

# Using Probability to do Statistics.

Al Nosedal.  
University of Toronto.

November 5, 2015

# Milk and honey and hemoglobin

- Animal experiments suggested that honey in a diet might raise hemoglobin level.
- A researcher designed a study involving four pairs of twins in a children's home.
- The researcher then observed which twin in each pair had the greater increase in hemoglobin.

# Conditional Probability

The conditional probability of  $B$ , given  $A$ , denoted by  $P(B|A)$ , is defined by

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \quad \text{if } P(A) > 0.$$

# Law of Total probability

If the events  $B_1, B_2, \dots, B_k$  constitute a partition of the sample space  $S$  such that  $P(B_i) \neq 0$  for  $i = 1, 2, \dots, k$ , then for any event  $A$  of  $S$ ,

$$P(A) = P(B_1 \cap A) + P(B_2 \cap A) + \dots + P(B_k \cap A)$$

which is equivalent to

$$P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \dots + P(A|B_k)P(B_k).$$

Assume that we have an urn that contains a "big" number of chips. We are told that the chips placed in this urn come in one of two colors, white or blue. Furthermore, we are informed that we are dealing with one of three possible scenarios:

- $B_1 =$  All chips are blue.
- $B_2 =$  Half of these chips are blue and the other half are white.
- $B_3 =$  All of them are white.

Suppose we are interested in assigning probabilities to each of these three possible scenarios.

# Prior probabilities

We will assign a probability to each of these three scenarios, according to our degree of belief. Given that we don't have any further information:

- $P(B_1) = 1/3$ .
- $P(B_2) = 1/3$ .
- $P(B_3) = 1/3$ .

These values are referred to as "a priori" probabilities and they form an "a priori" distribution for the random variable we are interested in. A priori probabilities are called like that due to the fact that they are determined prior to obtaining empirical evidence (data).

# Conditional Probabilities

Now, suppose that we are allowed to draw, randomly, one chip from the urn in question and the chip drawn turns out to be blue. Let  $D$  (for data) be the event "the chip drawn is blue". To incorporate  $D$  into our solution we need to find the following conditional probabilities:

$$P(B_1|D) = \frac{P(B_1 \cap D)}{P(D)}$$

$$P(B_2|D) = \frac{P(B_2 \cap D)}{P(D)}$$

$$P(B_3|D) = \frac{P(B_3 \cap D)}{P(D)}$$

$$P(D|B_1) = 1$$

$$P(D|B_2) = \frac{1}{2}$$

$$P(D|B_3) = 0$$



$$P(B_1 \cap D) = P(D|B_1)P(B_1) = (1)(1/3) = 1/3$$

$$P(B_2 \cap D) = P(D|B_2)P(B_2) = (1/2)(1/3) = 1/6$$

$$P(B_3 \cap D) = P(D|B_3)P(B_3) = (0)(1/3) = 0$$

Applying the Law of Total Probability,

$$\begin{aligned}P(D) &= P(B_1 \cap D) + P(B_2 \cap D) + P(B_3 \cap D) \\&= 1/3 + 1/6 + 0 \\&= 1/2.\end{aligned}$$

**P(D)** is sometimes referred to as the **predictive probability**.

Finally, we have:

$$P(B_1|D) = \frac{P(B_1 \cap D)}{P(D)} = \frac{1/3}{1/2} = 2/3$$

$$P(B_2|D) = \frac{P(B_2 \cap D)}{P(D)} = \frac{1/6}{1/2} = 1/3$$

$$P(B_3|D) = \frac{P(B_3 \cap D)}{P(D)} = \frac{0}{1/2} = 0$$

## Another way of finding posterior distribution

| Scenario | Prior $P(B_i)$ | $P(D B_i)$ | $P(D \cap B_i)$ | Posterior $P(B_i D)$ |
|----------|----------------|------------|-----------------|----------------------|
| $B_1$    | $1/3$          | 1          | $1/3$           | $2/3$                |
| $B_2$    | $1/3$          | $1/2$      | $1/6$           | $1/3$                |
| $B_3$    | $1/3$          | 0          | 0               | 0                    |
|          | 1              |            | $1/2$           | 1                    |

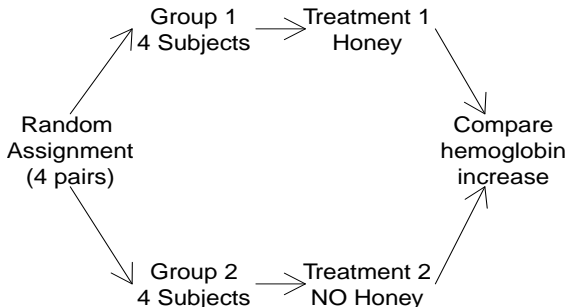
# Updating posterior distribution

Now, assume that we are allowed to draw a second chip and, again, the chip drawn is blue; What are the probabilities for the remaining two possible scenarios?

# Posterior probability after drawing a second chip

| Scenario | Prior $P(B_i)$ | $P(D B_i)$ | $P(D \cap B_i)$ | Posterior $P(B_i D)$ |
|----------|----------------|------------|-----------------|----------------------|
| $B_1$    | $2/3$          | 1          | $2/3$           | $4/5$                |
| $B_2$    | $1/3$          | $1/2$      | $1/6$           | $1/5$                |
|          | 1              |            | $5/6$           | 1                    |

# Milk and honey and hemoglobin



- $B_0$  = NONE of the honey twins have a greater hemoglobin increase.
- $B_1$  = 25% of honey twins have a greater hemoglobin increase.
- $B_2$  = 50% of honey twins have a greater hemoglobin increase.
- $B_3$  = 75% of honey twins have a greater hemoglobin increase.
- $B_4$  = ALL honey twins have a greater hemoglobin increase.



- a) What is the researcher's posterior probability of  $B_2$ ? (which can be interpreted as honey having no effect on hemoglobin level.)
- b) Consider a fifth set of twins (one receives honey in milk and the other receives only milk). Find the researcher's predictive probability that the honey twin's hemoglobin will have a greater increase.

The researcher recorded the following increases in hemoglobin:

| Pair     | 1  | 2  | 3 | 4  |
|----------|----|----|---|----|
| Honey    | 19 | 12 | 9 | 17 |
| No honey | 14 | 8  | 4 | 4  |

# Rethinking our Data

|       |      |      |      |      |
|-------|------|------|------|------|
| Chip  | 1    | 2    | 3    | 4    |
| Color | Blue | Blue | Blue | Blue |

# First posterior

| Scenario | Prior | Likelihood | Prior $\times$ Likelihood | Posterior |
|----------|-------|------------|---------------------------|-----------|
| $B_0$    | 1/5   | 0          | 0                         | 0         |
| $B_1$    | 1/5   | 1/4        | 1/20                      | 1/10      |
| $B_2$    | 1/5   | 2/4        | 2/20                      | 2/10      |
| $B_3$    | 1/5   | 3/4        | 3/20                      | 3/10      |
| $B_4$    | 1/5   | 4/4        | 4/20                      | 4/10      |
|          | 1     |            | 10/20                     | 1         |

## Second posterior

| Scenario | Prior | Likelihood | Prior $\times$ Likelihood | Posterior |
|----------|-------|------------|---------------------------|-----------|
| $B_1$    | 1/10  | 1/4        | 1/40                      | 1/30      |
| $B_2$    | 2/10  | 2/4        | 4/40                      | 4/30      |
| $B_3$    | 3/10  | 3/4        | 9/40                      | 9/30      |
| $B_4$    | 4/10  | 4/4        | 16/40                     | 16/30     |
|          | 1     |            | 30/40                     | 1         |

# Third posterior

| Scenario | Prior | Likelihood | Prior $\times$ Likelihood | Posterior |
|----------|-------|------------|---------------------------|-----------|
| $B_1$    | 1/30  | 1/4        | 1/120                     | 1/100     |
| $B_2$    | 4/30  | 2/4        | 8/120                     | 8/100     |
| $B_3$    | 9/30  | 3/4        | 27/120                    | 27/100    |
| $B_4$    | 16/30 | 4/4        | 64/120                    | 64/100    |
|          | 1     |            | 100/120                   | 1         |

# Final posterior

| Scenario | Prior  | Likelihood | Prior $\times$ Likelihood | Posterior |
|----------|--------|------------|---------------------------|-----------|
| $B_1$    | 1/100  | 1/4        | 1/400                     | 1/354     |
| $B_2$    | 8/100  | 2/4        | 16/400                    | 16/354    |
| $B_3$    | 27/100 | 3/4        | 81/400                    | 81/354    |
| $B_4$    | 64/100 | 4/4        | 256/400                   | 256/354   |
|          | 1      |            | 354/400                   | 1         |

## a) Solution

a) **Posterior probability of  $B_2 = 16/354 = 0.0452$**



## b) Solution

| Scenario | Prior   | Likelihood | Prior $\times$ Likelihood |
|----------|---------|------------|---------------------------|
| $B_1$    | 1/354   | 1/4        | 1/1416                    |
| $B_2$    | 16/354  | 2/4        | 32/1416                   |
| $B_3$    | 81/354  | 3/4        | 243/1416                  |
| $B_4$    | 256/354 | 4/4        | 1024/1416                 |
|          | 1       |            | 0.9180791                 |

**b) Predictive probability that a 'new' honey twin's hemoglobin will have a greater increase =  $1300/1416 = 0.9180791$**

**We have discussed the basic elements of Bayesian Statistics.**

- Prior Distribution
- Likelihood
- Predictive Probability
- Posterior Distribution

# Bayes' Rule

If the events  $B_1, B_2, \dots, B_k$  constitute a partition of the sample space  $\mathbf{S}$ , where  $P(B_i) \neq 0$  for  $i = 1, 2, \dots, k$ , then for any event  $A$  in  $S$  such that  $P(A) \neq 0$ ,

$$P(B_r|A) = \frac{P(B_r \cap A)}{P(B_1 \cap A) + P(B_2 \cap A) + \dots + P(B_k \cap A)}$$

which is equivalent to

$$P(B_r|A) = \frac{P(A|B_r)P(B_r)}{P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \dots + P(A|B_k)P(B_k)}$$

for  $r = 1, 2, \dots, k$ .

Binomial distribution with parameters  $n$  and  $p$ .

$$f(y|p) = \frac{n!}{y!(n-y)!} p^y (1-p)^{n-y}$$

for  $y = 0, 1, \dots, n$ .

# Our likelihoods

- $f(y = 4|p = 0, n = 4) = \frac{4!}{4!0!}(0)^4(1)^0 = 0.$
- $f(y = 4|p = 0.25, n = 4) = \frac{4!}{4!0!}(0.25)^4(0.75)^0 = \left(\frac{1}{4}\right)^4 = \frac{1}{256}.$
- $f(y = 4|p = 0.5, n = 4) = \frac{4!}{4!0!}(0.5)^4(0.5)^0 = \left(\frac{1}{2}\right)^4 = \frac{16}{256}.$
- $f(y = 4|p = 0.75, n = 4) = \frac{4!}{4!0!}(0.75)^4(0.25)^0 = \left(\frac{3}{4}\right)^4 = \frac{81}{256}.$
- $f(y = 4|p = 1, n = 4) = \frac{4!}{4!0!}(1)^4(0)^0 = 1 = \frac{256}{256}.$

# Posterior Probabilities

| Scenario | Prior | Likelihood | Prior $\times$ Likelihood | Posterior |
|----------|-------|------------|---------------------------|-----------|
| $B_0$    | 1/5   | 0          | 0                         | 0         |
| $B_1$    | 1/5   | 1/256      | 1/1280                    | 1/354     |
| $B_2$    | 1/5   | 16/256     | 16/1280                   | 16/354    |
| $B_3$    | 1/5   | 81/256     | 81/1280                   | 81/354    |
| $B_4$    | 1/5   | 256/256    | 256/1280                  | 256/354   |
|          | 1     |            | 354/1280                  | 1         |

Exactly the same probabilities that we obtained using our sequential approach.

# Definiton of mean

Let  $X$  be a random variable with probability distribution  $f(x)$ . The **mean** or **expected value** of  $X$  is

$$\mu = E(X) = \sum_x xf(x)$$

if  $X$  is discrete, and

$$\mu = E(X) = \int_{-\infty}^{\infty} xf(x)dx$$

if  $X$  is continuous.

## Another way of finding predictive probability

| Scenario | Model       | Posterior | Model $\times$ Posterior |
|----------|-------------|-----------|--------------------------|
| $B_0$    | $p_0 = 0$   | 0         | 0                        |
| $B_1$    | $p_1 = 1/4$ | 1/354     | 1/1416                   |
| $B_2$    | $p_2 = 2/4$ | 16/354    | 32/1416                  |
| $B_3$    | $p_3 = 3/4$ | 81/354    | 243/1416                 |
| $B_4$    | $p_4 = 4/4$ | 256/354   | 1024/1416                |
|          |             |           | 1300/1416                |

Posterior mean =  $\frac{1300}{1416} = 0.9180791$  (Exactly the same number we obtained before).



# Using a continuous uniform prior

Prior:  $f(p) = 1$  for  $0 \leq p \leq 1$ .

Likelihood:  $f(y|p) = \frac{n!}{y!(n-y)!} p^y (1-p)^{n-y}$  for  $y = 0, 1, \dots, n$ .

Product = Prior  $\times$  Likelihood =  $\frac{n!}{y!(n-y)!} p^y (1-p)^{n-y}$

# Product in our particular case

$$\text{Product} = \text{Prior} \times \text{Likelihood} = \frac{n!}{y!(n-y)!} p^y (1-p)^{n-y}$$

$$\text{Product} = \frac{4!}{4!(4-4)!} p^4 (1-p)^0 = p^4.$$

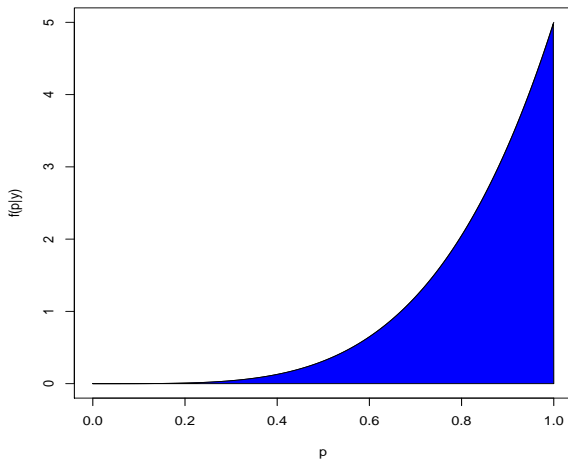
## Finding "sum" of all products

$$\int_0^1 p^4 dp = \frac{(1)^5}{5} - \frac{(0)^5}{5} = \frac{1}{5}.$$

Recall that our posterior distribution is given by:

$$\text{Posterior} = (\text{Prior} \times \text{Likelihood}) / (1/5) = 5p^4$$

# Graph of posterior distribution



$$E(p|data) = \int_0^1 p f(p|data) dp.$$

$$E(p|data) = \int_0^1 p(5p^4) dp = 5 \int_0^1 p^5 dp = 5 \left( \frac{1}{6} \right) = \frac{5}{6} = 0.8333$$

Posterior mean = Predictive probability = 83.33%.

On March 15, 2002, the Iowa City Press Citizen carried an article about the intended 19% tuition increase to go into effect at the University of Iowa (UI) for the next academic year.

Your research question is as follows: What is the unknown population parameter  $\pi$  - the proportion in the entire population of UI students who would be likely to quit school if tuition is raised 19%?

# Problem

You do not have the time or resources to locate and interview all 28,000+ students, so you cannot evaluate  $\pi$  exactly. Instead, you will pick a simple random sample of  $n = 50$  students from the student directory and ask each of them whether she or he would be likely to quit school if tuition were raised by 19%. You wish to use your sample data to estimate the population proportion  $\pi$  and to determine the amount of uncertainty in your estimate.

Define a random variable  $Y$  as the count of the number of successes in your sample.  $Y$  meets the definition of a binomial random variable - it is the count of the number of successes in  $n$ -independent Bernoulli trials, all with the same success probability. We can write

$$Y \sim \text{Binomial}(n, \pi).$$



If we knew  $\pi$ , we could use the binomial probability mass function to compute the probability of obtaining any one of the possible values  $y$  that the random variable  $Y$  could take on in our sample:

$$P(y|\pi) = \binom{n}{y} \pi^y (1 - \pi)^{n-y}, \quad y = 0, 1, 2, \dots, n.$$

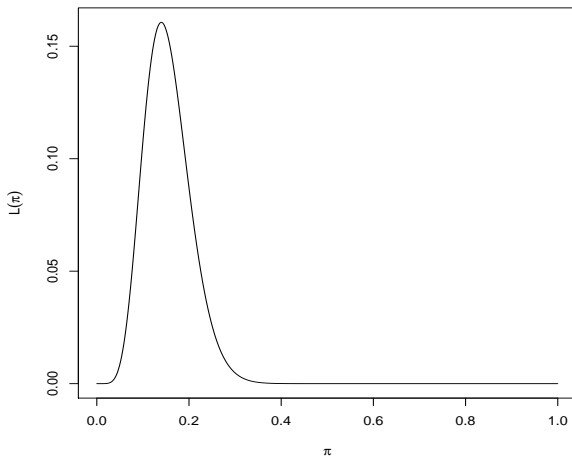
But we don't know  $\pi$ .

Instead, after you interview the 50 students, you will know how many said yes. That is, you will know which value  $y$  the random variable  $Y$  actually took on. Suppose this number turns out to be  $y = 7$ . We will want to use this information to estimate  $\pi$ . In this case, we may change perspective and regard the expression in  $P(y|\pi)$  as a function of the unknown parameter  $\pi$  given the now known (fixed) data value  $y$ . When viewed in this way, the expression is called the **likelihood function**. If  $y$  were 7, it would look like

$$L(\pi) = \binom{50}{7} \pi^7 (1 - \pi)^{43}, \quad 0 < \pi < 1.$$

We could compute this likelihood for different values of  $\pi$ . Intuitively, values of  $\pi$  that give larger likelihood evaluations are more consistent with the observed data.

# Likelihood



Now we are interested in evaluating the likelihood  $L(\pi)$  for changing values of  $\pi$ , not of  $y$ . Consequently, the  $\binom{50}{7}$  is now considered just a constant, and only the  $\pi^y(1 - \pi)^{43}$ , which varies with  $\pi$ , is the **kernel**. That is, we can write

$$L(\pi) \propto \pi^7(1 - \pi)^{43}, \quad 0 < \pi < 1.$$

or more generically

$$L(\pi) \propto \pi^y(1 - \pi)^{n-y}, \quad 0 < \pi < 1.$$

# The prior

To carry out a Bayesian analysis to learn about the unknown population proportion  $\pi$ , we need to assess our previous knowledge or belief about  $\pi$  before we observe the data from the survey. The Bayesian approach to expressing prior knowledge about a population parameter is to put a probability distribution on the parameter - that is, to treat the unknown population parameter as if it were a random variable.

# The prior

Because it is a proportion, the parameter  $\pi$  hypothetically could take on any value in the interval  $(0, 1)$ , although most of us realize that some ranges of values are much more likely than others. Because  $\pi$  can take on any of a continuum of values, we quantify our knowledge or belief most appropriately by means of a probability density function.

# The prior

A person who has little or no knowledge about university students might consider all values in  $(0, 1)$  equally plausible before seeing any data. A uniform density on  $(0,1)$  describes this belief (or state of ignorance!) mathematically

$$\pi \sim U(0, 1)$$

$$P(\pi) = 1, 0 < \pi < 1.$$

This continuous uniform distribution is called a "vague" or "noninformative" prior.

# Other Possible Prior Distributions

If a person has knowledge or belief regarding the value of  $\pi$ , his or her prior will be informative.



# Using the Data to Update the Prior: The Posterior Distribution

The possible ways of choosing a prior for an unknown proportion are endless. For the moment, let's see what happens if we use the "noninformative" continuous uniform prior for our analysis.

At last, you may select your sample and conduct your survey! It turns out that 7 of the 50 students say they would quit school if tuition were raised 19%. Thus, the sample proportion of yesses in your observed data is  $\hat{\pi} = \frac{7}{50} = 0.14$  Note that this is the value of  $\pi$  at which the likelihood function in  $L(\pi)$  attained its maximum.

For the quitting-school example and binomial likelihood, combining the prior density on  $\pi$  with the likelihood yields

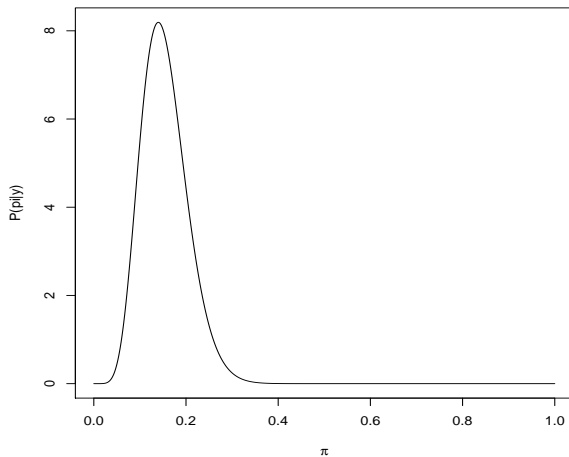
$$P(y|\pi) \propto p(\pi)\pi^y(1 - \pi)^{n-y}, \quad 0 < \pi < 1.$$

Having chosen the uniform prior,  $p(\pi) = 1$ ,  $0 < \pi < 1$ , and having observed  $y = 7$  "successes" out of  $n = 50$  people surveyed

$$P(\pi|y) \propto 1 \times \pi^7(1 - \pi)^{43}, \quad 0 < \pi < 1.$$

We can graph this function to see what the posterior density  $p(\pi|y)$  looks like

# Posterior



Note that the mode (highest peak) is at  $\pi = 0.14$ , and most of the area under the curve is above values of  $\pi$  in the interval (0.05, 0.35).

# Conjugate priors

A common way to construct a prior distribution is to designate that the prior is a member of a particular parametric family of densities. One then chooses the parameters of the prior density so as to reflect as closely as possible his or her beliefs about the unknown parameter being studied. When possible, it is very convenient analytically to choose the prior from a parametric family that has the same functional form as the likelihood function. Such a prior is called a conjugate prior.

A beta family of densities, with fixed parameters  $\alpha$  and  $\beta$  and with the random variable called  $\pi$  would be written as follows:

$$\pi \sim \text{Beta}(\alpha, \beta).$$

or

$$p(\pi) \propto \pi^{\alpha-1}(1-\pi)^{\beta-1}, \quad 0 < \pi < 1.$$

The kernel of the beta density is  $\pi^{\alpha-1}(1-\pi)^{\beta-1}$ .

# Computing the Posterior Distribution with a Conjugate Prior

Recall the relationship of the posterior distribution to the prior and likelihood:

$$P(y|\pi) \propto p(\pi)L(\pi; y).$$



# Computing the Posterior Distribution

So in the case of a beta prior and a binomial likelihood,

$$P(\pi|y) \propto \pi^{\alpha+y-1}(1-\pi)^{\beta+n-y-1}$$

This is the kernel of another beta density!

Since the beta family of priors is conjugate for the binomial likelihood, the posterior distribution is also beta - a member of the same parametric family as the prior distribution.

# Choosing the Parameters of a Beta Distribution to Match Prior Beliefs

Here are several ways to think about choosing the parameters of a beta distribution to express prior beliefs or knowledge about an unknown proportion:

Strategy 1: Graph some beta densities until you find one that matches your beliefs.

Strategy 2: Note that a  $Beta(\alpha, \beta)$  prior is equivalent to the information contained in a previously observed dataset with  $\alpha - 1$  successes and  $\beta - 1$  failures.

# Choosing the Parameters of a Beta Distribution to Match Prior Beliefs

Strategy 3: Solve for the values of  $\alpha$  and  $\beta$  that yield:

a) The desired mean (The mean of a  $Beta(\alpha, \beta)$  density is  $\frac{\alpha}{\alpha + \beta}$ ).

b) The desired equivalent prior sample size, which for a  $Beta(\alpha, \beta)$  prior is  $\alpha + \beta - 2$ . When you use this method, you are saying that your knowledge about  $\pi$  is as strong as if you'd seen a previous sample consisting of  $\alpha - 1$  successes and  $\beta - 1$  failures.

Strategy 4: Choose values of  $\alpha$  and  $\beta$  that produce a prior probability interval that reflects your belief about  $\pi$ .

We will apply the four strategies to the quitting-school-because-of-rising tuition example. We are attempting to construct a reasonable prior before we see the results of the actual survey of 50 UI students. (Forget about what we said earlier - you have not yet collected your data!)

We wish to use any relevant data available before we do our survey. Suppose that we read that such a survey has already been taken at Iowa State University in which:

- 50 students were interviewed.
- 10 said they would quit school; 40 said they would not.

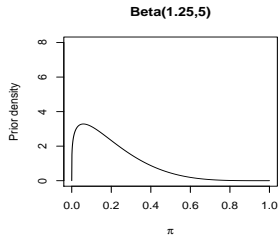
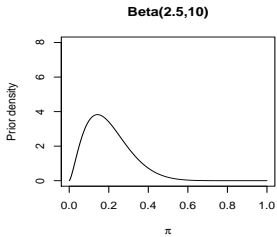
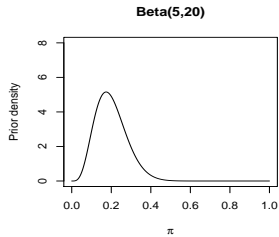
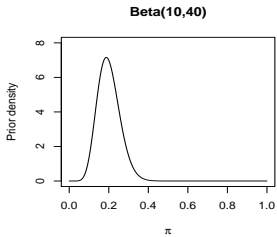
By strategy 2, this might suggest a  $\text{beta}(11, 41)$  prior. However, we need to be very cautious here because the sample on which the prior distribution is to be based was not drawn from the same population (UI students) in which we are interested and from which we will draw our sample. The question is whether ISU students might be different from UI students in ways that would be likely to affect their probability of dropping out in response to a tuition increase - that is, is a sample of ISU students exchangeable with a sample of UI student for our purposes?

Some relevant facts are that the UI has a medical school and a law school, which ISU does not, while ISU has a veterinary school and a College of Agriculture, which UI does not. At the UI, the majority of the student body (53.5%) are women, whereas at ISU, only 43.5% of students are women. Undergraduate tuition is a bit higher at the UI (\$6,544 for Iowa residents, \$20,658 for nonresidents) than at ISU (\$6,360 and \$17,350) for the 2008 - 2009 year. These and other factors suggest that, for drawing inference about UI students, the information in a sample of ISU students is not equivalent to the information contained in a sample of the same number of UI students. Thus, the  $\text{beta}(11,41)$  prior most likely is not appropriate.

On the other hand, there presumably is some relevant information for us in the ISU survey. We want to make use of that, but give it less weight than we would give to 50 observations from the UI population. One of many valid approaches to specifying a prior in this case is to say that we want a prior mean of 0.2, the same as the sample proportion  $\hat{\pi}_{ISU}$  from the ISU data, but an "equivalent prior sample size" (remember, for a beta prior that is  $\alpha - 1 + \beta - 1$ ) that is smaller than 50. One possibility is to look at the graphs of several different beta distributions, all with the same mean 0.2 but with smaller and smaller equivalent prior sample sizes, and seek one that matches our beliefs.



# Priors



```
# R Code;  
  
p=seq(0,1,by=0.001);  
  
prior1=dbeta(p,10,40);  
  
prior2=dbeta(p,5,20);  
  
prior3=dbeta(p,2.5,10);  
  
prior4=dbeta(p,1.25,5);
```

```
# R Code;
```

```
plot(p,prior1,type="l",xlab=expression(pi),  
ylab="Prior density",ylim=c(0,8));  
title("Beta(10,40)");
```

```
plot(p,prior2,type="l",xlab=expression(pi),  
ylab="Prior density",ylim=c(0,8));  
title("Beta(5,20)");
```

```
plot(p,prior3,type="l",xlab=expression(pi),  
ylab="Prior density",ylim=c(0,8));  
title("Beta(2.5,10)");
```

```
plot(p,prior4,type="l",xlab=expression(pi),  
ylab="Prior density",ylim=c(0,8));  
title("Beta(1.25,5)");
```

We can also consider whether the central prior intervals produced by any of these prior densities match our knowledge.

```
# R Code;
```

```
interval.1=c(qbeta(0.025,10,40),qbeta(0.975,10,40));
```

```
interval.1
```

```
interval.2=c(qbeta(0.025,5,20),qbeta(0.975,5,20));
```

```
interval.2
```

```
interval.3=c(qbeta(0.025,2.5,10),qbeta(0.975,2.5,10));
```

```
interval.3
```

```
interval.4=c(qbeta(0.025,1.25,5),qbeta(0.975,1.25,5));
```

```
interval.4
```

# Computing and Graphing the Posterior Distribution

Suppose you chose the Beta(10,40) prior because it best represented your beliefs. You then gathered your own data on  $n = 50$  UI students, and you got  $y = 7$  "successes" and  $n - y = 43$  "failures." Then your posterior distribution of  $\pi$  given your Beta prior and the new data is

$$P(\pi|y) \propto \pi^{17-1}(1 - \pi)^{83-1}$$

This is the kernel of a Beta(17, 83) density.

```
# R Code;

p=seq(0,1,by=0.001);

posterior=dbeta(p,17,83);

plot(p,posterior,type="l",xlab=expression(pi),
ylab="Posterior density");
title("Beta(17,83)");

interval=c(qbeta(0.025,17,83),qbeta(0.975,17,83));

interval
```



### Beta(17,83)

