LECTURE OUTLINE
Choosing Beta Priors

Professor Leibon

Math 10

MAY 2005

The $beta(a, b)$ random variable: Fact Review

1. a and b are positive numbers.
2. $rbeta(N, a, b)$ produces N random $beta(a, b)$-numbers.
3. $dbeta(x, a, b)$ is a formula for the curve that describes the shape of the $beta(a, b)$ random variable.
4. $pbeta(x, a, b) = P(rbeta(1, a, b) \leq x)$.
5. The mean of $beta(a, b)$ is $r = \frac{a}{a+b}$.
6. The standard deviation is $t = \sqrt{r(r^+ - r)}$ where $r^+ = \frac{a+1}{a+b+1}$.
7. If we use $\beta(a, b)$ as a prior and witness s successes and f failures, then the posterior (updated prior) is $\beta(a + s, b + f)$. In particular is has mean $\frac{a+s}{a+s+b+f}$ and can be interpreted as the *predictive probability* of success on the **next** observation.

When choosing $beta(a, b)$ *as a prior keep in mind:*

1. Open Minded Priors: Smaller values of a and b correspond to more open minded priors.
2. Choosing a beta distribution will require two *facts* (opinions).
3. Often, one *fact* is the mean probability of success.
4. Other *facts* can be:
   (a) How the mean probability of success would be effected by witnessing s success and f failures.
   (b) The probability of a certain event.
   (c) By fixing a (probably open minded) standard deviation.
5. If you have more than two *facts*, then you can do *consistency checks*. Consistency checks may help you choose better a and b or they may force you to abandon the notion that a $beta$ distribution is a good choice for your prior.

Example

Here we are viewing a success as *Black/Green* arising when we spin our Roulette wheel.

1. Suppose you'd like to use a beta prior and believe that the mean probability of success should be $20/38$. Further suppose you would like to be open mined enough to allow a $1/6$ chance that the actual probability of success is less than or equal to 50 %. Find good choices for a and b. (**Answer:** around $a = 180$ and $b = 162$ using $pbeta$).

2. Using your prior, compute the probability that the actual probability of success is greater than than or equal to 55 %. (**Answer:** 0.19) Is this answer acceptable to you?

3. What is the probability of $Black/Green$ arising on the next spin? (**Answer:** $\frac{20}{38} \approx 0.526$)

4. In this setting, what should we call the *null* hypothesis and why? How could we test whether or not the null hypothesis is supported?

Normal Approximation

For a and b "not moderately small"

$$rbeta(1, a, b) \approx t(rnorm(1)) + r.$$

and

$$rnorm(x) \approx \frac{rbeta(1, a, b) - r}{t}.$$

where $norm$ is the *standard normal* random variable. Notice that $pnorm(1)$ is computed on the opening pages of the book (if you don't like using R). This allows to replace questions concerning $beta$ with question concerning $norm$. You should try and redo the example using this approximation.

Probability Intervals

Fix a probability $Prob$. If c is the number such that

$$P(r - c \leq rbeta(1, a, b) \leq r + c) = Prob,$$

then the interval $r \pm c$ is called the $100(Prob)$ *% probability interval*.

If you assume that you start with the "no knowledge" prior $beta(1, 1)$ and run an experiment, then the posterior is $beta(1 + s, 1 + f)$ and the $100(Prob)$ *% probability interval* is called the $100(Prob)$ *% confidence interval*.

Example Continued

Suppose we played Roulette 20 times and that $Black/Green$ occurred only 3 times.

5. What is the probability of $Black/Green$ arising on the next spin? (The predictive probability). ($\frac{180+3}{162+17} \approx 0.508$)

6. What is the 95 % probability interval associated to the probability from part 5? (Using $pbeta$ 0.508 ± 0.0515. Do this **two** ways!). Compare it with the 95 % confidence interval (using $pbet$, we find $0.19 \pm .151$).

7. Does this data support the null hypothesis or can we safely "reject" the null hypothesis.?

A Hypothesis of a Proportion (A Two Sided Test)

Choices: Fix a *Null* hypothesis represented by a proportion p_n, choose of probability α (called the *level of significance*), a prior, and a number of trials N.

Find The Probability Interval: Run your experiment and find your posterior. Compute the $100(1-\alpha)$ % probability interval about your posterior's mean, $r \pm c$.

Interpreting the Test: If p_n is **not** in $r \pm c$, then you *reject the null hypothesis*. If p_n is in $r \pm c$, then you would **not** reject the Null Hypothesis and (might) say that the Null hypothesis is supported. (It is at least consistent with the experiment that you performed, for small c it is supported.)

Example Continued

8. Consider a hypothesis test to test our *Black/Green* bet in the spirit of the previous slide. Suppose $p_n = \frac{20}{38}$, $\alpha = 0.95$, and our prior is $beta(180, 162)$. If I perform 100 trials, then for what number of success will I reject the Null hypothesis? This is called the test's *critical region*. (using $pnorm$ we find ≤ 32 and ≥ 73 will do).

9. How big will N need to be so that if $s/N < .45$, then I will reject the Null hypothesis? (using $pnorm$, $N = 334$) This is an example of assessing our test's *Power*.

10. Design a test of the *Black/Green* bet in the spirit of the previous slide. (You will have to Choose α, p_n, a prior, and N.) What factors will you have to consider?

A solution to 8

Solution to 8: After s success we find our posterior equal to $beta(180+s, 162+100-s)$. We need to find the largest s so that of 95 % probablity interval is smaller that p_n and the smallest s so that of 95 % probablity interval is larger that p_n. You should draw a picture! One way to accomplish this is to use the normal approximation. To find our smaller s, we will need to solve $r + z_{95}t = p_n$ while to find our larger s we will need to solve $r - 1.96t = p_n$. Here $r = \frac{180+s}{162+180+100}$ and $r^+ = \frac{180+s+1}{162+180+100+1}$ and $t = \sqrt{r(r^+ - r)}$. So in both, cases we need a solution to the quadratic,

$$r(r^+ - r) = \left(\frac{p_n - r}{z_{95}}\right)^2$$

after substituting our values, as a function of s we find that we are solving the quadratic.

$$(0.1344)s^2 - 14.12s + 314.6 = 0.$$

Using our trusty quadratic equation, we find the solutions 72.99 and 32.07. So s must satisfy either $s \leq 32$ or $s \geq 73$.

A solution to 9

How big will N need to be so that if $s/N < .45$, then I will reject the Null hypothesis? (using $pnorm$, $N = 334$) This is an example of assessing our test's *Power*.

Solution to 9: Using the normal approximation, any s that works we will need to satisfy $r + z_{95}t \leq p_n$. So we need to find the smallest N so that $r + z_{95}t \leq p_n$. Here $s \approx \frac{9}{20}N$ the r and t come from the posterior $beta(180 + \frac{9}{20}N, 162 + \frac{11}{20}N)$. Once a again we need to solve $r + z_{95}t = p_n$, but now viewed as a function of N. Well we proceed as in the previous problem, but we find a cubic polynomial N

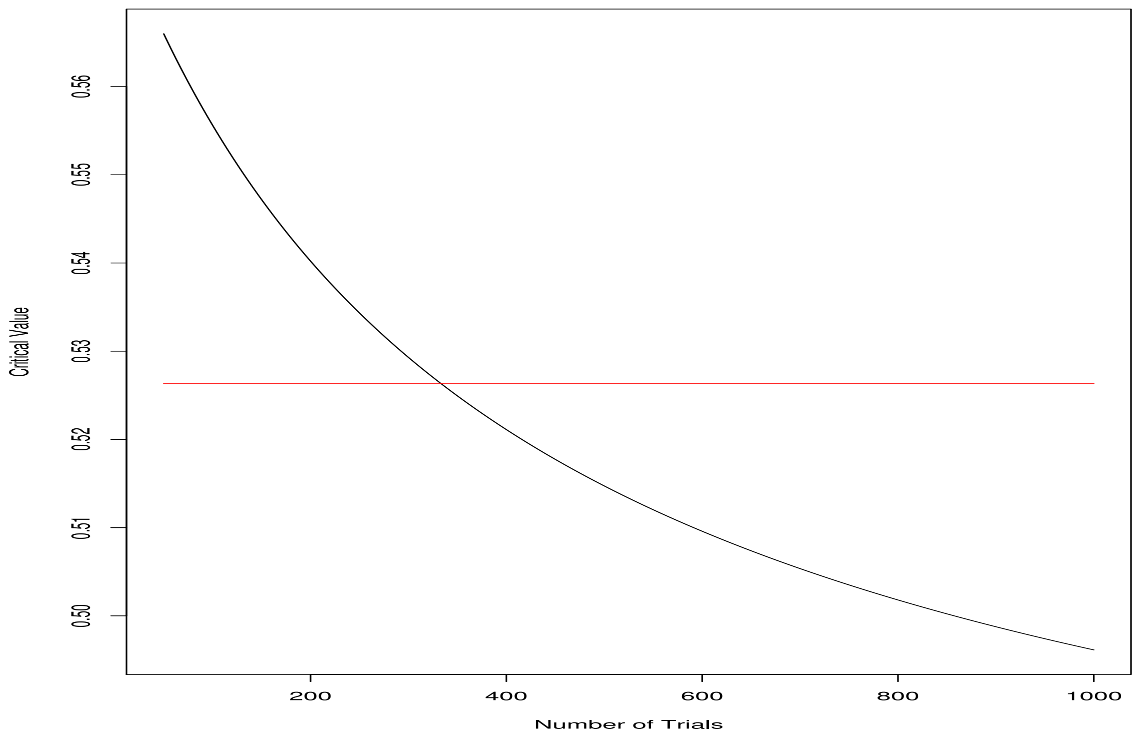
$$0.0002189N^3 + 0.03935N^2 - 24.82N - 4211 = 0.$$

This is of course very frightening since we don't remember (and possibly have never seen) the formulas for solving a cubic! If we did we would find the roots -346.8, -166.4, and 333.4. So our ours must correspond to the positive root and we need $N \geq 334$.

Alternate Solution: Due to the cubic, an alternate solution might be more tempting. In fact, there are **many** ways to solve this problem (and I encourage you to come up with your own!). One possibility is to Graph $r + z_{95}t$ as a functions of N, and see where it equal p_n. This graph is on the next page.

Alternate Solution to 9

Here we see the graph of $r + z_{95}t$ as a functions of N.



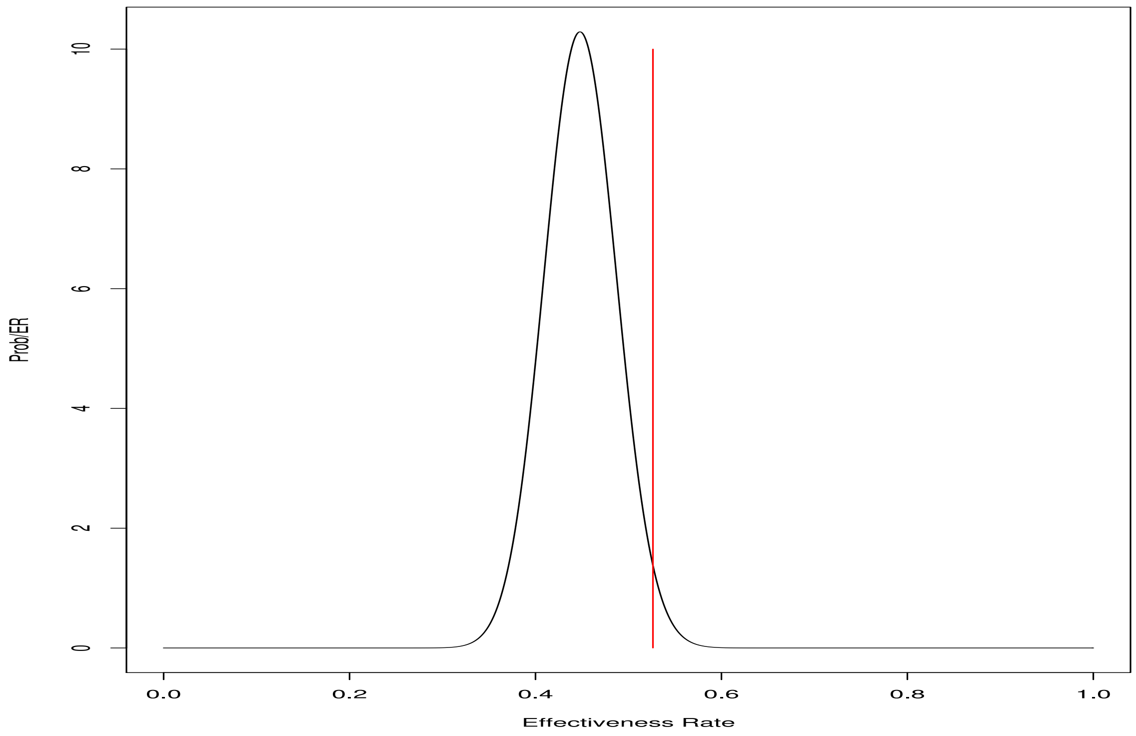
The need value is near 335, so we list the values near there to find the answer (recall $p_n \approx 0.52632$):
[1,] 330 0.5266061 [2,] 331 0.5265203 [3,] 332 0.5264347 [4,] 333 0.5263493 [5,] 334 0.5262642 [6,] 335 0.5261792 [7,] 336 0.5260946 [8,] 337 0.5260101 [9,] 338 0.5259259 [10,] 339 0.5258419 [11,] 340 0.5257581

A Question

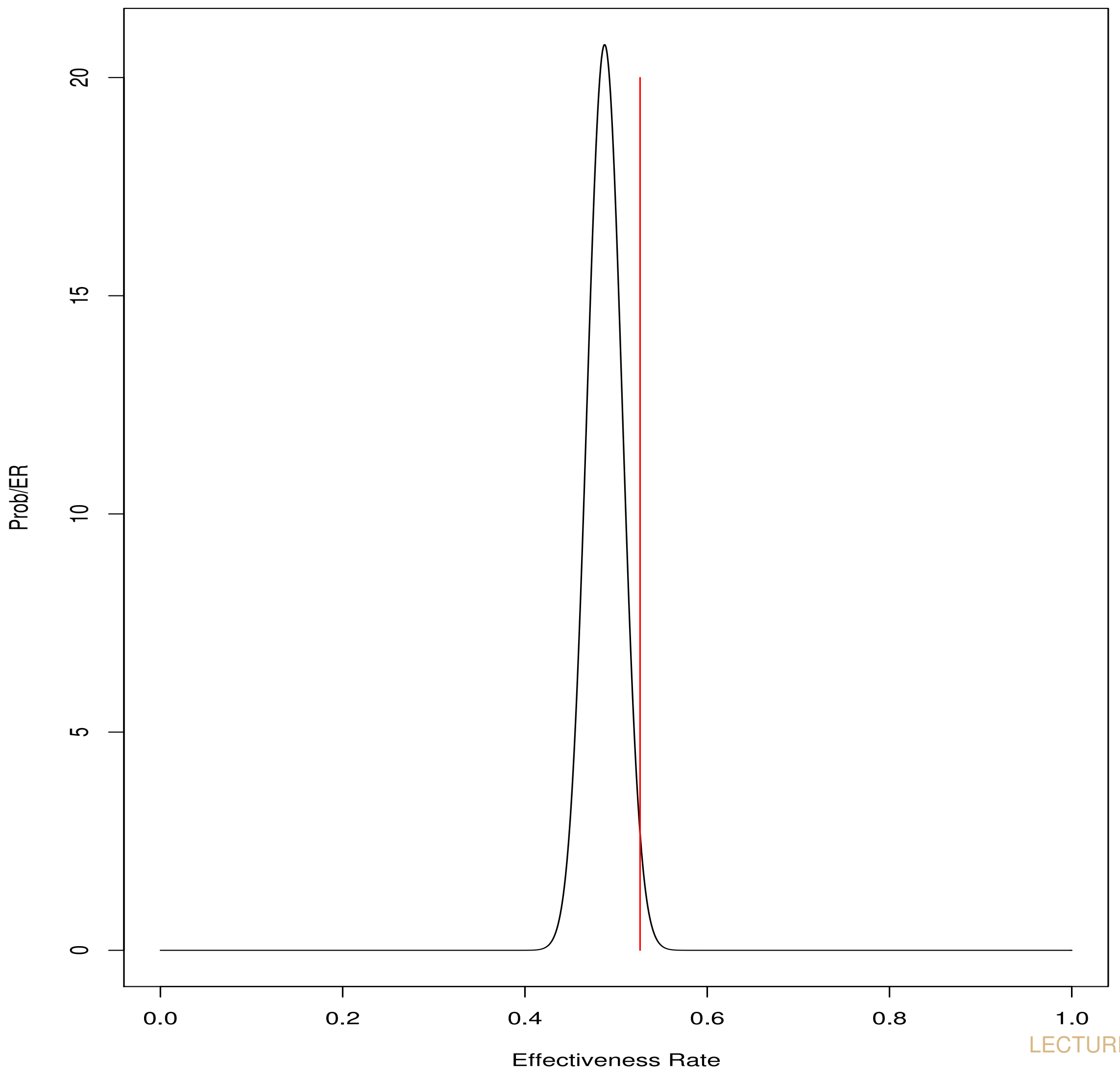
"What's up with all this hunting? I've seems some statistics before, and, as I recall, you just plug in and get an answer."

The real power of the hunting method is that it gives anyone who can use a computer a way of asking **and answering** complicated and interesting questions. For example, in medicine you often need to empirically beat the Null hypothesis by a fixed percent (medical significance). For example, we could imagine that with current techniques that the expression of a certain side effect is roughly $p_n = \frac{20}{38}$, and we want to to convince our selves that a certain drug has a significant effect on this expression rate. We'd like to produce an experiment such that if the result is empirically significant (say a rate $\leq .45$), then we will reject the Null hypothesis that the procedure has no effect (with a level of significance $\alpha = 0.05$). We have seen that if our prior is rather suspicious of this new drug ($beta(180, 162)$) then we would need to try the procedure on 334 patients in order to believe any empirically significant result. Clearly the prior plays a big role! For example, if we know nothing about this drug (prior $beta(1, 1)$), then it would only require 163 patients. In the next slide, we see the graph of the relevant posterior for the smaller trial when s/N is very near $.45$ together with a line representing p_n ($beta(74, 164)$ for $N = 163$). In the slide after the next, we see the posterior for the larger trial when s/N is very near $.45$ together with a line representing p_n ($beta(330, 347)$ for $N = 334$). You should think about the difference in these graphs!

The Posterior $beta(74, 164)$



The Posterior $beta(330, 347)$



Your Short Project

You will pick a topic of interest to you in which there is a hypothesis you wish to test, design an experiment to test this hypothesis, and implement a pretest of your experiment. You may work in group of 3 or smaller. If your topic is work intensive (like a Coke/Pepsi challenge), I recommend working in a group. Your project will have two parts a

PROPOSAL
(due by Monday May 23)

and a

CONCLUSION
(due by Noon on June 6th).

PROPOSAL

Your will clearly explain your your proposed implementation and as well as any other pertinent facts about your experiment in a 2+ page paper (double spaced) prior to the experiment. In this paper you should discuss your goal, protocol, and the hypothesis test you intend to run. A description of the hypothesis test should include:

1. A choice a prior **and** a careful to explanation of your choice. (Recall our discussion after analyzing Example part 1.) You **may** use a prior with known weakness, provided you articulate them. (Recall our discussion after analyzing Example part 2.)

2. A carefully justified choice of a Null Hypothesis. (Recall our discussion after analyzing Example part 4).

3. Choice of a level of significance α (It would be nice if you could view your result as at least statistically significant!)

4. Decide on a value of your primary statistic that you feel would be clearly be "empirically significant". Determine the number of trials that would be required for you to reject the null if this value arises. (Recall our discussion after analyzing Example part 9 and the slide titled "A Question".).

5. Find the critical region for your test (see Example part 8).

Disclaimer

Please keep your test as "innocuous" in nature as possible. The official word on hypothesis tests involving Human Subjects at Dartmouth is as follows: Dartmouth College has an office on campus referred to as the office of the Committee for the Protection of Human Subjects (CPHS). This office is federally mandated to review all research involving human participants at the College (and associated institutions). A research study will receive one of three levels of review: Exempt, Expedited or Full Committee review. The general scope of the project for this course has been approved with a designation of Exempt by the CPHS office because the information you plan to obtain is considered "innocuous" in nature. Your professor will be reviewing your particular project. It is possible your project may not fall into the exempt category, in which case your professor will contact the CPHS office. This is particularly true if your research involves minors or if it involves information that could place a participant at risk of criminal or civil liability or be damaging to the subject's financial standing, employability, or reputation. For our project, I highly recommend that you do not involve minors. Also, when designing your experiment please keep in mind the following guidelines: Prior to asking a participant to become involved in a research study s/he should be aware of: your name and affiliation with Dartmouth, the reason for the project, the level of confidentiality of responses, and the voluntary nature of the person's participation. This may be accomplished verbally or through a few sentences at the beginning of a survey instrument (see information sheet template in the "Forms" section of our web-site: www.dartmouth.edu/ cphs)

Keep in mind...

If you feel like being safe, then I recommended using a success failure rate as your primary statistic (since before this project was assigned, we had already finished a careful discussion of this case). If you choose to use a different sort of statistic, then great! But be warned that a good analysis may **require** some reading ahead and a fair dose of creativity.

Some questions that I will keep in mind when grading your Proposal and Conclusion are:

1. Are the notions of double blind, controlled and randomized appropriate to your experiment and implemented sensible way?

2. What was your method of choosing a sample population, and how did you test to see whether your sample population was an appropriate sample of the population that you made inferences about?

3. Was your choice of primary statistic relevant to your goal, well thought out, and analyzed correctly?

Conclusion

After running the experiment an analysis of your results will be performed and turned in as a second 2+ page (double spaced) paper. This paper will include a presentation of your results, an analysis of your results, and a conclusion concerning your goals. Also you should include a description of any problems that arose while implementing your experiment and a discussion of any improvements you would implement if you were to run a second more comprehensive pretest.

Comparing Controls an Treatments

We may find it convenient to use a pair of beta priors $dbeta(x, a_t, b_t)$ and $dbeta(x, a_c, b_c)$ for our control and treatment success rates (and view them as independent). The posteriors are then $dbeta(x, a_t + s_t, b_t + f_t)$ and $dbeta(x, a_c + s_c, b_c + f_c)$, and we can explore the success rate difference

$$d = rbeta(1, a_t + s_t, b_t + f_t) - rbeta(1, a_c + s_c, b_c + f_c)$$

by noting that for big enough posterior a and b parameters that

$$z = rnorm(x) \approx \frac{d - (r_t - r_c)}{\sqrt{t_t^2 + t_c^2}}.$$

Using the Approximation

For example, we can approximate

$$PdAL(x) = P(d \geq x)$$

via

$$P(d \geq x) \approx P\left(z \geq \frac{x - (r_t - r_c)}{\sqrt{t_t^2 + t_c^2}}\right)$$

and find the $perc$ % probability interval about the difference $r_t - r_c$ to be approximately

$$r_t - r_c \pm z_{perc}\sqrt{t_t^2 + t_c^2}.$$

Pepsi PdAL

Our Pepsi data for discrete and continuous *No Knowledge Priors* is pictured below. In our continuous model, we choose $a_t = b_t = a_c = b_c = 1$ and we found $s_t = 30$, $f_t = 20$, $s_c = 27$, $f_c = 23$. So our posterior estimate the success rate difference was $\frac{31}{51} - \frac{28}{51} = \frac{3}{51} \approx 0.059$, and the 95 % probability interval was $\frac{3}{51} \pm 1.96(0.096)$. In other words, we would expect the true percent to be in $[-0.130.25]$ 95 % of time.