. Roll two dice. Let X be the sum of the two numbers observed. The probability distribution of X is:

X	$P(X=x)$
2	$\frac{1}{36}$
3	$\frac{2}{36}$
4	$\frac{3}{36}$
5	$\frac{4}{36}$
6	$\frac{5}{36}$
7	$\frac{6}{36}$
8	$\frac{5}{36}$
9	$\frac{4}{36}$
10	$\frac{3}{36}$
11	$\frac{2}{36}$
12	$\frac{1}{36}$



We can also represent this distribution with a function: $P(X=x)=\frac{6-|x-7|}{36}$, for $x=2,3,\cdots,12$. This is called probability mass function and returns the probability for each possible value of the random variable X.

- **Expected value (or mean) of a discrete random variable**
  It is denoted with $E(X)$ or μ and it is computed as follows:
  Definition:

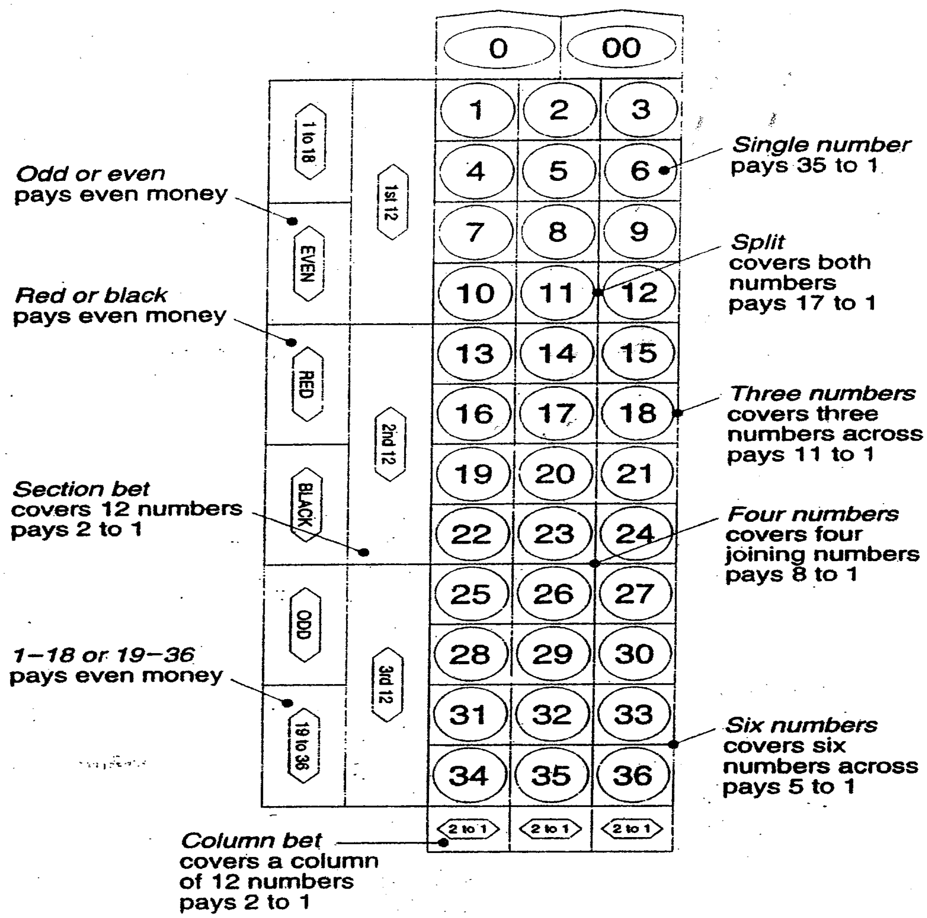
$$\mu = E(X) = \sum_x xP(X=x)$$

It is a weighted average. The weights are the probabilities.

Example:
Roll a die. Let X be the number observed. Find the expected value of X. The $E(X)$ must be somewhere between 1, 6: $E(X) = 1\frac{1}{6} + 2\frac{1}{6} + 3\frac{1}{6} + 4\frac{1}{6} + 5\frac{1}{6} + 6\frac{1}{6} = 3.5$.
What does this number mean?

Example: Casino roulette. Below you see the standard Nevada roulette table:

A NEVADA ROULETTE TABLE



Roulette is a pleasant, relaxed, and highly comfortable way to lose your money.
—JIMMY THE GREEK

FROM: "STATISTICS"
DAVID FREEDMAN
ROBERT PISANI
ROGER PURVES

A player will bet 1$ on four joining numbers. This bet pays 8 : 1. Let X be the player's payoff. Find the player's expected payoff.

- **Expected value of a sum of random variables**
  Let X and Y be 2 random varables. The expected value of the sum of these 2 random variables is:

  $$E(X + Y) = E(X) + E(Y)$$

  Example:
  Roll 2 dice. Let X be the number observed on the first die and Y be the number observed on the second die. Let W be the sum of the 2 dice. Find the expected value of W. There are 2 ways to solve this problem:

  a. $E(W) = E(X + Y) = E(X) + E(Y) = 3.5 + 3.5 = 7$

  b. Or using the distribution of the sum of the two dice (see page 2):

  $$E(W) = \sum_w wP(W = w) = 2\frac{1}{36} + 3\frac{2}{36} + \cdots + 12\frac{1}{36} = 7.$$

  The expected value of the sum can be extended to more than two random variables:

  $$E(X + Y + Z + \cdots) = E(X) + E(Y) + E(Z) + \cdots$$

- **Expected value of a function of a random variable**
  Let $g(X)$ a function of a discrete random variable X. Then its expected value is computed as follows:

  $$E[g(X)] = \sum_x g(x)P(X = x)$$

- **Variance and standard deviation of a discrete random variable**
  Consider the following 2 probability distributions:

X	$P(X)$	Y	$P(Y)$
0	$\frac{1}{2}$	5	$\frac{1}{2}$
1	$\frac{1}{2}$	-4	$\frac{1}{2}$

What do you observe?

Definition:

$$Var(X) = \sigma^2 = E(X-\mu)^2 = \sum_x (x-\mu)^2) P(X=x) = \sum_x x^2 P(X=x) - \mu^2$$

The standard deviation of a discrete random variable is the square root of the variance:

$$SD(X) = \sqrt{\sigma^2} = \sqrt{\sum_x (x-\mu)^2) P(X=x) = \sum_x x^2 P(X=x) - \mu^2}$$

It follows that:

$$\sigma^2 = EX^2 - \mu^2 \quad \text{or} \quad EX^2 = \sigma^2 + \mu^2$$

Example:
Roll a die. Let X be the number observed. Find the variance of X.

$$Var(X) = 1^2\frac{1}{6} + 2^2\frac{1}{6} + 3^2\frac{1}{6} + 4^2\frac{1}{6} + 5^2\frac{1}{6} + 6^2\frac{1}{6} - 3.5^2 = 2.917.$$

The standard deviation is:

$$SD(X) = \sqrt{2.917} = 1.708.$$

- **Some properties expectation and variance**
  Let X, Y random variables and a, b constants.

  a. $E(aX) = aE(X)$
  b. $E(aX+b) = aE(X) + b$
  c. $Var(X+a) = Var(X)$
  d. $Var(aX) = a^2 Var(X)$
  e. $Var(aX+b) = a^2 Var(X)$.
  f. If X, Y are independent then $Var(X+Y) = Var(X) + Var(Y)$

Example:
An insurance policy costs \$100, and will pay policyholders \$10000 if they suffer a major injury (resulting in hospitalization) or \$3000 if they suffer a minor injury (resulting in lost time from work). The company estimates that each year 1 in every 2000 policyholders may have a major injury, and 1 in 500 a minor injury.

a. Construct the probability distribution for the profit on a policy.

b. What is the company's expected profit on this policy?

c. Do you think the standard deviation is large or small. Why?

d. Compute the standard deviation.

e. Suppose that the company writes (a) 36, (b) 10000 of these policies per year. What are the mean and standard deviation of the annual profit for these 2 cases?

f. Comment!

Solution:

a. Let X the profit of the company on one of these insurance policies. The probability distribution of X is:

X	$P(x)$
-9900	0.0005
-2900	0.002
100	0.9975

b. Expected value of X:
$E(X) = -9900(0.0005) - 2900(0.002) + 100(0.9975) = 89$.

c. The standard deviation will be large.

d. Variance of X:
$var(X) = (-9900)^2(0.0005) + (-2900)^2(0.002) + (100)^2(0.9975) - 89^2 = 67879$. Therefore the standard deviation is: $sd(X) = \sqrt{67879} = 260.54$.

e. Here we need to find the expected value and variance of sum of random variables. In part (i) we have a sum of 36 random variables and in part (ii) a sum of 10000 variables.

Part(i):
$E(Y_1 + \ldots + Y_{36}) = 36(89) = 3204$.
$var(Y_1 + \ldots + Y_{36}) = 36(67879)$. The standard deviation is:
$sd(Y_1 + \ldots + Y_{36}) = \sqrt{36(67879)} = 1563.2$.

Part(ii):
$E(Y_1 + \ldots + Y_{10000}) = 10000(89) = 890000$.
$var(Y_1 + \ldots + Y_{10000}) = 10000(67879)$. The standard deviation is:
$sd(Y_1 + \ldots + Y_{10000}) = \sqrt{10000(67879)} = 26053.6$.

Some examples

Example 1
The number N of residential homes that a fire company can serve depends on the distance r (in city blocks) that a fire engine can cover in a specified (fixed) period of time. If we assume that N is proportional to the area of a circle r blocks from the firehouse, then $N = C\pi r^2$, where C is a constant, $\pi = 3.1416...$, and r, a random variable, is the number of blocks that a fire engine can move in the specified time interval. For a particular fire company, $C = 8$, the probability distribution for r is as shown in the accompanying table, and $p(r) = 0$ for $r \leq 0$ and $r \geq 27$.

r	21	22	23	24	25	26
$p(r)$	.05	.20	.30	.25	.15	.05

Find the expected value of N, the number of homes that the fire department can serve.

Example 2
Suppose that X takes on one of the values 0, 1, 2. If for some constant c, $P(X = i) = cP(X = i - 1), i = 1, 2$ find $E(X)$ in terms of c.

Example 3
Two balls are chosen randomly without replacement from and urn containing 8 white, 4 black, and 2 orange balls. Suppose that we win \$2 for each black ball selected and we lose \$1 for each white ball selected. We neither win nor we lose any money for selecting an orange ball. Let X denote our winnings.

a. What is the expected value of X?
b. What is the variance of X?
c. Given that our winnings are negative, what is the probability that we lost exactly \$2?

Example 4
To determine whether or not they have a certain disease, 100 people are to have their blood tested. However, rather than testing each individual separately, it has been decided first to group the people in groups of 10. The blood samples of the 10 people in each group will be pooled and analyzed together. If the test is negative, one test will suffice for the 10 people. If the test is positive each of the 10 people will also be individually tested. Suppose the probability that a person has the disease is 0.10 for all people independently from each other. Compute the expected number of tests necessary for each group.

Example 5
In exercise 5, the 100 people were placed in groups of 10. Try different group sizes in order to minimize the expected number of tests? For example, you can try 2, 4, 5, 20 etc? What group size gives the least expected number of tests?

Example 6

A man buys a racehorse for \$20000, and enters it in two races. He plans to sell the horse afterward, hoping to make a profit. If the horse wins both races, its value will jump to \$100000. If it wins one of the races, it will worth \$50000. If it loses both races, it will worth only \$10000. The man believes there is a 20% chance that the horse will win the first race and a 30% chance it will win the second one.

a. Assuming that the two races are independent events, find the man's expected profit.

b. Find the standard deviation of the man's profit.