Class 8: Expected Value and Variance of Random Variables (Text: Section 4.4)

A **random variable** is a number whose value depends on random process. If *X* is a random variable (RV), it is a function whose input is an experiment and whose output is a number, *x*.

EXPECTED VALUE OF A RANDOM VARIABLE

If we draw one of 10 slips of paper from a bowl, with \$0 written on nine of them and \$100 on the tenth, we will win nothing $^9/_{10}$ of the time and a hundred dollars $^1/_{10}$ of the time. On average, over many games, how much do you win per game?

$$\text{Average win per game} = {}^9/_{10} \cdot 0 + {}^1/_{10} \cdot 100 = \$10.$$

The \$10 is called the **expected value** of the game. Note that \$10 is not a value you can actually win—that is either \$0 or \$100—but it is the average winning per game over a very large number of plays.

Expected value is a weighted average, with more probable values being weighted more heavily.
For a random variable, X, it is written $E(X)$ or μ.
If x_1, x_2 .... are the values that X takes with probabilities p_1, p_2, ... then

$$E(X) = \mu_X = p_1x_1 + p_2x_2 + \cdots + p_nx_n$$

Ex: For these values of the prizes won in the Arizona Lottery,

(a) What is probability of winning nothing?
(b) Find the average win per ticket over many plays. What does it tell you about the price of a ticket?

Lottery	Arizona Pick Five[1]				
Win	\$50,000	\$500	\$5	\$1	\$0
Probability	$^1/_{575,757}$	$^1/_{3387}$	$^1/_{103}$	$^1/_{10}$	0.89

(a) $P(\text{no prize}) = 1 - \left({}^1/_{575,757} + {}^1/_{3387} + {}^1/_{103} + {}^1/_{10}\right) = 1 - 0.11 = 0.89.$

(b) Random variable is W = amount won from one ticket.
What is your average winning per game if you play for a long time?

$$E(W) = {}^1/_{575,757}(50{,}000) + {}^1/_{3387}(500) + {}^1/_{103}(5) + {}^1/_{10} = 0.383 \text{ or } 38.3 \text{ cents}$$

Since the state pays out 38.3 cents, on average, for every ticket sold, the price of ticket should be set above 38.3 cents $= E(W)$

Note:

- Expected value gives the amount the state pays out, on average, per ticket.
- In this case, $E(W)$, is not an amount you can actually win. It is the average value of your winnings, per ticket, over a large number of plays.
- Note that $E(X)$ does not include administrative costs.

[1] http://www.arizonalottery.com/Pick5.html?GameID=6&TokenID=150.135.120.6

Ex: How much should you pay per year for collision insurance for your car? Assume:

- Car is worth \$30,000.
- Chance of accident in a year is $1/20$
- If there is an accident, the car is totaled.
- Insurance pays the whole cost of replacing the car.

Random variable is the amount paid per policy per year for accident repairs.
With insurance: On average, each year the insurance company expects to pay for accident repairs

$$E(\text{cost}) = 1/20 \cdot \$30{,}000 + 19/20 \cdot \$0 = \$1500.$$

Thus, on average, the insurance company will pay out \$1500 per year per policy that it sells of this type; so the company will want at least \$1500 a year for the policy. Above this price "self-insuring" is cheaper, but you may want to buy the policy anyway because it brings you peace of mind—you can't be hit with a large unexpected cost. The policy was worth \$1500 per year to you, on average.

Ex: How much should you pay per year for collision car insurance for the same car assuming

- Same probability of accident
- If there is an accident, cost of repairs is uniformly distributed between \$0 and \$30,000.
- Insurance covers all repair costs

If you do have an accident,

$$\text{Average cost of repairs} = \text{Mean of uniform distribution from \$0 to \$30,000.}$$

Since pdf of a uniform distribution is a horizontal line and mean is balance point, we see

$$\text{Average cost of repairs} = \$15{,}000$$

Thus, on average, each year (including those when there is no accident), the insurance company expects to pay

$$E(\text{cost}) = 1/20 \cdot \$15{,}000 + 19/20 \cdot 0 = \$750.$$

So the insurance company will pay out on average \$750 a year for each such policy, and will charge at least \$750 a year for the policy. If you pay more, the additional over \$750 that you pay is for the peace of mind, not repairs. The policy was worth \$750 per year to you.

Ex: How much should you pay per year for collision car insurance with a \$1000 deductible? If there is an accident , the cost of repairs is uniformly distributed between \$0 and \$30,000. (Same car and probability of accident.)

Costs:
If the cost of repairs is \$1000 or below, we pay the whole cost; the average cost for a "small" accident is \$500.
If the repairs cost more than \$1000 (a "large" accident), we pay only \$1000.

Probabilities:
If there is an accident, the probability of it being "small" and costing less than \$1000 is $1/30$.
If there is an accident, the probability of its being "large" and costing more than \$1000 is $29/30$.
Since the probability of an accident is $1/20$:
The probability of small accident is $(1/20)(1/30) = 1/600$
The probability of large accident is $(1/20)(29/30) = 29/600$.
The probability of no accident is $19/20$

Expected cost:
The average yearly cost to you (the policy holder) is

$$E(\text{cost}) = 1/600 \cdot \$500 + 29/600 \cdot \$1000 + 19/20 \cdot 0 = \$49.17.$$

Insurance:
Without insurance, the average cost per year to you would have been \$750. With this insurance, you pay only \$49.17, so the insurance is worth $\$750 - \$49.17 = \$700.83$ to you. You should pay \$700.83 per year , and you may be willing to pay more for peace of mind.

VARIANCE AND STANDARD DEVIATION OF RANDOM VARIABLES

We define the **standard deviation** of a random variable in much the same way as the standard deviation of the sample: as the average distance from the expected value, with more probable values weighted more heavily.

The **variance** is the square of the standard deviation; it is useful when you need to find the standard deviation of sums and difference of random variables.

Interpretation and Notation:

Standard deviation is a weighted average of the distances from the mean (expected value), with more probable values being weighted more heavily. For a random variable, X, the standard deviation is written $SD(X)$ or σ; the variance is $Var(X)$ or $V(X)$ or σ^2. If X has expected value μ, then the variance is $$V(X) = \sigma^2 = p_1(x_1-\mu)^2 + p_2(x_2-\mu)^2 + \cdots + p_n(x_n-\mu)^2$$ $$SD(X) = \sqrt{V(X)}$$

Ex: Find the variance and standard deviation for Arizona Lottery

$$\text{Variance} = V(W)$$
$$= \left(1/_{575,757}\right)(50{,}000-0.383)^2 + \left(1/_{3387}\right)(500-0.383)^2 + \left(1/_{103}\right)(5-0.383)^2$$
$$+ \left(1/_{10}\right)(1-0.383)^2 + 0.89(0-0.383)^2 = 4416.12.$$
$$\text{Standard deviation} = SD(W) = \sqrt{4416.12} = 66.45.$$

GENERAL FORMULAS

For a Discrete Random Variable:

If the p.d.f of the random variable X is given by

Values of X	x_1	x_2	x_3	$\cdots$	x_k
pdf	p_1	p_2	p_3	$\cdots$	p_k

then

Expected value	***Variance***	***Standard deviation***
$E(X) = \mu_x = \sum_{i=1}^{i=k} p_i x_i$	$Var(X) = \sigma_X^2 = \sum_{i=1}^{i=k} p_i (x_i - \mu_x)^2$	$SD(X) = \sigma_X = \sqrt{V(X)}$

For a Continuous Random Variable:

In this course, for a continuous random variable, X ,you will be given $E(X)$ and $V(X)$ or $SD(X)$. If $p(x)$ is the pdf of a continuous random variable X, although these formulas are ***not needed*** in Math 263, they are:

Expected value	***Variance***	***Standard deviation***
$E(X) = \mu_x = \int_{-\infty}^{\infty} xp(x)\,dx$	$Var(X) = \sigma_X^2 = \int_{-\infty}^{\infty} (x-\mu_x)^2 p(x)\,dx$	$SD(X) = \sigma_X = \sqrt{V(X)}$

SUMS AND DIFFERENCES of RANDOM VARIABLES

Expected Value:

1. If a and b are numbers $E(a+bX) = a + bE(X)$ that is, $\mu_{a+bX} = a + b\mu_X$
2. If X and Y are random variables $E(X+Y) = E(X) + E(Y)$ that is, $\mu_{X+Y} = \mu_X + \mu_Y$

Variance and Standard Deviation:

1. If a and b are numbers: $\sigma^2_{a+by} = V(a+bY) = b^2 V(Y)$
2. If X and Y are independent random variable (so correlation $r = 0$)
$$\sigma^2_{x+y} = V(X+Y) = \sigma^2_x + \sigma^2_y$$
$$\sigma^2_{x-y} = V(X-Y) = \sigma^2_x + \sigma^2_y$$
3. (***Optional***) If X are Y are not independent (so correlation $r \neq 0$)
$$\sigma^2_{x+y} = V(X+Y) = \sigma^2_x + \sigma^2_y + 2r\sigma_x\sigma_y$$
$$\sigma^2_{x-y} = V(X-Y) = \sigma^2_x + \sigma^2_y - 2r\sigma_x\sigma_y$$

Ex: Ear lengths: For women, ear lengths are normally distributed with mean 2.06 inches and standard deviation 0.17 inches; for men they are normally distributed with mean 2.45 inches and standard deviation 0.17 inches

(a) A woman and a man are randomly selected. What are the mean and standard deviation of the difference in their ear lengths?
(b) Assuming that the difference in ear lengths is normally distributed (which it is), what is the probability that the woman's ear length is longer than the man?
(c) What difference does it make if the man and the woman are a couple?

Let M be the man's ear length and F be the woman's. Subtracting the woman's ear ength from the man's (but could have been the other way):

(a) The mean of difference is given by
$$E(M-F) = E(M) - E(F) = 2.45 - 2.06 = 0.39 \text{ inches.}$$
Since the man and the woman are randomly selected,
$$V(M-F) = V(M) + V(F) = (0.17)^2 + (0.17)^2 = 0.0578.$$
Thus
$$SD(M-F) = \sqrt{0.587} = 0.2404 \text{ inches.}$$
(b) We are looking for the probability that $M - F$is negative. Since the difference in lengths is normally distributed with mean 0.39 and standard deviation 0.76 inches, the z-value of 0 is
$$z = \frac{0 - 0.39}{0.24} = -1.625.$$
From the table, the probability that $z \leq -1.625$ is $0.0516 = 5.16\%$.
(c) Whether it matters that the man and woman are a couple depends on whether there is a correlation between their ear lengths—probably yes to some extent.

<u>Ex: What is variance and standard deviation of winnings from two lottery tickets, if the winnings are independent</u>, given $\sigma_x = \$66$ and $\sigma_y = \$50$.

We add the variances, so $(\sigma_x)^2 = 4357$ and $\left(\sigma_y\right)^2 = 2500$.
The variance of the combination is

$$\left(\sigma_{x+y}\right)^2 = \sigma_x^2 + \sigma_y^2 = 4356 + 2500 = 6856$$

The standard deviation is given by

$$\sigma_{x+y} = \sqrt{6856} = \$82.8$$

Note: Cannot add standard deviations; add variance.

<u>Ex: How do people determine how to invest in stock market</u>? (Optional)

<u>Stocks</u> have higher earnings, but carry more risk.
Let X be yearly earnings of an "index fund" which is a combination of many stocks, with $E(X) = 15\%$ and $\sigma(X) = 25\%$. (Earnings are given as a percentage of what you have invested.)

<u>T-bills</u>: Treasury bills earn less, but vary less.
Let Y be yearly earning of a treasury bill, with $E(Y) = 5\%$ and $\sigma(Y) = 3\%$

Suppose the correlation between X and Y is $r = -0.1$

(a) What fraction of investment should go to which to maximize earnings? What is the standard deviation in this case?
(b) What is the expected value if the money is split 50-50 between the stocks and the T-bill? What is the standard deviation in this case?

(a) Put all the money in stocks if the goal is to maximize earnings. The standard deviation is 25%.
(b) The random variable giving the annual return is $0.5X + 0.5Y$. The expected value is between the expected value of the stocks and T-bills and given by

$$E(0.5X + 0.5Y) = 0.5(0.15) + 0.5(0.05) = 0.1 = 10\%$$

Since there is no formula for the standard deviation of $aX + bY$, we use the formula for the variance:

$$V(aX + bY) = a^2V(X) + b^2V(Y) + 2(-0.1)ab\sigma_x\sigma_y$$

So

$$V(0.5X + 0.5Y) = 0.5^2(0.25)^2 + 0.5^2(0.03)^2 - 0.2(0.5)^2(0.25)(0.03) = 0.0155.$$

Thus

$$SD(0.5X + 0.5Y) = \sqrt{0.0155} = 0.124 = 12.4\%$$

So this combination less risky (because the standard deviation is less), but also earns less.