Stat 511 Bayes Overview

(Why You Must Take Stat 544 Before You Are Done)

Prof. Stephen Vardeman

ISU Statistics and IMSE

April 2008

In a Nutshell

Bayesian Statistics

- Replaces a family of models F_θ with a single probability model ... by treating the F_θ as conditional distributions of the data given θ and placing a "prior" probability distribution on θ (leading to a joint distribution for data and parameter)
- Replaces inference based on a likelihood with inference based on a "posterior" (the conditional distribution of parameter given data) ... ultimately

$$L(\theta) \text{ is replaced by } L(\theta)\, g(\theta)$$

- In its modern incarnation, is implemented primarily through *simulation* from the posterior instead of through pencil and paper Stat 542 calculus with joint and conditional distributions

A Simple Discrete Example (Likelihood/Data Model)

Suppose that a random quantity X with possible values $1, 2, 3,$ and 4, has a distribution that depends upon some parameter θ that has possible values $1, 2,$ and 3. We'll suppose the three possible distributions of X (appropriate under the three different values of θ) may be specified by probability mass functions $f(x|\theta)$ given in Table 1.

Table 1 Three Possible Distributions of X

x	$f(x\|1)$
1	.4
2	.3
3	.2
4	.1

x	$f(x\|2)$
1	.25
2	.25
3	.25
4	.25

x	$f(x\|3)$
1	.1
2	.2
3	.3
4	.4

("Big" parameter values will tend to produce "big" values of X.)

A Simple Discrete Example (Prior)

A "Bayes" approach to making inferences about θ based on an observation $X = x$ requires specification of a "prior" probability distribution for the unknown parameter θ. For sake of illustration, cosider a probability mass functions $g(\theta)$ given in Table 2.

Table 2 A "Prior" Distribution for θ

θ	$g(\theta)$
1	.5
2	.3
3	.2

Simple Discrete Example (Joint)

A given prior distribution together with the forms specified for the distribution of X given the value of θ (the "likelihood") leads to a joint distribution for both X and θ with probability mass function $g(\theta, x)$ that can be represented in a two-way table, where any entry is obtained as

$$\begin{aligned} g(\theta, x) &= f(x|\theta)\, g(\theta) \\ &= \text{likelihood} \cdot \text{prior} \end{aligned}$$

For example, using the prior distribution $g(\theta)$ in Table 2, one obtains the joint distribution $g(\theta, x)$ specified in Table 3.

A Simple Discrete Example (Joint cont.)

Table 3 Joint Distribution for X and θ Corresponding to $g(\theta)$

	$\theta = 1$	$\theta = 2$	$\theta = 3$	
$x = 1$	$g(1,1) = .4(.5) = .2$	$g(2,1) = .25(.3) = .075$	$g(3,1) = .1(.2) = .02$	$f(1) = .295$
$x = 2$	$g(1,2) = .3(.5) = .15$	$g(2,2) = .25(.3) = .075$	$g(3,2) = .2(.2) = .04$	$f(2) = .265$
$x = 3$	$g(1,3) = .2(.5) = .1$	$g(2,3) = .25(.3) = .075$	$g(3,3) = .3(.2) = .06$	$f(3) = .235$
$x = 4$	$g(1,4) = .1(.5) = .05$.	$g(2,4) = .25(.3) = .075$	$g(3,4) = .4(.2) = .08$	$f(4) = .205$
	$g(1) = .5$	$g(2) = .3$	$g(3) = .2$	

A Simple Discrete Example (Posterior)

The crux of the Bayes paradigm is that a *joint* probability distribution for X and θ, can be used not only to recover $f(x|\theta)$ and the marginal distribution of θ (the "likelihood" and the "prior distribution" that are multiplied together to get the joint distribution in the first place), but also to find conditional distributions for θ given possible values of X. In the context of Bayes analysis, these are called the *posterior* distributions of θ. For tables laid out as above, they are found by "dividing rows by row totals." We might use the notation $g(\theta|x)$ for a posterior distribution and note that for a given x, values of this are proportional to joint probabilities, i.e.

$$g(\theta|x) \propto f(x|\theta)\, g(\theta)$$

i.e.

$$\text{posterior} \propto \text{likelihood} \cdot \text{prior}$$

Take, for example the situation of the simple discrete example.

A Simple Discrete Example (Posterior cont.)

The four possible posterior distributions of θ (given the observed value of $X = x$) are as in Table 4.

Table 4 Posterior Distributions

	$\theta = 1$	$\theta = 2$	$\theta = 3$
$g(\theta\|1)$	$.2/.295 = .6780$	$.075/.295 = .2542$	$.02/.295 = .0678$
$g(\theta\|2)$	$.15/.265 = .5660$	$.075/.265 = .2830$	$.04/.265 = .1509$
$g(\theta\|3)$	.4255	.3191	.2553
$g(\theta\|4)$	.2439	.3659	.3902

A Simple Discrete Example (Inference)

The "Bayes paradigm" of inference is to base all formal inferences (plausibility statements) about θ *on a posterior distribution of* θ*. For example, an analyst who adopts prior* g *and observes* $X = 3$ *may correctly say that there is (posterior) probability* .2553 *that* $\theta = 3$*. Notice, that this is a different concept than the non-Bayesian concept of "confidence."*

A Simple Discrete Example (Prediction)

In many real contexts, one is doing inference based on data $X = x$ for the purpose of *predicting* the value of an as yet unobserved variable, X_{new}. If (given the value of θ) one is willing to model X and X_{new} as independent variables, one can extend the Bayes paradigm beyond inference for θ to the prediction problem. That is, conditioned on having observed $X = x$ one has a posterior distribution for both θ and X_{new} with a (joint) probability mass function that can be represented in a two way table, where each entry has the form

$$g\left(\theta, x_{\text{new}} | x\right) = f\left(x_{\text{new}} | \theta\right) g\left(\theta | x\right)$$

This can be added across values of θ to produce a *posterior predictive distribution* for X_{new} as

$$g\left(x_{\text{new}} | x\right) = \sum_{\theta} g\left(\theta, x_{\text{new}} | x\right) = \sum_{\theta} f\left(x_{\text{new}} | \theta\right) g\left(\theta | x\right)$$

A Simple Discrete Example (Prediction cont.)

That is, the posterior predictive distribution of X_{new} is what one gets upon weighting the three possible distributions for X_{new} in Table 1 according to the posterior probabilities in Table 4. Considering first the possibility that $X = 1$, note that the conditional distribution for X_{new} given this outcome is

x_{new}	$g(x_{\text{new}} \mid 1)$
1	$.4(.6780) + .25(.2542)$ $+.1(.0678) = .34153$
2	$.3(.6780) + .25(.2542)$ $+.2(.0678) = .28051$
3	$.2(.6780) + .25(.2542)$ $+.3(.0678) = .21949$
4	$.1(.6780) + .25(.2542)$ $+.4(.0678) = .15847$

The entire set of predictive distributions of X_{new} is then given in Table 5.

A Simple Discrete Example (Prediction cont.)

Table 5 Posterior Predictive Distributions for X_{new}

x_{new}	$g(x_{\text{new}}\|1)$
1	.34153
2	.28051
3	.21949
4	.15847

x_{new}	$g(x_{\text{new}}\|2)$
1	.31224
2	.27073
3	.22922
4	.18771

x_{new}	$g(x_{\text{new}}\|3)$
1	.27551
2	.25849
3	.24147
4	.22445

x_{new}	$g(x_{\text{new}}\|4)$
1	.22806
2	.24269
3	.25732
4	.27195

A Simple Discrete Example (Prediction cont.)

The predictive distributions for X_{new} to some degree "follow x" in the sense that "small" x leads to a predictive distribution shifted toward "small" values of x_{new}, while "large" x shifts the predictive distribution toward "large" values of x_{new}. The prior buffers this effect, in that all of the posteriors are in some sense shifted toward smaller values because of the form of g in Table 2 about the form of the different prior assumptions. Note too that in some sense the predictive distributions for X_{new} are qualitatively "flatter" than at least the first and last models for X represented in Table 1. This makes sense, as they are (posterior weighted) averages of the three distributions in that table.

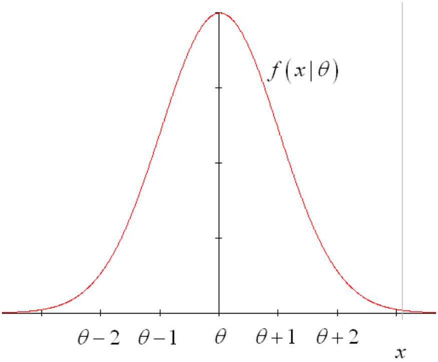
A Fairly Simple Continuous Example (Likelihood/Data Model)

The purpose of this example is to show precisely the operation of Bayes paradigm in a simple continuous context, where the calculus works out nicely enough to produce clean formulas. Suppose that a variable X is thought to be normal with standard deviation 1, but somehow the mean, θ, is unknown. That is, we suppose that X has a probability density function given by

$$f(x|\theta) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x-\theta)^2\right)$$

and pictured in Figure 1.

Figure 1 The normal probability density with mean θ and standard deviation 1



This model assumption is commonly written in "short-hand" notation as $X \sim \mathrm{N}(\theta, 1)$

A Fairly Simple Continuous Example (Prior)

Suppose that one wants to combine some "prior" belief about θ with the observed value of X to arrive at an inference for the value of this parameter that reflects both (the "prior" and the data $X = x$). A mathematically convenient form for specifying such a belief is to assume that *a priori*

$$\theta \sim \mathrm{N}\left(m, \gamma^2\right)$$

that is, that the parameter is itself normal with some mean, m, and some variance γ^2, i.e. θ has probability density

$$g\left(\theta\right) = \frac{1}{\sqrt{2\pi\gamma^2}} \exp\left(-\frac{1}{2\gamma^2}\left(\theta - m\right)^2\right)$$

(An analyst is here specifying not only a form for the distribution of X, but a complete description of his or her pre-data beliefs about θ in terms of "normal," "roughly m," "with uncertainty characterized by a standard deviation γ.")

A Fairly Simple Continuous Example (Posterior)

Intermediate level probability calculations then imply that X and θ have jointly continuous distribution with conditionals for $\theta|X = x$ (posterior densities for θ given $X = x$)

$$g(\theta|x) \propto f(x|\theta)\, g(\theta)$$

i.e.

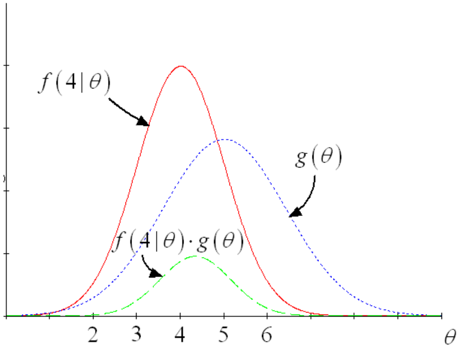
$$\text{posterior} \propto \text{likelihood} \cdot \text{prior}$$

where one thinks of x as fixed/observed and treats the expression above as a function of θ. This is formally exactly as in the simple discrete example of Module 1. The posterior density is proportional to the product of the prior density and the likelihood (the density of X treated as a function of the parameter). This is the continuous equivalent of making a table of products and dividing by "row totals" to get posterior distributions. As it turns out, the posterior (the conditional distribution of $\theta|X = x$) is again normal.

A Fairly Simple Continuous Example (Posterior)

Take, for a concrete example, a case where the prior is $\mathrm{N}(5,2)$ and one observes $X = 4$. It's easy to make the plots in Figure 2 of $f(4|\theta)$, $g(\theta)$, and the product $f(4|\theta)\,g(\theta)$.

Figure 2 A normal likelihood (red) a normal prior (blue) and their product (green)



A Fairly Simple Continuous Example (Posterior)

The product $f\left(4|\theta\right)g\left(\theta\right)$ is proportional to a normal probability density. Qualitatively, it seems like this normal density may have a mean somewhere between the locations of the peaks of of $f\left(4|\theta\right)$ and $g\left(\theta\right)$, and this graphic indicates how prior and likelihood combine to produce the posterior.

The exact general result illustrated by Figure 2 is that the posterior in this model is again normal with

$$\text{posterior mean} = \left(\frac{1}{1+\frac{1}{\gamma^2}}\right)x + \left(\frac{\frac{1}{\gamma^2}}{1+\frac{1}{\gamma^2}}\right)m$$

and

$$\text{posterior variance} = \frac{1}{1+\frac{1}{\gamma^2}}$$

A Fairly Simple Continuous Example (Posterior)

This is an intuitively appealing result in the following way. First, it is common in Bayes contexts to call the reciprocal of a variance a "precision." So in this model, the likelihood has precision

$$\frac{1}{1} = 1$$

while the "prior precision" is

$$\frac{1}{\gamma^2}$$

The sum of these is

$$1 + \frac{1}{\gamma^2}$$

so that the posterior precision is the sum of the precisions of the likelihood and the prior (overall/posterior precision comes from both data and prior in an "additive" way). Further, the posterior mean is a weighted average of the observed value, x, and the prior mean, m, *where the weights are proportional to the respective precisions of the likelihood and the prior*.

The Qualitative Nature of Bayes Analyses

The relationships between precisions of prior, likelihood and posterior, and how the first two "weight" the data and prior in production of a posterior are, in their exact form, special to this kind of "normal-normal" model. But they indicate how Bayes analyses generally go. "Flat"/ "uninformative"/large-variance/small-precision priors allow the data to drive an analysis. "Sharp"/"peaked"/"informative"/small-variance/ large-precision priors weight prior beliefs strongly and require strong sample evidence to move an analysis off of those beliefs.

A Fairly Simple Continuous Example (Prediction)

To continue with the example

$$X \sim \mathrm{N}(\theta, 1) \text{ and } \theta \sim \mathrm{N}\left(m, \gamma^2\right)$$

there remains the matter of a posterior predictive distribution for X_{new} (e.g. from the same distribution as X). If given θ it makes sense to model X and X_{new} as independent draws from the same $\mathrm{N}(\theta, 1)$ distribution, it's easy to figure out an appropriate predictive posterior for X_{new} given $X = x$. That is, we know what the posterior of θ is, and X_{new} is just "θ plus $\mathrm{N}(0, 1)$ noise." That is, the predictive distribution is

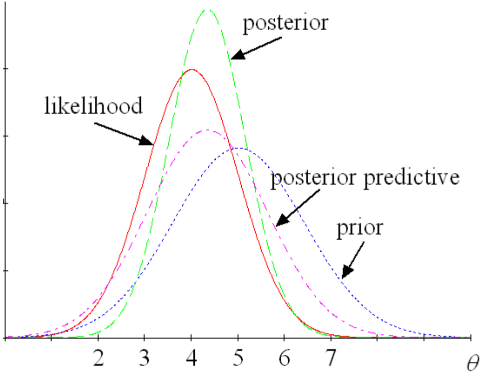
$$X_{\text{new}} | X = x \ \sim \ \mathrm{N}\left(\left(\frac{1}{1+\frac{1}{\gamma^2}}\right)x + \left(\frac{\frac{1}{\gamma^2}}{1+\frac{1}{\gamma^2}}\right)m, 1 + \frac{1}{1+\frac{1}{\gamma^2}}\right)$$

i.e. it is again normal with the same posterior mean as θ, but a variance increased from the posterior variance of θ by 1 (the variance of an individual observation). This is again qualitatively right. One can not know more about X_{new} than one knows about θ.

A Fairly Simple Continuous Example (Summary)

To make this all concrete, consider once again a case where the prior is $\mathrm{N}(5, 2)$ and one observes $X = 4$. Figure 3 summarizes this example.

Figure 3 A normal likelihood if $X = 4$ (red), a prior density for θ (blue), the corresponding posterior for θ (green), and the corresponding predictive distribution for X_{new} (violet)



Bayes Computation in Practice

Except for the very simplest models of the type we've just used as examples, people can't do the pencil and paper calculus needed to apply Bayes methods. Happily, in the last 20 years or so, modern methods of so-called "Markov chain Monte Carlo" (a type of stochastic simulation) have been developed and allow people to sample from posteriors

$$\text{posterior} \propto \text{likelihood} \cdot \text{prior}$$

even for very high-dimensional parameter vectors $\boldsymbol{\theta}$ and very complicated data models and priors *and use empirical properties of the samples to approximate theoretical properties of the posteriors*. In fact, truth be told, Bayesians can now fairly routinely handle huge problems for which no corresponding methods based only on likelihood theory are worked out or implemented in software.

"Real" professional Bayesians program their own MCMC algorithms, tailoring them to the models and data sets they face. The most widely used Bayes software available for non-specialists like you and me derives from the Biostatistics Unit at Cambridge University. The ^{TM}`Windows` version is `WinBUGS` and there is an open source version (that can be run in batch mode) called `OPENBUGS`. We'll illustrate `WinBUGS` in the balance of this introduction. `WinBUGS` has its own user manual and its own discussion list and you are referred to those sources for further information (and to the Stat 544 web pages).

Using WinBUGS

In order to make a `WinBUGS` analysis, one must

- write and have the software **check** the syntax of a **model** statement for the problem,
- **load** any **data** needed for the analysis not specified in the model statement,
- **compile** the program that will run the Gibbs sampler, and
- one way or another (either by supplying them or by generating them from the model itself) provide **initial values** for the sampler(s)/ chain(s) that will be run.

One then

- **updates** the sampler(s) as appropriate,
- **monitors** the progress of the sampler(s), and
- ultimately **summarizes** what the sampler(s) indicate about the posterior distribution(s).

WinBUGS Example 1

As a first example, we will do the `WinBUGS` version of the small normal-normal model. The code for this is in the file

`BayesASQEx1.odc`

on Vardmen's web page. Remember that the model is

$$\begin{aligned} X &\sim \mathrm{N}(\theta, 1) \\ \theta &\sim \mathrm{N}(5, 2) \end{aligned}$$

(the prior variance is 2, so that the prior precision is .5) and we are assuming that $X = 4$ is observed.

Example 1 (cont.)

Here is the code:

```
model {
X~dnorm(theta,1)
Xnew~dnorm(theta,1)
theta~dnorm(5,.5)
#WinBUGS uses the precision instead of the variance or
#standard deviation to name its normal distributions
#so the prior variance of 2 is expressed as a prior
#precision of .5
}
#here is a list of data for this example
list(X=4.0)
#here are 4 possible initializations for Gibbs samplers
list(theta=7,Xnew=3)
list(theta=2,Xnew=6)
list(theta=3,Xnew=10)
list(theta=8,Xnew=10)
```

Example 1 (cont.)

The model is specified using the `Specification Tool` under the `Model` menu. One first uses the `check model` function, then the `load data` function to enter the `list(X=4.0)`, then (for example choosing to run 4 parallel Gibbs samplers) employs the `compile` function. To initialize the simulation, one may either ask `WinBUGS` to generate initial values from the model, or one at a time enter 4 initializations for the chains like those provided above.

The `Update Tool` on the `Model` menu is used to get `WinBUGS` to do Gibbs updates of a current sampler state (in this case, a current θ and value for X_{new}). The progress of the iterations can be watched and summarizations of the resulting simulated parameters (and new observations) can be produced using the `Sample Monitor Tool` under the `Inference` menu.

Example 1 (cont.)

Here are screen shots of what one gets for summaries of a fairly large number of iterations using the `history`, `density`, and `stats` functions of the `Sample Monitor Tool`. (One must first use the `set` function before updating, in order to alert `WinBUGS` to the fact that values of θ and X_{new} should be collected for summarization.)

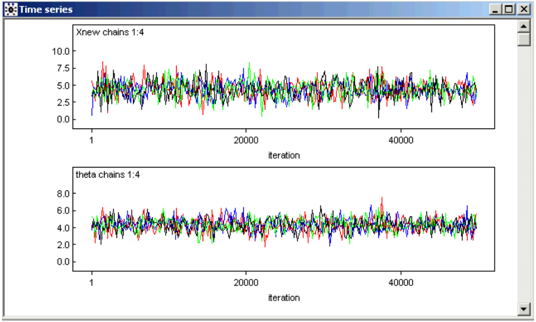
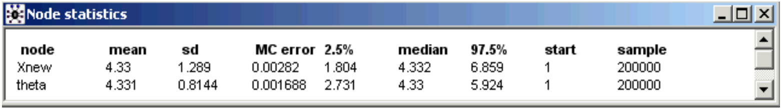


Figure: History plots for 50,000 iterations for 4 parallel chains (thinned to every 200th iteration for plotting purposes) for the toy normal-normal problem



Node statistics

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
Xnew	4.33	1.289	0.00282	1.804	4.332	6.859	1	200000
theta	4.331	0.8144	0.001688	2.731	4.33	5.924	1	200000

Figure: Summary statistics for $50,000$ iterations for 4 parallel chains for the toy normal-normal problem

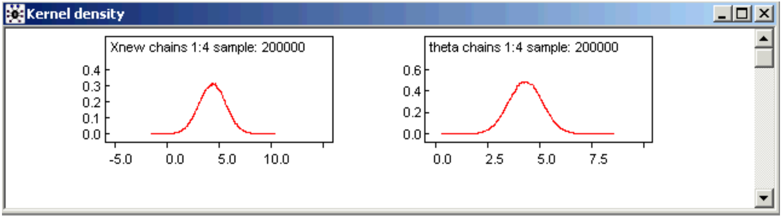


Figure: Approximate densities estimated from $50,000$ iterations for 4 parallel chains for the toy normal-normal problem

Example 1 (cont.)

The first figure shows no obvious differences in the behaviors of the 4 chains (started from the fairly "dispersed" initializations indicated in the example code), which is of comfort if one is worried about the possibility of Gibbs sampling failing. The last 2 figures are in complete agreement with the pencil and paper analyses of this problem. Both the posterior for θ and the posterior predictive distribution of X_{new} look roughly "normal" and the means and standard deviations listed in the "node statistics" are completely in line with posterior means and standard deviations. In fact, these can be listed in tabular form for comparison purposes as on the next panel.

Example 1 (cont.)

Table 1 Theoretical (Pencil and Paper Calculus) and MCMC (Gibbs Sampling) Means and Standard Deviations for the Toy Normal-Normal Example

	θ	X_{new}
Theoretical Posterior Mean	4.333	4.333
MCMC Posterior Mean	4.331	4.333
Theoretical Posterior Std Dev	$\sqrt{\frac{2}{3}} = .8165$	$\sqrt{\frac{2}{3}+1} = 1.291$
MCMC Posterior Std Dev	.8144	1.289

This first example is a very "tame" example, the effect of the starting value for the Gibbs sampling is not important, and the samplers very easily produce the right posteriors.

One of the real powers of simulation as a way of approximating a posterior is that it is absolutely straightforward to approximate the posterior distribution of any function of the parameter vector θ, say $h(\theta)$. One simply plugs simulated values of θ into the function observes the resulting relative frequency distribution.

Example 1 (cont.)

Continuing the normal-normal example, a function of θ that could potentially be of interest is the fraction of the X distribution below some fixed value, say 3.0. (This kind of thing might be of interest if X were some part dimension and 3.0 were a lower specification for that dimension.) In this situation, the parametric function of interest is

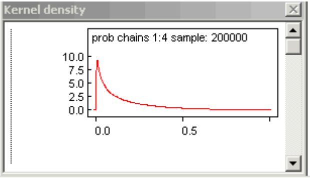
$$h(\theta) = \Phi\left(\frac{3.0-\theta}{1}\right) = P\left[Z \leq \frac{3.0-\theta}{1}\right]$$

and by simply adding the line of code

```
prob<-phi(3.0-theta)
```

to the previous model statement, it is easy to get a picture of the posterior distribution of the fraction of the X distribution below 3.0 similar to that in the figure on the next panel.

Example 1 (cont.)



Node statistics

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
prob	0.1511	0.1636	3.455E-4	0.001726	0.09175	0.6062	1	200000

Figure: The posterior distribution of $prob = \Phi(3.0 - \theta)$

The posterior mean for this fraction of the X distribution is about 15%, but very little is actually known about the quantity. If one wanted 95% posterior probability of bracketing $prob = \Phi(3.0 - \theta)$, a so-called 95% Bayes *credible interval* (running from the lower 2.5% point to the upper 2.5% point of the approximate posterior distribution) would be

$$(.001726, .6062)$$

WinBUGS Example 2

As a second simple well-behaved example, consider a fraction non-conforming context, where one is interested in

$$X = \text{the number non-conforming in a sample of } n = 50$$

and believes *a priori* that

$$p = \text{the process non-conforming rate}$$

producing X might be appropriately described as having mean .04 and standard deviation .04.

The model for the observable data here will be

$$X \sim \text{Binomial}\,(50, p)$$

Example 2 (cont.)

The figure below shows a convenient prior density for p that has the desired mean and standard deviation. This is the so-called Beta distribution with parameters $\alpha = .92$ and $\beta = 22.08$.

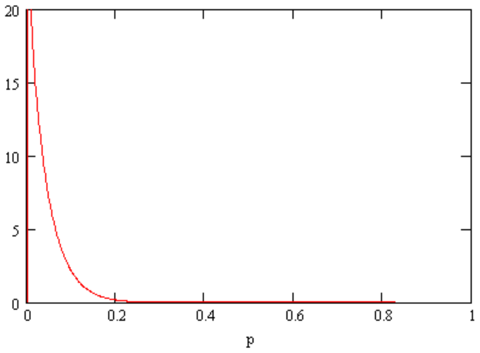


Figure: Beta density with parameters $\alpha = .92$ and $\beta = 22.08$

Example 2 (cont.)

The general Beta density is

$$g(p) \propto p^{\alpha-1}(1-p)^{\beta-1}$$

and it is common in Bayes contexts to think of such a prior as contributing to an analysis information roughly equivalent to α "successes" (non-conforming items in the present context) and β "failures" (conforming items in the present context). So employing a Beta$(.92, 22.08)$ prior here is roughly equivalent to assuming prior information that a single non-conforming item has been seen in 23 inspected items.

The code in the file

`BayesASQEx2.odc`

can be used to find a $X = 4$ posterior distribution for p and posterior predictive distribution for

$$X_{\text{new}} = \text{the number non-conforming in the next 1000 produced}$$

Example 2 (cont.)

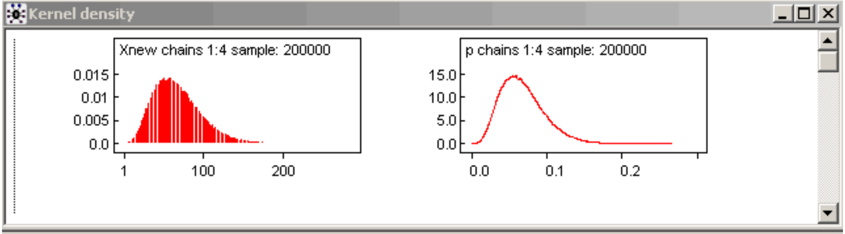
(One is, of course, assuming the stability of the process at the current p.) The code is

```
model {
 X~dbin(p,50)
 p~dbeta(.92,22.08)
 Xnew~dbin(p,1000)
}
#here are the data for the problem
list(X=4)
#here are 4 possible initializations for Gibbs samplers
list(p=.1,Xnew=50)
list(p=.5,Xnew=20)
list(p=.001,Xnew=30)
list(p=.7,Xnew=2)
```

Example 2 (cont.)

This is a "tame" problem (that could actually be solved completely by pencil and paper) and the 4 initializations all yield the same view of the posterior. The figure on the next panel provides an approximate view of the posterior distribution of p and the posterior predictive distribution of the number of non-conforming items among the next 1000 produced. (Note that in retrospect, it is no surprise that these distributions have essentially the same shape. $n = 1000$ is big enough that we should expect the sample fraction non-conforming among the 1000 to be about p, whatever that number might be. Thus X_{new} should have a posterior predictive distribution much like a posterior for $1000p$.)

Example 2 (cont.)



Node statistics

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
Xnew	67.39	30.19	0.06902	21.0	63.0	137.0	1	200000
p	0.0674	0.02914	6.648E-5	0.02232	0.06358	0.1347	1	200000

Figure: Approximate posterior distributions for p and for X_{new} (based on $n = 1000$) upon observing $X = 4$ non-conforming among 50 using a $\text{Beta}(.92, 22.08)$ prior

WinBUGS Example 3

We finish up with a several interesting multi-sample examples of Bayes analyses. The first is a "one way random effects analysis" example.

In a 2006 *Quality Engineering* paper "Calibration, Error Analysis, and Ongoing Measurement Process Monitoring for Mass Spectrometry," Vardeman, Wendelberger, and Wang discuss a Bayes analysis of 44 measurement of a spectrometer's sensitivity to Argon gas, made across 3 days. With

$$DS_{tj} = \text{the device sensitivity computed from specimen } j \text{ on day } t$$

they used a decomposition

$$DS_{tj} = \mu_{\text{S}} + \delta_t + \epsilon_{tj}$$

for μ_{S} a fixed unknown "true" device sensitivity, δ_t a random "day t" deviation in sensitivity, and ϵ_{tj} a random specimen deviation.

Example 3 (Data)

Here are the data for the study (device sensitivities, where units are $\mathrm{mol}/\ \mathrm{A\,s}$).

Day 1		Day 2		Day 3	
31.3	27.8	32.5	30.5	31.7	28.3
31.0	28.2	32.2	28.4	29.8	28.3
29.4	28.4	31.9	28.5	29.6	28.3
29.2	28.7	30.2	28.8	29.0	29.2
29.0	29.7	30.2	28.8	28.8	29.7
28.8	30.8	29.5	30.6	29.6	31.1
28.8	30.1	30.8	31.0	28.9	
27.7	29.9				
27.7					

Example 3 (Data Model)

Assuming that the δ_t are independent draws from a normal distribution with mean 0 and standard deviation σ_δ, independent of the ϵ_{tj} that are independent random draws from a normal distribution with mean 0 and standard deviation σ, this is a problem with parameters μ_S, σ_δ, and σ.

The model can be rephrased as

$$\mu_t \equiv \mu_S + \delta_t = \text{the day } t \text{ sensitivity} \quad \sim \mathrm{N}\left(\mu_S, \sigma_\delta^2\right) \text{ for } t = 1, 2, 3$$

and given the values of μ_t, the specimen sensitivities are

$$DS_{tj} = \mu_S + \delta_t + \epsilon_{tj} \quad \sim \mathrm{N}\left(\mu_t, \sigma^2\right)$$

Example 3 (Prior)

We considered the specification of a prior for the parameters $\mu_{\mathsf{S}}, \sigma_{\delta}$, and σ,

$$\begin{aligned} \mu_{\mathsf{S}} &\sim \mathsf{N}\left(0, 10^6\right) \text{ independent of} \\ \frac{1}{\sigma_{\delta}^2} &\sim \mathsf{Gamma}\left(.001, .001\right) \text{ independent of} \\ \frac{1}{\sigma^2} &\sim \mathsf{Gamma}\left(.001, .001\right) \end{aligned}$$

These were intended to be relatively non-informative (but nevertheless proper) priors for the parameters.

The `WinBUGS` code used in the paper is in the file

`BayesASQEx7.odc`

and listed on the next two panels.

Example 3 (WinBUGS Code)

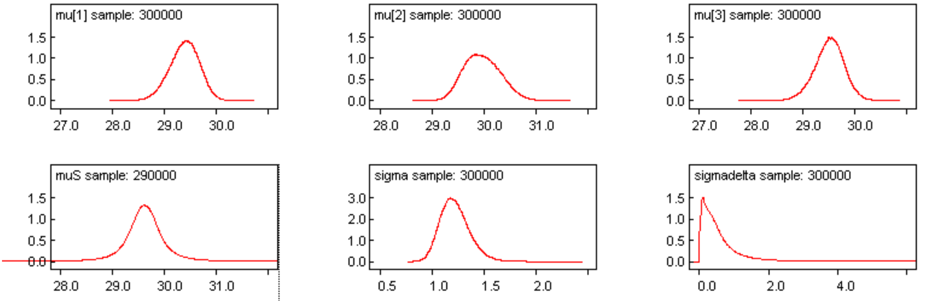
```
list(sens=c(31.3,31.0,29.4,29.2,29.0,28.8,28.8,27.7,
27.7,27.8,28.2,28.4,28.7,29.7,30.8,30.1,29.9,32.5,32.2,
31.9,30.2,30.2,29.5,30.8,30.5,28.4,28.5,28.8,28.8,30.6,
31.0,31.7,29.8,29.6,29.0,28.8,29.6,28.9,28.3,28.3,28.3,
29.2,29.7,31.1),ind=c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,
2,2,2,2,2,2,2,2,2,2,2,2,2,2,3,3,3,3,3,3,3,3,3,3,3,3,3),
N=44)

list(mu=c(3,3,3),tau=1,muS=0,taudelta=1)
```

Example 3 (WinBUGS Code cont.)

```
model {
 for(i in 1:N) {
 sens[i]~dnorm(mu[ind[i]],tau)
 }
 for(i in 1:3) {
 mu[i]~dnorm(muS,taudelta)
 }
 tau~dgamma(0.001,0.001)
 sigma<-1/sqrt(tau)
 muS~dnorm(0.0,1.0E-6)
 taudelta~dgamma(0.001,0.001)
 sigmadelta<-1/sqrt(taudelta)
}
```

The next figure shows some summaries from a `WinBUGS` session based on the code.



Node statistics

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
mu[1]	29.38	0.284	0.001202	28.8	29.4	29.9	11001	300000
mu[2]	29.97	0.347	0.001801	29.35	29.95	30.68	11001	300000
mu[3]	29.51	0.2842	9.391E-4	28.91	29.52	30.04	11001	300000
muS	29.62	0.9781	0.002014	28.55	29.62	30.68	21001	290000
sigma	1.221	0.1397	3.262E-4	0.9827	1.209	1.528	11001	300000
sigmadelta	0.6353	1.825	0.007307	0.03671	0.3903	2.685	11001	300000

Figure: Summaries of a Bayes analysis of a spectrometer's sensitivity to Argon gas

Notice that the mean(s) in this problem are far more precisely determined than are the standard deviations. That is a well known phenomenon. It takes a very large sample size to make definitive statements about variances or standard deviations ... and in the case of σ_δ, the appropriate "sample size" is 3! (Tests were made on only 3 days.)

Example 4

Two versions of an industrial process are run with the intention of comparing effectiveness. (There is an "old/#1" and a "new/#2" process.) Six different batches of raw material are used in the study. For

$$y_{ij} = \text{the yield of the } j\text{th run made using raw material batch } i \quad ,$$

process data are below.

i	j	process	y_{ij}	i	j	process	y_{ij}	i	j	process	y_{ij}
1	1	1	82.72	4	1	1	87.77	5	7	2	78.23
1	2	1	78.31	4	2	1	84.42	5	8	2	76.40
1	3	1	82.20	4	3	1	84.82	6	1	2	81.64
1	4	1	81.18	5	1	1	78.61	6	2	2	83.04
2	1	1	80.06	5	2	1	77.47	6	3	2	82.40
2	2	1	81.09	5	3	1	77.80	6	4	2	81.93
3	1	1	78.71	5	4	1	81.58	6	5	2	82.96
3	2	1	77.48	5	5	1	77.50				
3	3	1	76.06	5	6	2	78.73				

We'll consider an analysis based on the model

$$y_{ij} = \mu_{\text{process}(ij)} + \beta_i + \varepsilon_{ij}$$

where $\text{process}(ij)$ takes the value either 1 or 2 depending upon which version of the process is used, μ_1 and μ_2 are mean yields for the two versions of the process, the β_i are $\text{N}\left(0, \sigma_\beta^2\right)$ independent of the ε_{ij} which are $\text{N}\left(0, \sigma^2\right)$, and the parameters of the model are $\mu_1, \mu_2, \sigma_\beta,$ and σ.

This is (by the way) a so-called "mixed effects model." (The μ's are "fixed effects," of interest in their own right. The β's are "random effects" primarily of interest for what they tell us about σ_β, that measures random batch-to-batch variability.)

Example 4 (Priors)

What was presumably a fairly "non-informative" choice of prior distribution for the model parameters was

$$\begin{aligned}
\mu_1 &\sim \text{N}\left(0, 10^6\right) \text{ independent of} \\
\mu_2 &\sim \text{N}\left(0, 10^6\right) \text{ independent of} \\
\ln\left(\sigma_\beta\right) &\sim \text{"Uniform } \left(-\infty, \infty\right)\text{"/"flat" independent of} \\
\ln\left(\sigma\right) &\sim \text{"Uniform } \left(-\infty, \infty\right)\text{"/"flat"}
\end{aligned}$$

`WinBUGS` code for analysis of this situation is in the file

`BayesASQEx8.odc`

and listed on the next panel two panels.

Example 4 (WinBUGS Code)

```
model {
for (i in 1:2) {process[i]~dnorm(0,.000001)}
diff<-process[2]-process[1]
logsigb~dflat()
taubatch<-exp(-2*logsigb)
sigmabatch<-exp(logsigb)
for (j in 1:7) {batch[j]~dnorm(0,taubatch)}
logsig~dflat()
tau<-exp(-2*logsig)
sigma<- exp(logsig)
for (l in 1:27) {mu[l]<-process[p[l]]+batch[b[l]]}
for (l in 1:27) {y[l]~dnorm(mu[l],tau)}
}
```

Example 4 (WinBUGS Code cont.)

```
list(b=c(1,1,1,1,2,2,3,3,3,4,4,4,5,5,5,5,
5,5,5,5,6,6,6,6,6,7,7),
p=c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,2,2,
2,2,2,2,2,2,1,2),
y=c(82.72,78.31,82.20,81.18,80.06,81.09,
78.71,77.48,76.06,87.77,84.42,84.82,78.61,
77.47,77.80,81.58,77.50,78.73,78.23,76.40,
81.64,83.04,82.40,81.93,82.96,NA,NA))
```

Note that the code calls for simulation of new responses from a 7th batch under both of the possible process conditions (the 26th and 27th y's). These come from posterior predictive distributions.

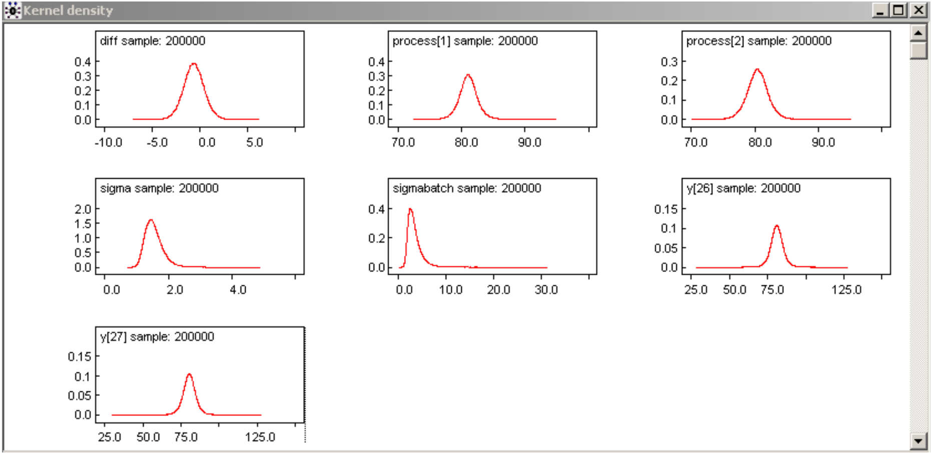
Example 4 (cont.)

Here is WinBUGS output that shows clearly that not enough has been learned from these data to say definitively how the processes compare.

Node statistics

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
batch[1]	0.01095	1.678	0.02521	-3.434	0.03386	3.276	7001	200000
batch[2]	-0.4584	1.77	0.02425	-4.081	-0.4266	2.965	7001	200000
batch[3]	-3.348	1.735	0.02495	-6.991	-3.278	-0.1003	7001	200000
batch[4]	4.165	1.729	0.02479	0.7914	4.142	7.657	7001	200000
batch[5]	-2.459	1.615	0.02603	-5.798	-2.421	0.6801	7001	200000
batch[6]	1.851	1.818	0.02661	-1.755	1.841	5.487	7001	200000
batch[7]	-0.007046	3.744	0.008558	-7.521	-0.01226	7.522	7001	200000
diff	-0.6555	1.082	0.005232	-2.794	-0.6595	1.497	7001	200000
process[1]	81.09	1.567	0.02558	78.03	81.06	84.34	7001	200000
process[2]	80.43	1.755	0.02696	76.96	80.42	83.96	7001	200000
sigma	1.541	0.2764	8.625E-4	1.112	1.504	2.187	7001	200000
sigmabatch	3.429	1.5	0.01263	1.646	3.091	7.238	7001	200000
y[26]	81.08	4.348	0.02703	72.52	81.04	89.84	7001	200000
y[27]	80.43	4.426	0.0282	71.65	80.42	89.28	7001	200000

Example 4 (cont.)



Example 5

As a final example of the power of Bayes analysis to handle what would otherwise be quite non-standard statistical problems, consider the following situation.

A response, y, has mean known to increase with a covariate/predictor, x, and is investigated in a study where (coded) values $x = 1, 2, 3, 4, 5, 6$ are used and there are 3 observations for each level of the predictor. Suppose the form of the dependence of the mean of y on x (say $\mathrm{E}y_{xj} = \mu_x$) is not something that we wish to specify beyond the restriction that

$$\mu_1 \leq \mu_2 \leq \mu_3 \leq \mu_4 \leq \mu_5 \leq \mu_6$$

(So, for example, simple linear regression is not an appropriate statistical methodology for analyzing the dependence of mean y on x.) We consider an analysis based on a model

$$y_{xj} = \mu_x + \varepsilon_{xj}$$

where the (otherwise completely unknown) means $\mu_1, \mu_2, \ldots, \mu_6$ satisfy the order restriction and the ε_{xj} are independent $\mathrm{N}\left(0, \sigma^2\right)$ variables.

Example 5 ("Data")

Here are some hypothetical data and summary statistics for this example.

x	y_{xj}'s	$\bar{y}_x$
1	$0.9835899, -0.5087186, 1.0450089$	$.51$
2	$0.6815755, -2.1739497, 1.0464128$	$-.15$
3	$1.3717484, 1.1350734, 0.3384970$	$.95$
4	$6.5645035, 5.0648255, 6.0209295$	5.88
5	$6.5766160, 5.8730637, 7.4934093$	6.65
6	$7.8030626, 8.2207331, 6.7444797$	7.59

and $\sqrt{MSE} = .99$.

Example 5 (Prior)

One way to make a simple WinBUGS analysis of this problem is to set

$$\delta_i = \mu_i - \mu_{i-1} \text{ for } i = 2, 3, \ldots, 6$$

and use priors

$$\begin{aligned} \mu_1 &\sim \text{N}\left(0, 10^4\right) \text{ independent of} \\ \delta_i &\sim \text{Uniform}\left(0, 10\right) \text{ for } i = 2, \ldots, 6 \text{ independent of} \\ \ln(\sigma) &\sim \text{"Uniform } (-\infty, \infty)\text{"/"flat"} \end{aligned}$$

WinBUGS code for analysis of this situation is in the file

`BayesASQEx9.odc`

and is listed on the next two panels.

Example 5 (WinBUGS Code)

```
model {
 mu1 ~dnorm(0,.0001)
 mu[1] <- mu1
 for (i in 2:6) {
 delta[i] ~dunif(0,10)}
 for (i in 2:6) {
 mu[i] <- mu[i-1]+delta[i]}
 logsigma ~dflat()
 sigma <- exp(logsigma)
 tau <- exp(-2*logsigma)
 for (j in 1:N) {
 y[j] ~dnorm(mu[group[j]],tau)}
}
```

Example 5 (WinBUGS Code cont.)

```
list(N=18,group=c(1,1,1,2,2,2,3,3,3,4,4,4,5,5,5,6,6,6),
y=c(0.9835899,-0.5087186,1.0450089,0.6815755,-2.1739497,
1.0464128,1.3717484,1.1350734,0.3384970,6.5645035,
5.0648255,6.0209295,6.5766160,5.8730637,7.4934093,
7.8030626,8.2207331,6.7444797))
```

```
list(mu1=0,logsigma=0)
```

Output on the next two panels shows how easy it is to get sensible inferences in this nonstandard problem from a `WinBUGS` Bayes analysis.

Example 5 (cont.)

Node statistics

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
delta[2]	0.4471	0.4005	0.001279	0.01344	0.3385	1.485	10001	500000
delta[3]	0.8565	0.5797	0.002075	0.04612	0.7695	2.199	10001	500000
delta[4]	4.557	0.7806	0.002791	2.88	4.605	5.96	10001	500000
delta[5]	0.9683	0.6314	0.001778	0.05662	0.8845	2.405	10001	500000
delta[6]	1.048	0.6514	0.001226	0.06984	0.973	2.504	10001	500000
mu[1]	-0.1484	0.4775	0.001551	-1.181	-0.1174	0.7098	10001	500000
mu[2]	0.2987	0.4397	0.00115	-0.5765	0.2995	1.172	10001	500000
mu[3]	1.155	0.5353	0.001638	0.2069	1.119	2.314	10001	500000
mu[4]	5.712	0.5521	0.001483	4.542	5.742	6.714	10001	500000
mu[5]	6.681	0.4983	8.5E-4	5.701	6.679	7.666	10001	500000
mu[6]	7.729	0.5604	6.954E-4	6.707	7.701	8.916	10001	500000
sigma	1.062	0.2289	5.746E-4	0.7249	1.025	1.61	10001	500000

Example 5 (cont.)