The Hat Model

The random process associated with the hat model consists of drawing a ticket from the hat. Typically that ticket is then put back into the hat before another ticket is drawn. Each ticket has an outcome written on it. Multiple tickets may have the same outcome written on them. Each *ticket* in the hat has an equal likelihood (equally likely outcomes page 306) of being drawn: $\frac{1}{\text{total number of } tickets}$. However, *outcomes* which appear on more tickets will be more likely than outcomes which appear on fewer tickets.

Example 1 Suppose a hat contains the following tickets: $\boxed{1}$, $\boxed{2}$, $\boxed{2}$, $\colorbox{lightgray}{1}$, $\colorbox{lightgray}{2}$, $\colorbox{lightgray}{2}$. Here, each outcome is a number and a ticket color. Since each ticket is equally likely, we can compute the probability of drawing a "2" (regardless of color) by counting the number of tickets with a "2" printed on it and divide by the total number of tickets: $P(2) = \frac{4}{6}$ which reduces to $\frac{2}{3}$.

Conditional Probability

As a notational convenience we use $P(E)$ to mean the probability of a particular event, E, occurring. A conditional probability $P(E_1 \,|\, E_2)$, which is read "the probability of E_1 *given* E_2", is the probability that an event, E_1, occurs given the knowledge that another event, E_2, has or will occur. A condition such as this changes the sample space of the probability.

Example 2 We return to our hat example (Example 1) above: $\boxed{1}$, $\boxed{2}$, $\boxed{2}$, $\colorbox{lightgray}{1}$, $\colorbox{lightgray}{2}$, $\colorbox{lightgray}{2}$. If we wish to compute $P(2 \,|\, \colorbox{lightgray}{})$, the probability that we draw a two given that we know that the ticket is shaded, we consider only the restricted sample space of the *three* tickets which are shaded. Thus, $P(2 \,|\, \colorbox{lightgray}{}) = \frac{2}{3}$

Example 3 It can be convenient to arrange our information into a table before attempting to compute probabilities:

Suppose you have a large hat with 50 $\boxed{1}$ tickets, 25 $\boxed{2}$ tickets, 24 $\colorbox{lightgray}{1}$ tickets, and 38 $\colorbox{lightgray}{2}$ tickets. When each outcome has two variables like this (number and color here) We can arrange this information in a table.

	$\boxed{}$	$\colorbox{lightgray}{}$	totals
"1"	50	24	74
"2"	25	38	63
totals	75	62	137

Then probabilities can be read off of the table easily: $P(1) = \frac{74}{137}$, $P(1 \,|\, \colorbox{lightgray}{}) = \frac{24}{62}$, $P(\colorbox{lightgray}{} \,|\, 2) = \frac{38}{63}$.

Probability Formulas

Notation:
The intersection, $E_1 \cap E_2$, indicates that *both* of the events, E_1 and E_2, must occur.
The union, $E_1 \cup E_2$, indicates that E_1, E_2, or both E_1 and E_2 occur.

$$P(E_1 \cap E_2) = P(E_1) \cdot P(E_2 \,|\, E_1)$$
$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$$

If $P(E_1)$ is the same as $P(E_1|E_2)$ we say that E_1 and E_2 are *independent*.

If $P(E_1|E_2)$ is zero we say that E_1 and E_2 are *mutually exclusive*.

Standard Deviation

There are two different standard deviations we will have use of.

Sample Standard Deviation This is the first standard deviation that we had come across. It can be computed from sample data by the formula (page 196),

$$s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}}.$$

As its name suggests, we use this standard deviation when working with data from a sample. In practice, this is the standard deviation that we most often use.

Population Standard Deviation If we happen to have data about the full population (*not just a sample!*) then the standard deviation is computed slightly differently.

$$\sigma = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n}}$$

The only difference is that we divide by n rather than $n-1$. One instance where we have data about the full population is when we are drawing tickets from a hat. In the hat model all possible outcomes (the entire population) are printed on a ticket in the hat so we should compute σ (and not s) when we compute the standard deviation of a hat.

We can, and usually do, write the equation in terms of the probability model rather than in terms of individual tickets (page 316).

$$\sigma = \sqrt{\sum(x_i - \mu)^2 \cdot p_i}$$

Example 4 The back of a particular scratch-off lottery ticket lists the following prizes and their corresponding chance of winning — \$10,000: 1 in 100,000; \$500: 1 in 5,000; \$5: 1 in 100. What is the probability model, mean, and standard deviation of this game if each ticket costs \$1 to buy? *Do the computation from the perspective of the seller of the ticket.*

We first list the possible outcomes (each way for the player to win less the \$1 ticket cost; also the losing outcome) and their corresponding probabilities. This is the probability model of this game.

outcome	probability
−\$9,999	.00001
−\$499	.0002
−\$4	.01
+\$1	.98979

We compute the winning probabilities by dividing the chances: "1 in 100,000" $= \frac{1}{100{,}000}$. The probability of losing is (1 − probability of winning *any* prize).

The mean (page 313) is computed directly from the probability model.

$$\begin{aligned}\mu &= \sum x_i \cdot p_i \\ &= -\$9{,}999 \cdot .00001 - \$499 \cdot .0002 - \$4 \cdot .01 + \$1 \cdot .98979 \\ &= \$0.75\end{aligned}$$

We compute the standard deviation using a table very similar to the table we used to compute the sample standard deviation.

outcome	$x_i - \mu$	$(x_i - \mu)^2$	$(x_i - \mu)^2 \cdot p_i$
−$9,999	−9999.75	99995000.0625	999.950000625
−$499	−499.75	249740.0676	49.94801352
−$4	−4.75	22.5625	0.225625
+$1	0.25	0.0625	0.061861875
		sum:	1050.18550102
		square root:	32.4065657085

The standard deviation of this probability model is $\sigma \approx \$32$.

Multiple Draws

What does the mean (μ) and standard deviation (σ) tell us if we draw repeatedly from a hat? If we assume that we replace our drawn ticket each time so that the hat is the same each time we draw, we can compute an *expected value of the average* of the draws using the mean. We use the standard deviation to compute a *standard error* that determines how much we should be prepared for our actual average to vary from the expected value. In fact, repeatedly drawing from a hat follows a normal distribution with the expected value as mean and the standard error as standard deviation!

The Background Story

Suppose each student in the class is given a hat containing tickets. The hats are all identical and the mean value of the tickets in any given hat is μ with standard deviation σ.

Each student then goes home, draws 100 times (replacing the ticket after each draw) and records only the average of their 100 draws. Each student has then computed an average ($\bar{x}$) of a sample (100 draws), and each student is likely to have a different average value.

Expected Value of the Average

Each student has their own average of 100 draws. What would we expect that average to be? The average of any number of draws should be close to the mean value of the hat. The Expected Value of the Average is always simply the mean value.

$$\text{EV} = \text{mean}$$

Thus, we expect each student's computed average of 100 draws to be close to the original μ.

Standard Error of the Average

The standard error is a bit more interesting. The standard deviation of the original hat (σ) measures the variation in the values of the individual tickets. The standard error however, is a measure of the variation in the average values computed by the students. The variation in the averages should be smaller than the variation in the tickets since in the average there is some cancellation between large-valued tickets and small-valued tickets. The actual formula for the variation in the averages (Standard Error of the Average) is

$$\text{SE} = \frac{\sigma}{\sqrt{n}}$$

where n is the number of draws made.

Students computing the average of 100 draws should follow a normal distribution with Expected Value μ and Standard Error $\frac{\sigma}{\sqrt{100}} = \frac{\sigma}{10}$.

Example 5 Suppose I poll all of my acquaintances (my population is all acquaintances and I sample the entire population). Each acquaintance places a ticket with a "1" ("yes") or a "0" ("no") on it into a hat. After counting the tickets I find that there are 75 "1" tickets and 45 "0" tickets for a total (population) of 120 tickets.

a. What is the probability model of the hat?

outcome	probability
0	$\frac{45}{120} = 0.375$
1	$\frac{75}{120} = 0.625$

b. What is the mean and standard deviation of the hat?

The mean of the hat is $\mu = \sum x_i \cdot p_i = 0 \cdot 0.375 + 1 \cdot 0.625 = 0.625$. Notice that this is the percentage of students who voted "yes" (1).

outcome	$x_i - \mu$	$(x_i - \mu)^2$	$(x_i - \mu)^2 \cdot p_i$
0	−0.625	0.390625	0.146484375
1	0.375	0.140625	0.087890625
		sum:	0.234375
		square root:	0.484122918276

The standard deviation of this probability model is $\sigma \approx 0.484$. (Note: We will learn a shortcut for computing standard deviations of 0/1 hats later.)

c. What is the expected value and standard error of the average of 50 draws? How can this number be interpreted (other than as an average)?

$$\text{EV} = \mu = 0.625$$
$$\text{SE} = \frac{\sigma}{\sqrt{50}} \approx 0.0684653196882$$

The expected value of the average is the percentage of "1" votes we expect to find in our 50 draws. In the context of the original story, if I were to randomly sample 50 of my acquaintances I would expect 62.5% of those 50 (about 31 people) to have voted "yes". The standard error of the average tells us how much variation we can expect. For example, I would have a 68% chance of 50 draws consisting of between 55.65% and 69.34% "yes" votes (a random sample of 50 people has a 68% chance of having 27 to 35 "yes" voters).

0/1 Hats

In Example 5 the tickets in our hat only had either a "0" or a "1" on them. We had pointed out there that if we consider "1" tickets to be "yes" votes and "0" tickets to be "no" votes the expected values and standard errors could be interpreted as the percentage of supporters.

Standard Deviation of 0/1 Hats

With a little bit of algebra, the standard deviation of a hat (when applied to a 0/1 hat) simplifies to:

$$\sigma = \sqrt{p(1-p)}$$

where p is the percentage of supporters in the hat.

Example 6 In Example 5, 62.5% of the tickets in the hat were supporters. Using the shortcut computation, $\sigma = \sqrt{.625(1-.625)} = \sqrt{.625(.375)} = \sqrt{.234375} \approx .484122918276$.

What is the largest that σ can be for a 0/1 hat? Plug different values of p into the formula to develop intuition (Problem 7).

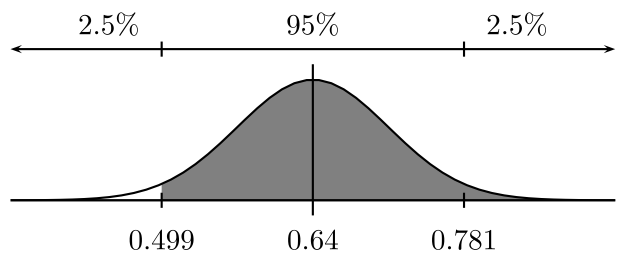
Confidence Intervals

Our goal now is to reverse the direction of our predictions. In Example 5 we made predictions of the form, "Since the *true proportion of supporters* is .625. We can be fairly sure that a draw of 50 tickets will be within a few standard errors of this value." We now want to say something like, "we have made 50 draws and found that 64% of those drawn are supporters. We can be fairly sure that the *true proportion of supporters* lies within a few standard errors of this value." The problem is that if we do not know the true proportion of supporters in the hat we are unable to compute a standard deviation. This is where the question in Example 6 and Problem 7 comes in. The largest possible σ for a 0/1 hat is $\sigma = 0.5$ when $p = .5$. Any other value for p causes σ to be smaller. Using this "worst case scenario" value for the standard deviation allows us to compute what is called a *confidence interval*.

Example 7 Suppose that you draw 50 times from a 0/1 hat. You get 32 "yes" votes out of your 50 draws, but *you do not know the actual percentage of supporters in the hat!* We will compute a 95% confidence interval for the true percentage of supporters in the hat.

If the standard deviation of the hat were .5, then the standard error of the average would be SE Average $= \frac{.5}{\sqrt{50}} \approx 0.07071$.

Thus, we are at least 68% confident that the true percentage of supporters is within 0.07071 of our measured support level of $\frac{32}{50} = 0.64$ and at least 95% confident that the true percentage of supporters is within $2 \cdot 0.07071 = 0.14142$ of our measured support level of 0.64. Therefore, our 95% confidence interval is 0.499 to 0.781. In particular, we can be 97.5% confident (95% plus 2.5% from one tail) that the majority of the people in the hat are supporters.



Problems

1. A common test for AIDS is called the ELISA test. Among 1,000,000 people who are given the test, we can expect results similar to those given in the following table:[1]

	Carry AIDS virus (B_1)	Do not Carry AIDS virus (B_2)	totals
Test positive (A_1)	4,885	73,630	78,515
Test negative (A_2)	115	921,370	921,485
totals	5,000	995,000	1,000,000

[1]Data from "Data and Statistical Inference" by R. Hogg and E. Tanis.

If one of these 1,000,000 people is selected randomly, compute the probabilities: $P(A_1)$, $P(B_1)$, $P(A_1 \mid B_2)$, $P(B_1 \mid A_1)$. In words, what is the meaning of the two conditional probabilities you just computed?

2. The following table classifies 1456 people by their gender and by whether or not they favor a gun law:[2]

	Male (G_1)	Female (G_2)	totals
Favor (A_1)	392	649	1041
Oppose (A_2)	241	174	415
totals	633	823	1456

Is support for the gun law independent of gender?

3. Go to the California Lottery "Scratchers" website (`http://www.calottery.com/Games/Scratchers/`). Choose a ticket and look at the "full odds" of winning (treat any "free ticket" prizes as a prize at the value of the ticket).

   a. What is the "value" of a single ticket?
   b. The retailer has a 99.7% chance of making from $____ to $____ per ticket.
   c. How much would a retailer expect to end up with by selling (and distributing winnings for) 1,000 of these tickets?
   d. How might the answer to part a. affect our treatment of "free ticket" prizes? How much is the value of 1,000 tickets altered by a more realistic estimate of the value of a "free ticket"?

4. a. What is the mean and standard deviation of a hat with $2000 \times \boxed{0}$ and $6000 \times \boxed{1}$?

   b. What is the mean and standard deviation of a hat with $2 \times \boxed{0}$ and $6 \times \boxed{1}$?

   c. If you know that a population consists of 58% supporters, is that enough information to construct a hat to model the population? Why or why not?

5. What is the expected value and standard error of the average of 500 draws from a 0/1 hat if 50% of the *population* are supporters? If 5% of the *population* are supporters?

6. a. If *65% of a population* are supporters what is the expected value and standard error of the average of 100 draws from the 0/1 hat modeling that population?

   b. What is the probability that a sample of 100 people would consist of at least 50% supporters?

7. Using the shortcut formula for the standard deviation of a 0/1 hat, fill in the following table:

supporters	0%	10%	30%	50%	70%	90%	100%
σ							

What two things of significance do you notice about the values of σ?[3]

8. Suppose we choose a sample of 100 people and 60 of those people support a particular statement.

   a. What is the proportion of supporters *in our sample*?

   b. How bad could the standard deviation of the *population* be? (What is the largest possible value of the standard deviation of a 0/1 hat modeling the population? Keep in mind that we do not know the proportion of supporters in the population.)

   c. What would be the corresponding standard error of the average in that case?

   d. Give the 95% and 99.7% confidence intervals.

[2] Ibid.

[3] Hint: The answer is the symmetry and the fact that there is a maximum value for σ.

e. What is our *confidence* that our sample was "good" and correctly predicts that the majority of the population are supporters? (Our confidence that the proportion of supporters in the population is greater than 50%.)[4]

9*. that the true proportion of supporters and standard deviation of a 0/1 hat is not known so that we must use the worst-case estimate $\sigma = .5$.

a. What would be the standard error of the average for a sample of 100 people? What value would we add and subtract to the proportion of supporters, $\hat{p}$, in a sample of 100 people to obtain a 95% confidence interval?

b. How many people would we need to sample so that our standard error was only 3%?

c. How many people would we need to sample so that we are 95% confident that our error is at most 2%? (That is, how many people do we need to sample so that we add and subtract .02 to $\hat{p}$ in order to compute our 95% confidence interval?)

d. Notice that we can control the error of our study simply by altering the size of our sample and that *the size of the population did not come in to play*.

10*. Typically when samples are chosen an individual's vote will not be recorded more than once. However, so far we have assumed that each ticket is replaced in-between each draw in the hat model. Doing so allows an individual's vote to theoretically be recorded more than once. In this problem we explore why this assumption is not typically a problem.

We say that we draw from a hat *with replacement* if after each draw we return the ticket to the hat. Thus, for each draw the hat is in the same state as it was originally. Draws made with replacement are independent of each other and the probability of drawing a particular outcome on the n^{th} draw is the same as the original probability of drawing that outcome.

We say that we draw from a hat *without replacement* if we do not return each ticket to the hat after each draw. Thus, for each draw the hat is in a different state than it was originally. Draws made without replacement are not independent of each other and the probability of drawing a particular outcome on the n^{th} draw must be computed from the tickets which are actually still left in the hat when the n^{th}draw is made.

For the following problems we will use the following two hats:

The Small Hat	**The Big Hat**
$\boxed{0}\,\boxed{0}\,\boxed{0}\,\boxed{1}\,\boxed{1}\,\boxed{1}\,\boxed{1}$	$3000 \times \boxed{0},\ 4000 \times \boxed{1}$

Recall that $P(E_1 \cap E_2) = P(E_1) \cdot P(E_2 \,|\, E_1)$. We will use the shortcut $P(\boxed{0}\boxed{1})$ to mean $P(\text{draw } \boxed{0} \text{ first and then draw } \boxed{1} \text{ second})$.

a. What is $P(\boxed{1}\boxed{1})$ and $P(\boxed{1}\boxed{1}\boxed{1})$ when we draw from the small hat with replacement?

b. What is $P(\boxed{1}\boxed{1})$ and $P(\boxed{1}\boxed{1}\boxed{1})$ when we draw from the large hat with replacement?

c. What is $P(\boxed{1}\boxed{1})$ and $P(\boxed{1}\boxed{1}\boxed{1})$ when we draw from the small hat without replacement?

d. What is $P(\boxed{1}\boxed{1})$ and $P(\boxed{1}\boxed{1}\boxed{1})$ when we draw from the large hat without replacement?

---

[4]Note: Our use of the term confidence, rather than probability, is intentional since the proportion of supporters in the population is a fixed and *determinable* quantity. Since the true proportion of supporters is not a random quantity it can not have a probability of being in a particular range — it either is in that range or it isn't. The confidence percentage that you compute is a lower bound for the probability *before the sample was drawn* of choosing a sample in which the supporters form a majority. Thus, we are computing our confidence in an arbitrary sample being a good representative of the population. The Central Limit Theorem tells us that the confidence can be computed using $SE = \frac{\sigma}{\sqrt{n}}$ and the 68/95/99.7 rule.

e. What does this tell us about replacement and the total number of tickets in the hat? Under what conditions must we be concerned whether we are drawing with or without replacement? How does this interact with the observation made in problem 9d?