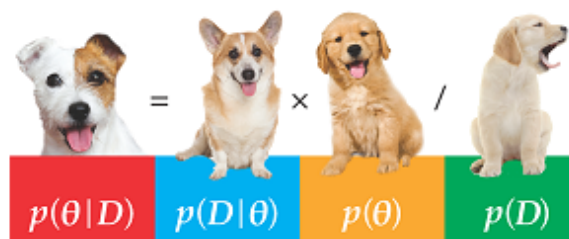


# Doing Bayesian Data Analysis

A Tutorial with R, JAGS, and Stan




Search this site

2ND EDITION:  
WHAT'S NEW

SAMPLE CHAPTER

PURCHASE WITH  
DISCOUNT

PARAPHERNALIA

SOFTWARE, WITH  
PROGRAMS FOR  
BOOK

LIKE THE BOOK?  
REVIEW IT :-)

**EXERCISES AND  
SOLUTIONS**

FIGURES FOR  
INSTRUCTORS

VIDEOS

WORKSHOPS

BLOG

ABOUT THE  
AUTHOR

CONTACT

CORRIGENDA

1ST EDITION

## Exercises and solutions

**Solutions to the exercises in the 2nd edition:** The solutions for exercises in Chapters 1 - 18 can be retrieved by clicking [here](#). Solutions for later chapters will be available here at a future date.

For solutions to the exercises in the 1st edition, please click the 1st Edition link in the left margin.

**Additional exercises for the 2nd edition:** Please contact the author if you have suggested exercises to add to the list below.

Ch.	Purpose	Exercise Statement	D
2	Thinking about noise in a posterior predictive check.	Consider Figure 2.6, p. 29, in DBDA2E. Two of the data points fall above the vertical bars. Does this mean that the model does not describe the data well? Briefly explain your answer.	201
2	Thinking about prior probabilities in reallocation of credibility in disease diagnosis.	Suppose there are two factories of trick coins. One factory creates tail-biased coins, such that the probability of T (an outcome we'll call "negative") is 95%. We'll call this the "healthy" factory. The second factory creates head-biased coins, such that the probability of H (an outcome we'll call "positive") is 95%. We'll call this the "disease" factory. (Just like real diagnostic tests, "positive" suggests having the disease.) Now, suppose you grab a coin at random from a sack of coins. You flip the coin and it comes up "positive." Our goal is to think about the probability that the coin came from the healthy factory or the disease factory.  A. Suppose that the sack of coins has 50% healthy coins. (This is your initial, prior allocation of credibility across the	201

		<p>two types of coins.) Given the “positive” outcome of the single flip, should your re-allocation of credibility end up with more credibility on the healthy factory or the disease factory? Why?</p> <p>B. Suppose that the sack of coins has 99% healthy coins. (This is your initial, prior allocation of credibility across the two types of coins, similar to real life.) Given the “positive” outcome of the single flip, should your re-allocation of credibility end up with more credibility on the healthy factory or the disease factory? To answer this question, do the following steps:</p> <p>* Suppose the sack has 10,000 coins. How many of them would you expect to be healthy? (Hint: Of 10,000 coins, we expect 99% of them to be healthy. Compute the actual number.) How many would you expect to be diseased?</p> <p>* If you tested every <i>healthy</i> coin, how many would you expect to come up positive? (Hint: For the number computed in the previous part, we expect 5% to come up positive. Compute the actual number.)</p> <p>* If you tested every <i>diseased</i> coin in the sack, how many would you expect to come up positive?</p> <p>* Out of all the coins in the sack that you would expect to come up positive, what proportion of them are diseased? (Hint: Compute the total number of coins that you would expect to come up positive. Of that total, what proportion is from diseased coins? Your answer should be 16.1%; show your work. By the way, this exercise is similar to Exercise 5.2 of DBDA2E.)</p>	
3	Experience with <code>cata</code> files and indexing in R.	<p>Invoke RStudio with the DBDA2E program folder as its working directory. Read a csv file into a data frame as follows:</p> <pre>datFram = read.csv( "HGN.csv" )</pre> <p>A. Display the first three rows of the data frame using explicit numerical indexing of rows in bracket notation.</p> <p>B. Display the first three rows of the data frame using the <code>head()</code> function.</p>	20'

		<p>C. What is the class (e.g., factor, integer, etc.) of the \$group component of the data frame? Use the class() function.</p> <p>D. Change the \$group component to a factor, using the factor() function, and display the result with its levels.</p> <p>E. Display the Hair component of the data frame by referring to it three different ways: Using \$ notation, using brackets with column name, and using brackets with column number.</p>	
3	Creating functions in R.	<p>Create a function in R that converts Fahrenheit temperatures to Celsius temperatures. Set the default input temperature to 72. Hint: <math>C = (F - 32) * 5/9</math>.</p> <p>A. Show the specification of the function.</p> <p>B. Run the function with no explicit input, to reveal its default output.</p> <p>C. Run the function with 98.6 as its input.</p> <p>D. Run the function with the vector <code>c(32, 72, 98.6, 212)</code> as its input. Very briefly explain why the function produces a vector as output.</p>	201
4	Hands-on experience approximating probability density from a large sample, as in Figures 4.2 and 4.3.	<p>Consider a normal probability density function that has mean of 10.0 and standard deviation of 0.20.</p> <p>A. What is the exact probability density at <math>x=9.9</math>? Hint: <code>dnorm(x,m,sd)</code></p> <p>B. Generate a random sample of 100,000 values from the normal distribution. Hint: <code>rnorm(N,m,sd)</code></p> <p>C. Compute the approximate probability density at <math>x=9.9</math> by counting the proportion of sampled points that falls between <math>x=9.8</math> and</p>	201

		<p><math>x=10.0</math>, and then dividing by the width of the bin. Show your code and explain what it does. Is the result approximately the same as Part A? (It should be.) Hint: Name the sample of values <code>heapOdata</code>; try <code>sum( heapOdata &gt;= 9.8 &amp; heapOdata &lt; 10.0 )</code>.</p>	
6	Experience converting from central tendency and concentration of beta distribution to a and b shape parameters	<p>A. Suppose a colleague tells you she read a research report in which "there were about 50 patients and two-thirds of them survived at least one year after surgery." Suppose you want to use this information in an informal prior for subsequent data. (You'll get the exact numbers later.) What are the corresponding a and b shape parameters? Hint: Convert from mean and concentration; why use the mean?</p> <p>B. Suppose you are given a six-sided die, and you want to find out whether it's biased by assessing the probability of rolling a 1-dot outcome. You know the die just came out of a brand new package from a reputable manufacturer, so you believe that the most probable probability of a 1-dot outcome is <math>1/6</math>. Suppose it would take 50 rolls to begin to sway you from your prior belief that the die is fair. What are the corresponding a and b shape parameters? Hint: Convert from mode and concentration; why use the mode?</p>	201
6	Seeing that various broad priors (all beta distributions) yield similar posterior distributions for moderately large data.	<p>For all these cases, suppose the data show 63 heads in 97 flips.</p> <p>A. Suppose the prior is <math>\text{beta}(\theta a=0.01, b=0.01)</math>, which is (approximately) the so-called Haldane prior. What is mode and 95% HDI of the posterior distribution? Show the result graphically. Hint: Use <code>BernBeta.R</code>.</p> <p>B. Suppose the prior is <math>\text{beta}(\theta a=1, b=1)</math>, which is uniform. What is mode and 95% HDI of the posterior distribution? Show the result graphically. Hint: Use <code>BernBeta.R</code>.</p> <p>C. Suppose the prior is <math>\text{beta}(\theta a=2, b=4)</math>, which is gently biased toward tails. What is mode and 95% HDI of the posterior distribution?</p>	201

		<p>Show the result graphically. Hint: Use <code>BernBeta.R</code>.</p> <p>D. Is there much difference in the posterior distributions (modes and HDI's) across the different priors? Hint: See the purpose of the exercise.</p>	
7	Another look at autocorrelation, similar to Figure 7.12; and, an R exercise.	<p>Consider the following R code:</p> <pre> source("DBDA2E-utilities.R")  openGraph(width=5,height=7.5)  layout( matrix(1:6,nrow=3) )  par(mar=c(4,4,3,1),mgp=c(2,0.7,0))  maxLag=10  for ( inert in c(0.99,0.80) ) {    x = rep(0,500)    for ( i in 2:length(x) ) {      x[i] = inert*x[i-1] + rnorm(1,0,1-inert)    }    x = (x-mean(x))/sd(x)    #    plot( x , ylab="x" , cex.lab=1.5 , type="l" )    #    plot( x[1:(500-maxLag)] , x[(maxLag+1):500] ,          xlab=bquote("x[i]" ) , ylab=bquote("x[i+"*. (maxLag) *"]" ) ,          cex.lab=1.5 , cex.main=1.5 ,          main=bquote(r==. (round(cor(x[1:(500- maxLag)],x[(maxLag+1):500]),2))) )    #    acfInfo = acf( x , lag.max=maxLag+2 , main="" )    title( main=bquote("acf["*. (maxLag) *"] = "*. (round(acfInfo\$acf[maxLag+1],2)) ) ,          cex.main=1.5 )    points( maxLag , acfInfo\$acf[maxLag+1] , cex=2 , lwd=3 ) </pre>	

		<p>}</p> <p>A. Explain, briefly, what every line does, perhaps by adding comments before or at the end of each line.</p> <p>B. Run the code. It produces a graphical output; include the graph in your write-up.</p> <p>C. Explain the scatterplots in the second row, including the relation of the scatterplot to the traces in the first row. Also explain the number in the title of the scatterplot.</p> <p>D. Explain the bar graphs in the third row, including the circle on one of the bars, and the relation of that bar to the scatterplot in the second row.</p> <p>E. What is the key difference between the columns of the graph? (Only a brief answer is wanted here.)</p>	
8	See that the JAGS answer matches the beta distribution from an analytical solution.	<p>Immediately after Exercise 8.1: Here our goal is to superimpose an analytically correct beta distribution curve on the MCMC histogram produced by JAGS. After the end of the script of Exercise 8.1, run the R code below, with the boxes <code>□</code> replaced with the appropriate values:</p> <pre> <b>openGraph()</b> <b>plotPost(</b>   <b>mcmcCoda[, "theta[1]" ] ,</b>   <b>xlim=c(0,1) )</b> <b>a = □ ; b = □ # constants</b> <b>from prior in Jags-Ydich-</b> <b>XnomSsubj-MbernBeta.R</b> <b>H = □ ; T = □ # heads and</b> <b>tails from your data</b> <b>thetaGrid=seq(0,1,length=201)</b> <b>lines( thetaGrid , dbeta(</b>   <b>thetaGrid , a+H , b+T ) )</b> </pre> <p>Explain how you determined the a and b values.</p> <p>Explain the meaning of the lines() command, being sure to relate it to Equation 6.8.</p> <p>Include the resulting graph in your write-up. (The superimposed curve should closely match the histogram.)</p>	202
8	Getting familiar with some details of	<p>A. Consider the data and script you used for Exercise 8.1. Re-run it two ways:</p>	202

	<p>the JAGS program, “under the hood.”</p>	<p>genMCMC( data=myData , numSavedSteps=2000 )  genMCMC( data=myData , numSavedSteps=50000 )</p> <p>For both, include the diagnostic plot of theta[1] in your write-up. Explain what the numSavedSteps argument does, and describe its effect on the diagnostic plot.</p> <p>B. In the program Jags-Ydich-XnomSsubj-MbernBeta.R, change these lines</p> <pre>jagsModel = jags.model(   "TEMPmodel.txt" ,   data=dataList ,   inits=initsList ,   n.chains=nChains ,   n.adapt=adaptSteps )</pre> <p>to this:</p> <pre>jagsModel = jags.model(   "TEMPmodel.txt" ,   data=dataList , #   inits=initsList ,   n.chains=nChains ,   n.adapt=adaptSteps )</pre> <p>Then save the modified program and re-run the high-level script. What did the change in the program line do to JAGS? That is, what did JAGS do differently? Did the MCMC output change in any noticeable way?</p>	
8	<p>Preview of the ROPE argument in plotPost(), and prelude of the Savage-Dickey method for Bayes factors in Section 12.3 and Exercise 12.1.</p>	<p>A. Consider the data and script you used for Exercise 8.1. This time, after running the main script, run these lines:</p> <pre>openGraph() plotPost( mcmcCoda["theta[1]" , xlim=c(0,1) , ROPE=c(0.45,0.55) )</pre> <p>Notice the ROPE argument; describe what it does to the graph and what the annotation in the graph means. In particular, how much of the posterior distribution falls inside the ROPE?</p> <p>B. Re-run the script but with the data commented out, so that JAGS samples from the prior distribution (recall Exercise 8.4). Again run the</p>	202

		<p>lines of part A. How much of prior distribution falls inside the ROPE?</p> <p>C. What is the ratio of the posterior probability inside the ROPE to the prior probability inside the ROPE? (You can think of this as the shift in credibility of theta values near chance.)</p>	
9	Hands-on experience with the gamma distribution	<p>Make plots like those in Figure 9.8, p. 237. You don't have to make the annotation exactly match the look of Figure 9.8, but make sure that the shape, rate, mean, sd, and mode show up somewhere in the figure. Hint: Start with mean or mode and sd, and use Equations 9.7 or 9.8 (or corresponding R functions) to determine the shape and rate. Extra special bonus hint: Embellish code like the following</p> <pre>kappa = seq( 0 , 150 , length=1001 ) source("DBDA2E-utilities.R") sr = gammaShRaFromModeSD( mode=18 , sd=40 ) plot( kappa , dgamma( kappa , shape=sr\$shape , rate=sr\$rate ) , type="l" )</pre>	20%
9	Hands-on computation of likelihood in a hierarchical model, to see that shrinkage really does yield higher likelihood	<p>In this exercise, you'll use R to compute the likelihood value in Equation 9.10, p. 247.</p> <p>Consider four coins, each flipped 4 times, with these outcomes:</p> <p><math>\{y_{i s=1}\}=1,0,0,0</math> <math>\{y_{i s=2}\}=1,1,0,0</math>  <math>\{y_{i s=3}\}=1,1,0,0</math> <math>\{y_{i s=4}\}=1,1,1,0</math></p> <p>A. What are the proportions of heads for each coin?</p> <p>B. What is the value of the likelihood when <math>w=0.5</math>, <math>k=2</math>, <math>q_1=0.25</math>, <math>q_2=0.50</math>, <math>q_3=0.50</math>, and <math>q_4=0.75</math>? Show your computations, perhaps in R. Do these parameter values constitute any shrinkage relative to the data proportions? Is the shape of the beta distribution flat or peaked?</p> <p>C. What is the value of the likelihood when <math>w=0.5</math>, <math>k=20</math>, <math>q_1=0.35</math>, <math>q_2=0.50</math>, <math>q_3=0.50</math>, and <math>q_4=0.65</math>? Show your computations, perhaps in R. Do these parameter values constitute any</p>	20%



		shrinkage relative to the data proportions? Is the shape of the beta distribution flat or peaked? Which set of parameter values, these or the previous part, yield a higher likelihood value for the data?	
10	Practice and intuition building with the model-comparison example in Section 10.2.1, using different data and priors	<p>Consider the example in Section 10.2.1, regarding comparison of head-biased and tail-biased factories as models for a coin. Re-do the example with the following data sets, factory biases, and prior probabilities of the factories. For each case, show the R commands you used and their results, and show the algebra you used to compute the Bayes factor and the posterior probability of each factory (in R, if you like).</p> <p>A. <math>z=3</math> , <math>N=9</math> , <math>w_1=0.25</math> , <math>w_2=1-w_1</math> , <math>k=12</math> , <math>p(m=1)=0.5</math> . State what is different about this example relative to the book example. How do the results compare with the results in the book example? (Hint: Your results should be just the “flip” of the results in the book example.)</p> <p>B. <math>z=6</math> , <math>N=9</math> , <math>w_1=0.25</math> , <math>w_2=1-w_1</math> , <math>k=120</math> , <math>p(m=1)=0.5</math> . State what is different about this example relative to the book example. How do the results compare with the results in the book example?</p> <p>C. <math>z=6</math> , <math>N=9</math> , <math>w_1=0.025</math> , <math>w_2=1-w_1</math> , <math>k=12</math> , <math>p(m=1)=0.5</math> . State what is different about this example relative to the book example. How do the results compare with the results in the book example?</p> <p>D. <math>z=6</math> , <math>N=9</math> , <math>w_1=0.25</math> , <math>w_2=1-w_1</math> , <math>k=12</math> , <math>p(m=1)=0.05</math> . State what is different about this example relative to the book example. How do the results compare with the results in the book example?</p>	202

In the file list below, click only on the download arrow at the right, not on the file name or version number.





KRUSCHKE-DBDA..JOHN K. KRUSCH... V.3



[Sign in](#) | [Recent Site Activity](#) | [Report Abuse](#) | [Print Page](#) | Powered By **Google Sites**