

Stat 1001  
Winter 1998  
Geyer

## Homework 6

### Problem 16.1

The answer is (iii).

(i) is false because you can get anywhere from 0 to 10,000  $\boxed{1}$ 's. The probability of being far from 60%  $\boxed{1}$ 's is very small, but not zero.

(iii) is false because the probability of *zero error* is very unlikely when the number of draws is large. You are likely to get somewhere near 60%  $\boxed{1}$ 's, but not exactly 60%.

### Problem 16.3

Both are wrong. The notion of independent repetitions says the gambler's probability of winning the next bet is *exactly the same as always*.

The gambler's mistake is thinking that averages even out *in the short run*. But that is not what the law of averages says. It says *in the long run*. Its working does not require probabilities changing over time.

The bystander's mistake is thinking that the roulette wheel goes on streaks. It doesn't. The roulette wheel does not remember the past. It can't change the probabilities either way.

### Problem 16.5

False. We could use the same logic as we used for Problem 16.1 here, but this problem is even simpler. The percentage of heads is exactly 50% if and only if the number of heads is 50. So the two probabilities must be *exactly the same*. There is no way one can be "likely" and the other "not likely." The right answer is that both are "not likely" but we don't need to know that to see that the answer must be "false."

### Problem 16.6

Both of these are about the same *percentage error* (relative error). What is the probability that the percentage error will be more than  $2/3 - 1/2 = 1/6$ ? (10/15 and 20/30 are both 2/3.)

The *percentage error* goes down as the number of draws goes up, so (ii) is less likely and (i) is more likely.

### Problem 16.7

... like the sum of 25 draws from the box 

4	-1	-1	-1	-1
---	----	----	----	----

.

### Problem 17.1

(a) The draws could be all  $\boxed{1}$ 's and the sum of draws 100. The draws could be all  $\boxed{10}$ 's and the sum of draws 1000.

(b) The average of the box is  $(1 + 6 + 7 + 9 + 9 + 10)/6 = 7$ .

The deviations from average are  $-6, -1, 0, 2, 2, 3$ . The squared deviations are 36, 1, 0, 4, 4, 9. The sum of the squared deviations is 54. The average squared deviation is  $54/6 = 9$  and the SD is  $\sqrt{9} = 3$ .

The *expected value* for the sum of draws is

$$(\text{number of draws}) \times (\text{average of box}) = 100 \times 7 = 700$$

The *standard error* for the sum of draws is

$$\sqrt{\text{number of draws}} \times (\text{SD of box}) = \sqrt{100} \times 3 = 30$$

The range the question asks about, 650 to 750, is  $50/30 = 1.67$  SE to either side of the expected value. Using the normal curve table we see that 90.8% of the area under the normal curve is in this range (between 90.11 and 91.09, closer to the latter).

### Problem 17.4

This is a *classifying and counting* problem (Section 5) so we need to recode the tickets  $\boxed{0}$  and  $\boxed{1}$ . We are counting aces, so we recode that  $\boxed{1}$  and the rest  $\boxed{0}$ .

The box is  $\boxed{\boxed{1} \boxed{0} \boxed{0} \boxed{0} \boxed{0} \boxed{0}}$ . Using the short cut (Section 4) the SD of the box is

$$\begin{aligned} \left( \begin{array}{c} \text{bigger} \\ \text{number} \end{array} - \begin{array}{c} \text{smaller} \\ \text{number} \end{array} \right) \times \sqrt{\begin{array}{c} \text{fraction with} \\ \text{bigger number} \end{array} - \begin{array}{c} \text{fraction with} \\ \text{smaller number} \end{array}} \\ = (1 - 0) \times \sqrt{\frac{1}{6} \times \frac{5}{6}} = 0.372678 \end{aligned}$$

The average of the box is  $1/6 = 0.1666667$ .

The *expected value* for the sum of draws is

$$(\text{number of draws}) \times (\text{average of box}) = 180 \times 0.1666667 = 30$$

The *standard error* for the sum of draws is

$$\sqrt{\text{number of draws}} \times (\text{SD of box}) = \sqrt{180} \times 0.372678 = 5$$

The range the question asks about, 15 to 45 is  $15/5 = 3$  SE to either side of the expected value. Using the normal curve table we see that 99.73% of the area under the normal curve is in this range.

### Problem 17.7

(a)  $321/100 = 3.21$ .

(b)  $3.78 \times 100 = 378$ .

(c) We haven't learned anything about the average of draws (yet). But using part (b) as a hint we can rephrase the problem so it is about the sum of draws. If the average of draws is 3, the sum is 300, and if the average is 4, the sum is 400. So the problem "estimate the chance that the sum of draws is between 300 and 400" has the same answer as the one originally asked.

The average of the box is  $(1 + 2 + 3 + 4 + 5 + 6)/6 = 3.5$ .

The deviations from the average are  $-2.5, -1.5, -0.5, 0.5, 1.5, 2.5$ . The squared deviations are  $6.25, 2.25, 0.25, 2.25, 6.25$ . The sum of the squared deviations is  $17.5$ . The average squared deviation is  $17.5/6 = 2.916667$  and the SD is  $\sqrt{2.916667} = 1.707825$ .

The *expected value* for the sum of draws is  $100 \times 3.5 = 350$ .

The *standard error* for the sum of draws is  $\sqrt{100} \times 1.707825 = 17.07825$ .

The range the question asks about, 300 to 400, is  $50/17.07825 = 2.93$  SE to either side of the expected value. Using the normal curve table we see that 99.66% of the area under the normal curve is in this range (between 99.63 and 99.68, closer to the latter).

### Problem 17.8

(a) Like (ii) because we code +1 for heads and  $-1$  for tails for the sum of draws to be heads  $-$  tails.

(b) The average of the box is  $[(-1) + (+1)]/2 = 0$ .

The SD of the box, using the short cut formula is

$$[(+1) - (-1)] \times \sqrt{.5 \times .5} = 2 \times .5 = 1$$

The *expected value* for the sum of draws is  $100 \times 0 = 0$ .

The *standard error* for the sum of draws is  $\sqrt{100} \times 1 = 10$ .

### Problem 17.9

The point of this question is that you have to calculate; intuition won't get it.

The box for (i) has 12 2's and 26 -1's. The box for (ii) has 1 35 and 37 -1's.

The average of box (i) is

$$\frac{12 \times 2 + 26 \times (-1)}{38} = -\frac{2}{38} = -\frac{1}{19}$$

and the average of box (ii) is

$$\frac{1 \times 35 + 37 \times (-1)}{38} = -\frac{2}{38} = -\frac{1}{19}$$

Both boxes have the same average.

The SD of box (i) is

$$[2 - (-1)] \times \sqrt{\frac{12}{38} \times \frac{26}{38}} = 3 \times 0.4648295 = 1.394489$$

and the SD of box (ii) is

$$[35 - (-1)] \times \sqrt{\frac{1}{38} \times \frac{37}{38}} = 36 \times 0.1600727 = 5.762617$$

The SDs are a lot different.

For both boxes, the *expected value* for the sum of draws is  $1,000 \times -\frac{1}{19} = -52.63158$ .

The *standard error* for the sum of draws is  $\sqrt{1,000} \times 1.394489 = 44.0976$  for box (i) and  $\sqrt{1,000} \times 5.762617 = 182.23$  for box (ii).

Now we are ready to answer the questions. There is a quick way to answer all three with no further calculations. Since the SE is larger for (ii), the result will typically be farther from the expected value, hence one is more likely to win or lose big with (ii). Thus the answers are false, true, true.

For those who are curious about the exact numbers . . . .

**(a)** False. Not the same because the SEs are different. The break even point (zero) is  $52.63158/44.0976 = 1.19$  SEs above the expected value for (i) and  $52.63158/182.23 = 0.29$  SEs above the expected value for (ii). The table of normal curve tail areas tells us the probability of coming out ahead is 11.6% (between 12.51 and 11.51 and closer to the latter) for (i) and is 38.6% (between 40.13 and 38.21 and closer to the latter) for (ii). Quite a difference.

**(b)** True.  $+100$  is  $152.63158/44.0976 = 3.46$  SEs above the expected value for (i) and  $152.63158/182.23 = 0.84$  SEs above the expected value for (ii). The probabilities come out to be essentially zero for (i) and about 20% for (ii).

**(c)** True.  $-100$  is  $47.36842/44.0976 = 1.07$  SEs below the expected value for (i) and  $47.36842/182.23 = 0.26$  SEs below the expected value for (ii). The probabilities come out to be about 14% for (i) and about 40% for (ii).