```
# Suppose we are giving two students a multiple-choice exam with 40 questions,
# where each question has four choices. We don't know how much the students
# have studied for this exam, but we think that they will do better than just
# guessing randomly.
# 1) What are the parameters of interest?
# 2) What is our likelihood?
# 3) What prior should we use?
# 4) What is the prior probability P(theta>.25)? P(theta>.5)? P(theta>.8)?
# 5) Suppose the first student gets 33 questions right. What is the posterior
#    distribution for theta1? P(theta1>.25)? P(theta1>.5)? P(theta1>.8)?
#    What is a 95% posterior credible interval for theta1?
# 6) Suppose the second student gets 24 questions right. What is the posterior
#    distribution for theta2? P(theta2>.25)? P(theta2>.5)? P(theta2>.8)?
#    What is a 95% posterior credible interval for theta2?
# 7) What is the posterior probability that theta1>theta2, i.e., that the
#    first student has a better chance of getting a question right than
#    the second student?

############
# Solutions:

# 1) Parameters of interest are theta1=true probability the first student
#    will answer a question correctly, and theta2=true probability the second
#    student will answer a question correctly.

# 2) Likelihood is Binomial(40, theta), if we assume that each question is
#    independent and that the probability a student gets each question right
#    is the same for all questions for that student.

# 3) The conjugate prior is a beta prior.
set up columns (starting in Column B): theta  f(theta)  L(theta1)  f(theta1|Y)
start theta at 0.01 in cell B2
> Edit > Fill > Series  -- Columns -- Step .01, Stop 0.99
set prior parameters:  label alpha in A2, value 1 in A3
                       label beta in A4, value 1 in A5
prior density in C3
= (FACT($A$3+$A$5-1)/FACT($A$3-1)/FACT($A$5-1))*B2^($A$3-1)*(1-B2)^($A$5-1)
copy and paste to the rest of Column C
> Insert > Chart > Line
change prior parameters, try alpha=4, beta=2, then try alpha=8, beta=4

# 4) Find probabilities using the BETADIST function.
=1-BETADIST(.25,8,4)
=1-BETADIST(.5,8,4)
=1-BETADIST(.8,8,4)

# 5) Posterior is Beta(8+33,4+40-33) = Beta(41,11)
# posterior mean and MLE
=41/(41+11)
=33/40

L(theta1) in D3
=BINOMDIST(33,40,B2,FALSE)
posterior density in E3
= (FACT(41+11-1)/FACT(41-1)/FACT(11-1))*B2^(41-1)*(1-B2)^(11-1)
> Insert > Chart > Line
plotting together doesn't work well because of difference in scale

# posterior probabilities
=1-BETADIST(.25,41,11)
=1-BETADIST(.5,41,11)
=1-BETADIST(.8,41,11)

# equal-tailed 95% credible interval
```

```
=BETAINV(0.025,41,11)
=BETAINV(0.975,41,11)

# 6) Posterior is Beta(8+24,4+40-24) = Beta(32,20)
# posterior mean and MLE
=32/(32+20)
=24/40

L(theta2) in Column F
=BINOMDIST(24,40,B2,FALSE)
f(theta2|Y) in Column G
= (FACT(32+20-1)/FACT(32-1)/FACT(20-1))*B2^(32-1)*(1-B2)^(20-1)
> Insert > Chart > Line

=1-BETADIST(.25,32,20)
=1-BETADIST(.5,32,20)
=1-BETADIST(.8,32,20)

=BETAINV(0.025,32,20)
=BETAINV(0.975,32,20)

# 7) Estimate by simulation: draw 500 samples from each and see how often
#    we observe theta1>theta2

theta1
=BETAINV(RAND(),41,11)
theta2
=BETAINV(RAND(),32,20)
=IF(H2 > I2, 1, 0)
get sum, divide by 500


# Note for other distributions:
# GAMMA.DIST,GAMMA.INV,GAMMA.INV(RAND(),a,1/b)
# NORM.DIST,NORM.INV,NORM.INV(RAND(),mu,sigma)
```