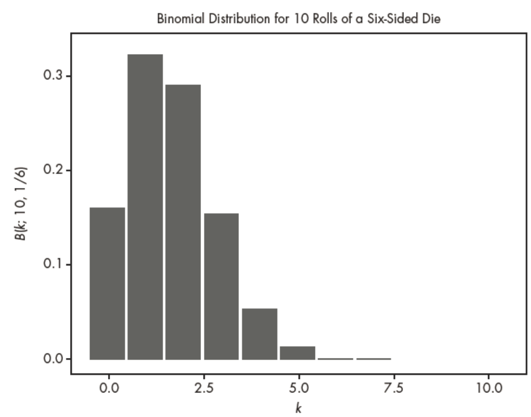
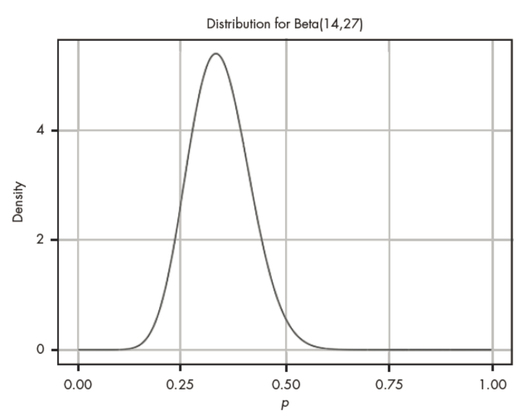
Bayesian Statistics the Fun Way

Understanding Statistics and Probability with Star Wars, LEGO, and Rubber Ducks

by Will Kurt

errata updated to print 6

Page	Error	Correction	Print corrected
17	Equation replacement	$P(H_{\text{no article}}) = 20 \times (1 - P(H_{\text{no article}}))$	Print 2
29	So, using our die roll and coin toss example, the probability of rolling a number **less than** 6 or flipping a heads is:	So, using our die roll and coin toss example, the probability of rolling a number **equal to** 6 or flipping a heads is:	Print 3
40	Figure replacement	 *Figure 4-2: The probability of getting 6 k times when rolling a six-sided die 10 times*	Print 3
51	Figure replacement	 *Figure 5-3: Visualizing the beta distribution for our data collected about the black box*	Print 3

<table>
<tr><td>51</td><td>What we get in the end is a function that describes the probability of each possible hypothesis for our true belief in the probability of getting two heads from the box . . .</td><td>What we get in the end is a function that describes the probability of each possible hypothesis for our true belief in the probability of getting two coins from the box . . .</td><td>Print 5</td></tr>
<tr><td>53</td><td>Here we calculate the probability that the chance of getting two coins from the box is 0.5, given the data:</td><td>Here we calculate the probability that the chance of getting two coins from the box is less than or equal to 0.5, given the data:</td><td>Print 3</td></tr>
<tr><td>71</td><td>numberOfRedStuds = P (yellow | red) × numberOfRedStuds = 1/5 × 20 = 4</td><td>numberOfRedUnderYellow = P(yellow | red) × numberOfRedStuds = 1/5 × 20 = 4</td><td>Print 5</td></tr>
<tr><td>87</td><td>We just add the alphas for our prior and posterior and the betas for our prior and posterior, and we arrive at a normalized posterior. Because this is so simple, working with the beta distribution is very convenient for Bayesian statistics. To determine our posterior for Han making it through the asteroid field, we can perform this simple calculation:

Beta (20002,7401) = Beta (2 + 20000, 7400 + 1)</td><td>We just add the alphas for our prior and posterior and the betas for our prior and likelihood and we arrive at a normalized posterior. Because this is so simple, working with the beta distribution is very convenient for Bayesian statistics. To determine our posterior for Han making it through the asteroid field, we can perform this simple calculation:

Beta (20002,7441) = Beta (2 + 20000, 7440 + 1)</td><td>Print 5</td></tr>
<tr><td>88</td><td>Figure replacement</td><td>

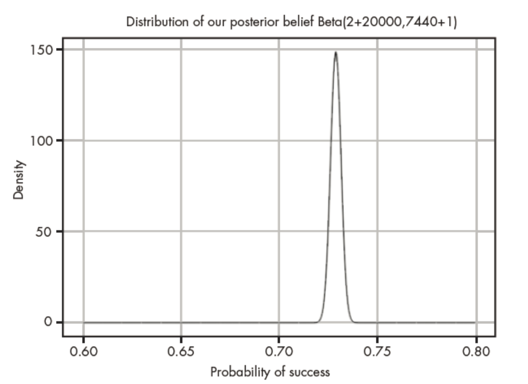


Figure 9-3: Combining our likelihood with our prior gives us a more intriguing posterior.

</td><td>Print 5</td></tr>
<tr><td>105</td><td><table><tr><td>Observation</td><td>Difference from mean</td></tr><tr><td>Group b</td><td></td></tr><tr><td>2.80</td><td>–0.16</td></tr></table></td><td><table><tr><td>Observation</td><td>Difference from mean</td></tr><tr><td>Group b</td><td></td></tr><tr><td>2.80</td><td>–0.2</td></tr></table></td><td>Print 5</td></tr>
<tr><td>105</td><td>Equation replacement</td><td>$\sum_{i=1}^{5} a_i - \mu_a = 0 \qquad \sum_{i=1}^{5} b_i - \mu_b = 0$</td><td>Print 5</td></tr>
<tr><td>106</td><td>Equation replacement</td><td>$\frac{1}{5} \times \sum_{1}^{5} \lvert a_i - \mu_a \rvert = 0.04 \qquad \frac{1}{5} \times \sum_{1}^{5} \lvert b_i - \mu_b \rvert = 0.416$</td><td>Print 5</td></tr>
</table>

116	Equation replacement	$N(\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \times e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	Print 3
127	`xs <- seq(0.005,0.01,by=0.00001)` `xs.all <- seq(0,1,by=0.0001)`	`xs <- seq(0.005,0.01,by=0.00001)` ~~`xs.all <- seq(0,1,by=0.0001)`~~	Print 5
130	As Figure **3**-5 illustrates, the point where this line intersects the x-axis gives us our median!	As Figure **13**-5 illustrates, the point where this line intersects the x-axis gives us our median!	Print 5
163	P $(D \mid H_1)$ = 0.94 × **0.89** = 0.78	P $(D \mid H_1)$ = 0.94 × **0.83** = 0.78	Print 3
164	The **prior** odds look like this:	The **probabilities** look like this:	Print 5
164	Equation replacement	$O(H_1) \times \frac{P(D \mid H_1)}{P(D \mid H_2)} = \frac{11}{37,000} \times 2.23 = \frac{245}{370,000}$	Print 5
178	Equation replacement	$PO = O(H_2)' \times \frac{P(D_{15} \mid H_2)}{P(D_{15} \mid H_3)} = \frac{1}{1,000} \times \frac{\left(\frac{9}{10}\right)^{14} \times \left(1-\frac{9}{10}\right)^{1}}{\left(\frac{9}{10}\right)^{14} \times \left(1-\frac{9}{10}\right)^{1}} = \frac{1}{1,000}$	Print 3
224	Since you've run half **a mile**, using this simple formula, we can figure out:	Since you've run half **an hour**, using this simple formula, we can figure out:	Print 5
234	A3. This is the same as B(5; 10, 1/**23**). As expected, the probability of this is **extremely low: about 1/32,000**.	A3. This is the same as B(5; 10, 1/**13**). As expected, the probability of this is **low: about 1/2,200**.	Print 6
236	Luckily we already did all this work earlier in the chapter, so we know that (A) = **4/1,000** and P(B) = 3/(100,000).	Luckily we already did all this work earlier in the chapter, so we know that (A) = **8/100** and P(B) = 3/(100,000).	Print 5
237	Plugging in our numbers, we get an answer of **100,747/25,000,000 or 0.00403**.	Plugging in our numbers, we get an answer of **800,276/10,000,000 or 0.0800276**.	Print 5
242	`temp.sd <- my.sd(temp.data)`	`temp.sd <- sd(temp.data)`	Print 4
250	P (D ǀ H2) = **0.63 × 0.55 × 0.49 = 0.170**	P (D ǀ H2) = **0.94 x 0.83 x 0.49 = 0.382**	Print 5
250	This means that given the Bayes factor alone, vestibular schwannoma is a roughly **two** times better explanation than labyrinthitis. Now we have to look at the odds ratio:	This means that given the Bayes factor alone, vestibular schwannoma is a roughly **four** times better explanation than labyrinthitis. Now we have to look at the **prior** odds ratio:	Print 5
251	The end result is that **labyrinthititis** is only a slightly better explanation than **vestibular schwannoma**.	The end result is that **vestibular schwannoma** is only a slightly better explanation than **labyrinthititis**.	Print 5

254	Equation replacement	$$50 = \frac{9}{19} \times 950$$ $$\mathrm{BF} = 950$$	Print 5
254	`dx <- 0.01` `hypotheses <- seq(0,1,by=0.01)`	`dx <- 0.01` `hypotheses <- seq(0,1,by=dx)`	Print 5