Random Variables

A random variable X is a variable whose value is determined by the result of a random phenomenon. When there are many possible outcomes ("events") it is convenient to define a random variable (r.v.) rather than naming all of the different events. For example, if we flip a coin 10 times, there are $2^10 = 1024$ possible sequences of heads and tails we might observe (two possible outcomes for each of the 10 flips). We could define the random variable X to be the number of heads. Then the event "$X = 2$" (e.g.) is the union of all outcomes where there are exactly 2 heads and exactly 8 tails (there are 45 such outcomes). More generally, we can write "$X = x$" to represent an arbitrary value of x heads in 10 tosses. I use capitol X to represent the variable itself, and lower case x to represent a possible value (in this example, possible values are $x = 0, 1, \ldots, 10$).

The Mean of a Random Variable The mean μ_x of a random variable X is the long-run average value you would expect to see if you repeated the random phenomenon many times. To find μ_x, take a weighted average of all possible values, weighting by the probability of each value.

$$\mu_x \;=\; \sum_{\text{all } x} xP(X = x).$$

Lottery example: Quick Draw - 3 spot game This New York state lottery game is also called *Keno*. The rules for the 3 spot game are as follows:

1. You pay \$1 to play, and select 3 distinct numbers from $1, 2, \ldots, 80$.
2. The State randomly selects 20 distinct numbers from $1, 2, \ldots, 80$.
3. Your payout is determined by how many of your numbers match the state's numbers:

# of matches	Payout (x)	Winnings (y)	Probability
0 , 1	\$0	−\$1	0.847
2	\$2	\$1	0.139
3	\$23	\$22	0.014

The mean payout is given by $\mu_x = 0(0.847) + 2(0.139) + 23(0.014) \;=\; \$0.60$.

The mean winnings is given by $\mu_y = (-1)(0.847) + 1(0.139) + 22(0.014) \;=\; -\$0.40$.

This tells us that, on average, New York pays out only 60 cents for every dollar bet, meaning you lose 40 cents on average for each dollar bet. This is why NY can count on earning money with this lottery game.

Means for Linear Combinations of random variables
If a r.v. X has mean μ_x and $Y = a + bX$ is a linear function of X, then the mean of Y is

$$\mu_y \;=\; a + b\mu_x.$$

For example, the winnings is the payout minus one dollar: $Y = -1 + X$. So

$$\mu_y \;=\; -1 + \mu_x \;=\; -1 + 0.6 \;=\; -0.4.$$

Special Promotion: DOUBLE PAYOUTS ON WEDNESDAY!!

When this game was first introduced in September 1998, NY had a special promotion where they had doubled payouts on each Wednesday of that month. The new distribution of payouts and winnings is given by

# of matches	Payout (x^*)	Winnings (y^*)	Probability
0 , 1	\$0	$-$\$1	0.847
2	\$4	\$3	0.139
3	\$46	\$45	0.014

The mean payout with doubled payouts is $\mu_{x^*} \;=\; 0(0.847) + 4(0.139) + 46(0.014) \;=\; \$1.20.$

The mean winnings with doubled payouts is $\mu_{y^*} \;=\; 0(0.847) + 3(0.139) + 45(0.014) \;=\; \$0.20.$

We could also figure these out using the rule for linear combinations:

$$X^* = 2X, \quad \text{so} \quad \mu_{x^*} = 2(\mu_x) \;=\; 2(0.6) = 1.20.$$
$$Y^* = -1 + 2X, \quad \text{so} \quad \mu_{y^*} = -1 + 2(\mu_x) \;=\; -1 + 2(0.6) = 0.20.$$

With doubled payouts, the mean winnings changes from -0.40 to $+0.20$, so instead of losing 40 cents on average, you make 20 cents on average. This makes it a "winning game" for the player (and a losing game for NY).

Notice that there is still probability 0.847 of losing your dollar on any single play. But because the mean winnings is positive, you would expect to earn money if you played the game enough times to establish a "long run." Four graduates from St. Lawrence University in upstate NY recognized the potential to win money with these rules, and set about playing the game as many times as they could on the four Wednesdays of the promotion. They ended up winning over \$100, 000!!!

Standard Deviation of a Random Variable

The law of large numbers says that the average of a large number of independent values will be very close to the mean value. Another way of saying that is that average becomes less variable as we increase the number of values being averaged. We measure variability using standard deviation.

Suppose a random variable has mean outcome μ_x and standard deviation σ_x. If we collect n independent values of the variable and average them, the mean and standard deviation of the average $\bar{x}$ are given by

$$\mu_{\bar{x}} \;=\; \mu_x \quad \text{and} \quad \sigma_{\bar{x}} \;=\; \frac{\sigma_x}{\sqrt{n}}.$$

The mean of the average of n is the same as the mean for each individual variable. But the standard deviation decreases as n increases, meaning that the average of a large number of values will tend to be close to the mean.

For example, the mean winning for Quick draw under doubled payouts is $\mu_x = \$0.20$, with a standard deviation of $\sigma_x = \$5.52$. If you play 100 times, the mean of your average winnings is $\mu_{\bar{x}} = \$0.20$, with a standard deviation of $\sigma_{\bar{x}} = 5.52/\sqrt{100} = \$0.552$.