

Module 6

Some one-sample normal examples

Prof. Stephen B. Vardeman
Statistics and IMSE
Iowa State University

March 5, 2008

Example 4

We continue with our illustration of what is possible using WinBUGS to do Bayes analyses, in this module treating some problems involving normal observations with both μ and σ unknown. A real example on page 771 of Vardeman and Jobe's *Basic Engineering Data Collection and Analysis* concerns measurements on a critical dimension of $n = 5$ consecutive parts produced on a CNC Lathe. The measurements in units of .0001 in over nominal were

4, 3, 3, 2, 3

We note that the usual sample mean for these values is $\bar{x} = 3.0$ in and the usual sample standard deviation is $s = .7071$ in.

To begin, we're going to first do what most textbooks would do here, and consider analysis of these data under the model that says that the data are realized values of

X_1, X_2, \dots, X_5 independent $N(\mu, \sigma^2)$ random variables

Example 4 (cont.)

Standard "Stat 101" confidence limits for the process mean μ are

$$\bar{x} \pm t_{n-1} \frac{s}{\sqrt{n}},$$

while confidence limits for the process standard deviation σ are

$$\left(s \sqrt{\frac{n-1}{\chi_{\text{upper}, n-1}^2}}, s \sqrt{\frac{n-1}{\chi_{\text{lower}, n-1}^2}} \right),$$

and prediction limits for an additional dimension generated by this process, x_{new} , are

$$\bar{x} \pm t_{n-1} s \sqrt{1 + \frac{1}{n}}.$$

In the present problem, these yield 95% limits as in Table 1 on the next panel.

Example 4 (cont.)

Table 1 Standard (non-Bayesian) 95% Confidence and Prediction Limits Based on a Normal Assumption for the Part Measurements

Quantity	lower limit	upper limit
μ	$3.0 - 2.776 \frac{.7071}{\sqrt{5}} = 2.12$	3.88
σ	$.7071 \sqrt{\frac{4}{11.143}} = .42$	$.7071 \sqrt{\frac{4}{.484}} = 2.03$
x_{new}	$3.0 - 2.776 (.7071) \sqrt{1 + \frac{1}{5}} = .85$	5.15

Bayesians often try to find relatively "non-informative" priors for which inferences based on their posteriors are in substantial agreement with non-Bayes methods of inference for problems where the latter are available. This gives hope that when such priors are used as parts of more complicated models (some for which no non-Bayesian methods of inference are even known) they will give inferences not too heavily affected by the choice of prior. We now consider what are fairly standard choices of priors in one-sample normal problems.

Example 4 (cont.)

A standard "noninformative" choice of prior distribution for a one-sample normal problem is

$$\begin{aligned}\mu &\sim \text{"Uniform } (-\infty, \infty) \text{" independent of} \\ \ln(\sigma) &\sim \text{"Uniform } (-\infty, \infty) \text{"}\end{aligned}$$

Strictly speaking this is nonsense, as there is no such thing as a uniform probability distribution on $(-\infty, \infty)$. Precisely what is meant is that with $\delta = \ln(\sigma)$ one is using as a "prior" for (μ, δ) a function

$$g(\mu, \delta) = 1$$

This turns out to be equivalent to using as a "prior" for (μ, σ^2) a function

$$g(\mu, \sigma^2) \propto \frac{1}{\sigma^2}$$

Example 4 (cont.)

These don't actually specify a joint probability distribution for (μ, σ^2) , but can be thought of as approximately equivalent to the "honest" priors

$$\begin{aligned}\mu &\sim \text{Uniform}(-large, large) \text{ independent of} \\ \ln(\sigma) &\sim \text{Uniform}(-large, large)\end{aligned}$$

or

$$\begin{aligned}\mu &\sim N(0, large) \text{ independent of} \\ \frac{1}{\sigma^2} &\sim \text{Gamma}(small, small)\end{aligned}$$

WinBUGS allows the use of the "improper" "Uniform $(-\infty, \infty)$ " prior through the notation

`dflat()`

Example 4 (cont.)

The file

BayesASQEx4A.odc

contains WinBUGS code below for implementing a Bayes analysis for these data based on independent "flat" priors for μ and $\ln(\sigma)$.

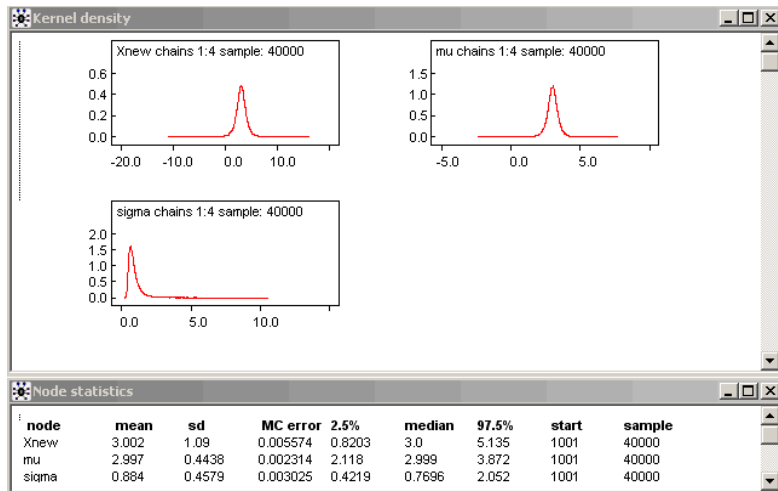
```
model {
  mu~dflat()
  logsigma~dflat()
  sigma<-exp(logsigma)
  tau<-exp(-2*logsigma)
  for (i in 1:5) {
    X[i]~dnorm(mu,tau)
  }
  Xnew~dnorm(mu,tau)
}
```

```
list(X=c(4,3,3,2,3))
#here are 4 starts
list(mu=7,logsigma=2,Xnew=7)
list(mu=7,logsigma=-2,Xnew=7)
list(mu=2,logsigma=2,Xnew=2)
list(mu=2,logsigma=-2,Xnew=2)
```

The figure on the next panel shows the result of running this code.

Example 4 (cont.)

I checked that burn-in had occurred by 1000 iterations. What is pictured are then results based on iterations 1001 through 11000.



Example 4 (cont.)

Bayes results for this "improper prior" analysis are in almost perfect agreement with the "ordinary Stat 101" results listed in Table 1. This analysis says one has the same kind of precision of information about μ , σ , and x_{new} that would be indicated by the elementary formulas.

The files

BayesASQEx4B.odc and BayesASQEx4C.odc

contain code that the reader can verify produces results not much different from those above. The priors used in these are respectively

$$\begin{aligned}\mu &\sim \text{Uniform}(-10000, 10000) \text{ independent of} \\ \ln(\sigma) &\sim \text{Uniform}(-100, 100)\end{aligned}$$

and

$$\begin{aligned}\mu &\sim N(0, 10^6) \text{ independent of} \\ \tau &= \frac{1}{\sigma^2} \sim \text{Gamma}(.01, .01)\end{aligned}$$

Example 4 (cont.)

The code for the first is:

```
model {
mu~dunif(-10000,10000)
logsigma~dunif(-100,100)
sigma<-exp(logsigma)
tau<-exp(-2*logsigma)
for (i in 1:5) {
  X[i]~dnorm(mu,tau)
}
Xnew~dnorm(mu,tau)
}
list(X=c(4,3,3,2,3))
#here are 4 starts
list(mu=7,logsigma=2,Xnew=7)
list(mu=7,logsigma=-2,Xnew=7)
list(mu=2,logsigma=2,Xnew=2)
list(mu=2,logsigma=-2,Xnew=2)
```

and the code for the second is:

```
model {
mu~dnorm(0,.000001)
tau~dgamma(.01,.01)
sigma<-sqrt(1/tau)
for (i in 1:5) {
  X[i]~dnorm(mu,tau)
}
Xnew~dnorm(mu,tau)
}
list(X=c(4,3,3,2,3))
#here are 4 starts
list(mu=7,tau=1,Xnew=7)
list(mu=7,tau=.0001,Xnew=7)
list(mu=2,tau=1,Xnew=2)
list(mu=2,tau=.0001,Xnew=2)
```

A real potential weakness of standard analyses data like those in Example 4 is that they treat what are obviously "to the nearest something" values as if they were exact (infinite number of decimal places) numbers. The "4" in the data set is treated as if it were 4.0000000000 ... Sometimes this interpretation is adequate, but other times ignoring the "digitalization"/"quantization" of measurement produces inappropriate inferences. This is true when the quantization is severe enough compared to the real variation in what is being measured that only a very few values appear in the data set. This is potentially the case in the real situation of Example 4. We consider how one might take account of quantization evident in data.

Digitalization/Quantization/Rounding (cont.)

Consider a model for recorded part dimensions that says they are (infinite number of decimal places) normal random variables rounded to the nearest integer (unit of .0001 in above nominal). A cartoon of a possible distribution of the real part dimension and the corresponding distribution of the digital measurement is below.

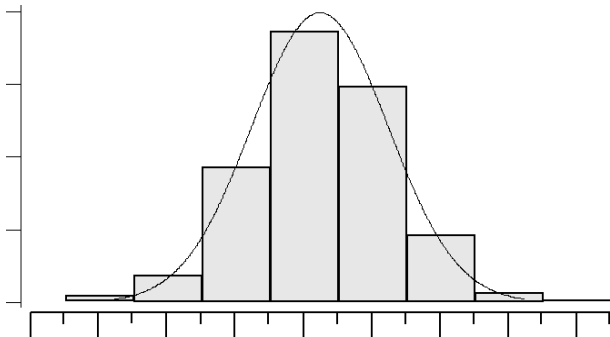


Figure:

Digitalization/Quantization/Rounding (cont.)

To model this, we might suppose that real part dimensions are realized values of

X_1, X_2, \dots, X_n independent $N(\mu, \sigma^2)$ random variables

but that data in hand are Y_1, Y_2, \dots, Y_n where

$Y_i =$ the value of X_i rounded to the nearest integer

In this understanding, the value $Y_1 = 4.0$ in a data set doesn't mean that the first measurement was exactly 4.000000000... but rather only that it was somewhere between 3.500000... and 4.500000.... What is especially pleasant is that WinBUGS makes incorporating this recognition of the digital nature of measurements into a Bayes analysis absolutely painless.

Example 5

Let's do an analysis for the situation of Example 4 with priors

$$\begin{aligned}\mu &\sim \text{"Uniform } (-\infty, \infty) \text{" independent of} \\ \ln(\sigma) &\sim \text{"Uniform } (-\infty, \infty) \text{"}\end{aligned}$$

and suppose that of interest are not only the parameters μ and σ , but also an X_{new} from the process and its rounded value Y_{new} , and in addition the fraction of the X distribution below 1.000, that is

$$h(\mu, \sigma^2) = P\left[Z < \frac{1.0 - \mu}{\sigma}\right] = \Phi\left(\frac{1.0 - \mu}{\sigma}\right)$$

There are no widely circulated non-Bayes methods of inference for any of these quantities. (Lee and Vardeman have written on non-Bayesian interval estimation for μ and σ , but these papers are not well-known.)

The file

BayesASQEx5.odc

contains some simple WinBUGS code for implementing a Bayes analysis.

Example 5 (cont.)

The code is:

```
model {  
  mu~dflat()  
  logsigma~dflat()  
  sigma<-exp(logsigma)  
  tau<-exp(-2*logsigma)  
  X1~dnorm(mu,tau) I(3.5,4.5)  
  X2~dnorm(mu,tau) I(2.5,3.5)  
  X3~dnorm(mu,tau) I(2.5,3.5)  
  X4~dnorm(mu,tau) I(1.5,2.5)  
  X5~dnorm(mu,tau) I(2.5,3.5)  
  Xnew~dnorm(mu,tau)  
  Ynew<-round(Xnew)  
  prob<-phi((1.0-mu)/sigma)  
}
```

Example 5 (cont.)

```
#here are 4 possible initializations
list(mu=7,logsigma=2,Xnew=7,X1=4,X2=4,X3=4,X4=2,X5=3)
list(mu=7,logsigma=-2,Xnew=7,X1=4,X2=4,X3=4,X4=2,X5=3)
list(mu=2,logsigma=2,Xnew=2,X1=2,,X1=4,X2=4,X3=4,X4=2,X5=3)
list(mu=2,logsigma=-2,Xnew=2,,X1=4,X2=4,X3=4,X4=2,X5=3)
```

The figure on the next panel summarizes an analysis that results from using this code.

Example 5 (cont.)

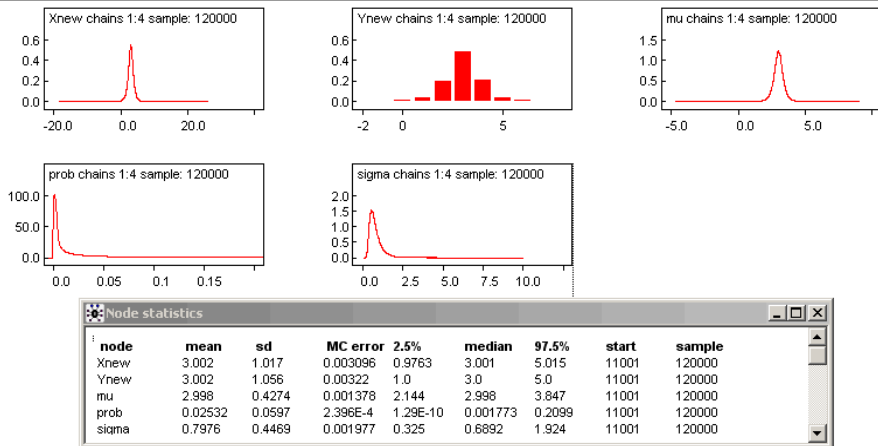


Figure: Results of a Bayes analysis for $n = 5$ part dimensions, treating what is observed as integer-rounded normal variables

Example 5 (cont.)

Perhaps the most impressive part of all this is the ease with which something as complicated as $\Phi\left(\frac{1.0-\mu}{\sigma}\right)$ is estimated and some sense of the precision with which it is known is provided ... and this taking account of rounding in the data. Note that of the parts of the analysis that correspond to things that were done earlier, it is the estimation of σ that is most affected by recognizing the quantized nature of the data. The posterior distribution here is shifted somewhat left of the one in Example 4.

This is probably a good point to remark on the fact that for a Bayesian, learning is always sequential. Today's posterior is tomorrow's prior. As long as one is willing to say that the CNC turning process that produced the parts is physically stable, the posterior from one of the analyses in this module would be appropriate as a *prior* for analysis of data at a subsequent point. So, for example, instead of starting with a flat prior for the process mean, μ , a normal distribution with mean near 3.00 and standard deviation near .44 could be considered.