The Law of Averages

In this class we will consider the law of averages. We will introduce the concepts of expected value and standard error. We will see that random variables can be represented using a box model. This in turn can be used to calculate expected values, standard errors and approximate the probabilities of relevant intervals.

Averages

When tossing a fair coin the chances of tails and heads are the same: 50% and 50%. So if the coin is tossed a large number of times, the number of heads and the number of tails should be approximately equal.

This is the law of averages.

The number of heads will be off half the number of tosses by some amount. That amount is called chance error. So we have

number of heads = half the number of tosses + chance error

The chance error increases with the number of tosses in absolute terms, but it decreases in relative terms.

Q: A coin is tossed and you win a dollar if there are more than 60% heads. Which is better: 10 tosses or 100?

A: 10 tosses is better. As the number of tosses increase you are more likely to be close to 50%, according to the law of averages.

Q: Same as before, but you win one dollar if there are exactly 50% heads.

A: 10 tosses is better, since in absolute terms you are more likely to be off the expected value when the number of tosses is large.

Q: 100 tickets are drawn with replacement from one of two boxes: one contains two tickets with -1 and two with 1. The other contains one ticket with -1 and one with 1. One hundred tickets will be drawn at random with replacement from one of them and the amount on the ticket will be paid to you, which box do you prefer?

A: In this case the expected payoff is the same for both boxes, since they both have 50% 1 and 50% -1.

A box model

A box model is a useful device to understand the properties of random variables. These are variables whose results depend on the outcome of a random experiment. Two quantities that describe the behavior of a random variable are the Expected Value and the Standard Error.

A box model relates to a process that can be simulated by considering that there are a number of tickets in a box, each of them has a number and they are drawn at random. The numbers that result are added.

Consider a box model for a roulette.

A roulette wheel has 38 pockets. 1 through 36 are alternatively colored red and black, plus 0 and 00 which are colored green. So, there are 18 red pockets and 18 black ones.

Suppose you win $1 if red comes out and loose $1 if either a black number or 0 or 00 come out.

Your chance of winning is 18 to 38 and your chance of loosing is 20 to 38. A box representation is

18 tickets	+$1	20 tickets	-$1

Suppose now that you bet a dollar on a single number and that, if you win, you get your $1 plus $35, but you loose your $1 if any other number comes up.

A box model of this bet is given by

1 ticket	+$35	37 tickets	-$1

What is your net gain after 100 plays?

This correspond to the amount of money you are left with after 100 draws with replacements from the box.

To calculate this amount we need the concept of Expected value.

Expected value

Count the number of heads in 100 tosses of a coin. The law of large numbers tells you that the expected value of the number of heads is 50. You actually toss he coin and the results are

- 57 heads, you are off by 7
- 46 heads, you are off by -4
- 47 heads, you are off by -3

The amounts off are similar in size to the standard error, which we will define in a couple of slides.

The expected value and the standard error depend on the random process that generates the numbers.

Consider the box [1] [1] [1] [5] and draw a ticket at random with replacement 100 times. What is the expected value of the sum of the tickets?

The chance of a 1 is 75% and the chance of a 5 is 25%. So we expect to see

$$25 \times 5 + 75 \times 1 = 200$$

Notice that this number is equal to

$$100 \times 2 = 200$$

which is the number of draws times the average number in the box.

As a general rule we have that the expected value is given by

(number of draws) × (average of box)

Q: Suppose you play Keno, a game where you win \$2 and pay \$1 to play. You have 1 chance in 4 to win. How much should you expect to win if you play 100 times?

A: A box representation of the game is given by

\$2	-\$1	-\$1	-\$1

so the average in the box is $\$2 - 3 \times \$1 = -\$0.25$.

So you are expected to 'win' -\$25. Of course, if you keep playing you are expected to loose more money!

Q: Consider the box

0	2	3	4	6

and suppose 25 draws with replacement are made from the box. What is the expected sum?

A: Each number should appear 1/5 of the time, that is 5 on average. So the expected value of the sum is

$$5 \times 0 + 5 \times 2 + 5 \times 3 + 5 \times 4 + 5 \times 6 = 75 = 25 \times 3$$

The standard error

In the last example we will not see each ticket appearing exactly 5 times. The actual sum we observe will be off by a chance error

$$\text{sum} = \text{observed value} = \text{expected value} + \text{chance error}$$

The standard error gives a measure of how large the chance error is likely to be.

When drawing at random with replacement from a box we can calculate the standard error for the sum of the draws as

$$\sqrt{(\text{number of draws})} \times (\text{SD of box})$$

where SD of box stands for the standard deviation of the list of numbers in the box.

Notice that the SE increases as the number of draws increases, but only by a factor equal to the square root of the number of draws.

Back to the example of the box with 5 numbers. to compute the SD we need the average, which is equal to 3. The deviations from the average are

$$-3 \quad -1 \quad 0 \quad 1 \quad 3$$

then the SD

$$\sqrt{\frac{9+1+0+1+9}{5}} = \sqrt{4} = 2.$$

If we do 25 draws from the box the SE is $\sqrt{25} \times 2 = 10$. So we expect the sum of the draws from the be around 75 plus or minus 10.

> Observed values are rarely more than 2 or 3 SEs away from the expected value

Can we make the previous statements more precise?

Consider again the box

0	2	3	4	6

. We know that in 25 draws the expected value is 75 and the SE is 10. Also, we observe that in 25 draws the sum ranges from 0 to 150 (all 0 through all 6). What are the chances that the sum will be between 50 and 100?

To answer this question we observe that

$$50 - 75 = -25 = -2.5 \times 10 \quad \text{and} \quad 100 - 75 = 25 = 2.5 \times 10$$

So that 50 and 100 are 2.5 times SEs away from the expected value. We say that 25 is 2.5 standard units.

We can apply the following approximation that will be better justified by the use of the Central Limit Theorem:

- 68% of the draws will be within one standard unit of the expected value.
- 95% of the draws will be within two standard units of the expected value.
- 99% of the draws will be within 2.5 standard units of the expected values.

For our example the ranges we get for one, two and 2.5 standard units are 75 ± 10; 75 ± 20 and 75 ± 25.

Q: Suppose there are 10,000 independent plays on a roulette wheel in a casino. Suppose all plays are of \$1 on red at each play. What are the chances that the casino will win more than \$250 from these plays?

A: The box model for this problem is

20 tickets	+\$1	18 tickets	-\$1

The expected value of the casino's net gain is the average

$$\frac{\$20 - \$18}{38} = \frac{\$2}{38} \approx \$0.05$$

times the number of plays:

$$10,000 \times \$0.05 = \$500$$

So the casino expects to win \$500 a month.

We now need to calculate the SD. We have 20 deviations of 0.95

and 18 deviations of -1.05 from the average. So

$$\text{SD} = \sqrt{\frac{20 \times (0.95)^2 + 18 \times (1.05)^2}{38}} = 0.9986175$$

being conservative we can approximate this number to 1. So the SE is approximately $\sqrt{10,000} = 100$.

\$250 is 2.5 SE units from the expected gain of \$500, so we see that there is 99% chance that the net gain for the casino will be between \$250 and \$750.