Bayesian Statistics
Estimation of a Single Mean and Variance
MCMC Diagnostics and Missing Data

Michael Anderson, PhD

Department of Biostatistics and Epidemiology
The University of Oklahoma Health Sciences Center

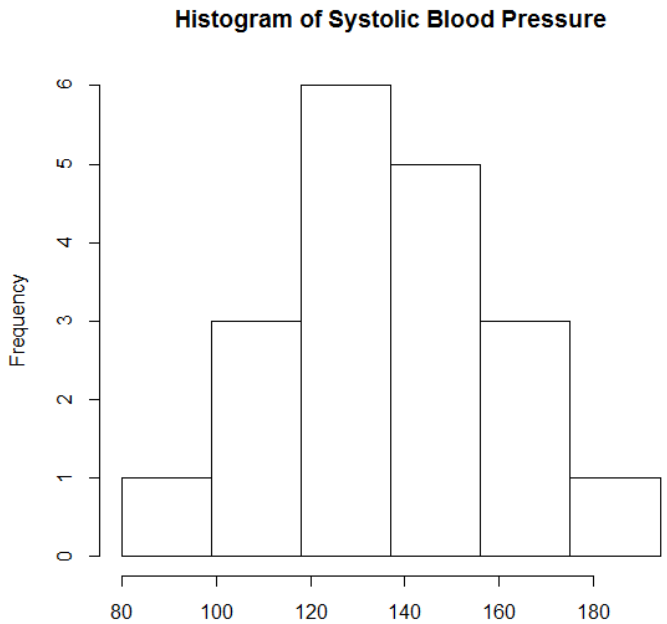
Sept 15, 2016

Outline

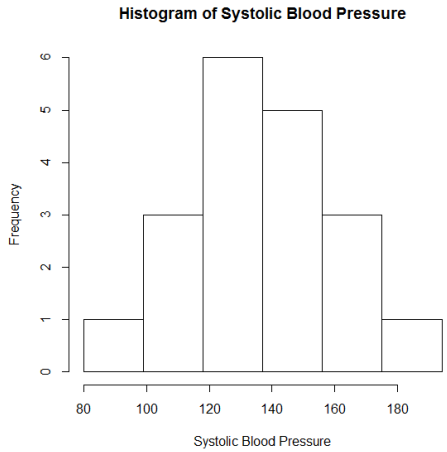
High Blood Pressure Treatment

Suppose a study examines the systolic blood pressure (SBP) of hypertensive subjects (SBP $>$ 140) after 3 months of using blood pressure medication. The SBP for 19 subjects using this medication for 3 months is given below.

```
list(N=19,sbp=c(121,94,119,122,142,168,116,172,155,
     107,180,119,157,101,145,148,120,147,125))
```



High Blood Pressure Treatment



- Q: What seems like a reasonable distribution of the **data**?
  - SBP is a continuous measure.
  - Histogram above shows rough symmetric bell-shaped form.
- A: Normal distribution seems to be a reasonable fit.
  - Shape of Normal is determined by two parameters: μ and σ^2.

We might reasonably conclude $p(x_1, \ldots, x_{19}|\mu, \sigma^2)$ is Normal(μ,σ^2).
We will seek to obtain the posterior $p(\mu, \sigma^2|x_1, \ldots, x_{19})$.
This requires specification of a joint prior, $p(\mu, \sigma^2)$.

Specification of the Likelihood and Prior

$$p(\mu, \sigma^2|x_1, \ldots, x_n) = \frac{p(x_1, \ldots, x_n|\mu, \sigma^2)p(\mu, \sigma^2)}{\int\int p(x_1, \ldots, x_n|\mu, \sigma^2)p(\mu, \sigma^2)d\mu d\sigma^2}$$

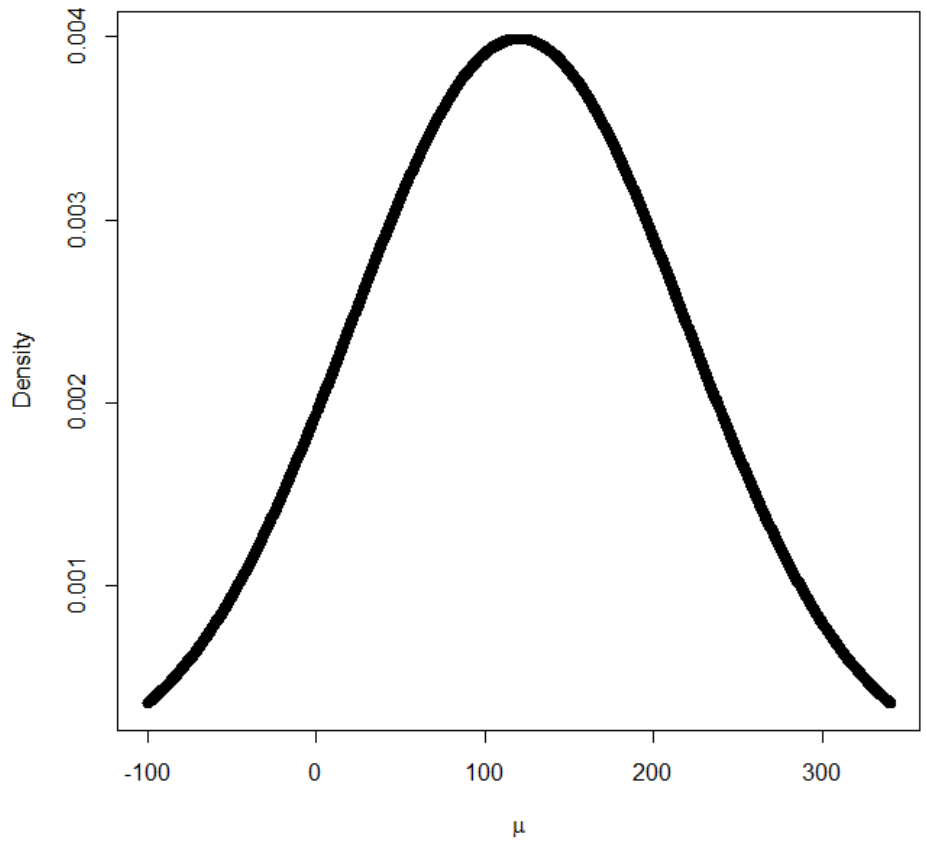
- Note if μ and σ^2 are independent then
  - $p(\mu, \sigma^2) = p(\mu)p(\sigma^2)$.
  - This means we specify a prior for μ and a separate prior for σ^2

$$p(\mu, \sigma^2|x_1, \ldots, x_n) = \frac{p(x_1, \ldots, x_n|\mu, \sigma^2)p(\mu)p(\sigma^2)}{\int\int p(x_1, \ldots, x_n|\mu, \sigma^2)p(\mu)p(\sigma^2)d\mu d\sigma^2}$$

Specification of the Likelihood and Prior

Q: What makes a reasonable prior for μ?

A: Diffuse prior on μ could be $Normal(120, 100^2)$

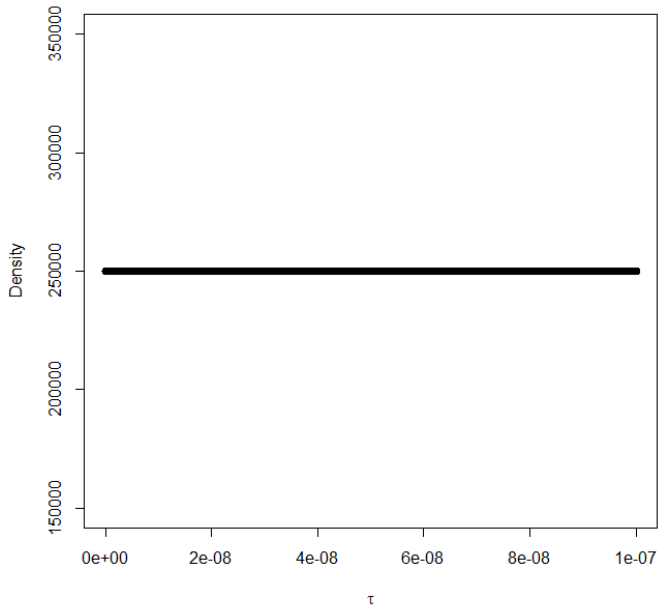


Specification of the Likelihood and Prior

Q: What makes a reasonable prior for σ^2?

A: Diffuse prior on σ could be $Unif(0, 500)$ but another popular option is a Gamma with wide variance.

- Mean of this uniform is 250.
- Variance of this uniform is $500^2/12 = 20833$.
- NOTE: JAGS requires precision, τ where $\tau = 1/\sigma^2$.



Specification of the Likelihood and Prior

Putting this altogether we have:

- $p(x_1, \ldots, x_n|\mu, \sigma^2) \sim dnorm(\mu, \tau)$
- $p(\mu) \sim dnorm(120, 0.0001)$.
- $p(\sigma) \sim unif(0, 500)$ and $\tau = 1/\sigma^2$

Let's do this for the Systolic Blood Pressure Example
See the "SBP Example.R" file

MCMC Diagnostics

- There are NO diagnostic tools in JAGS to assess how well the MCMC process has worked.
- We will use output from MCMC in R to assess these.
- The trick here will be to run 2 or more chains with `jags()`and compare behavior.
- It will be helpful to save the `jags()` object `as.mcmc()`

MCMC Diagnostics Formal

There are a few diagnostics tools in R to assess samples that have been drawn.

- Trace Plots: `plot()` Patterns are bad.
- Raftery-Lewis: `raftery.diag()` Posterior tails explored?
- Geweke: `geweke.diag()` Start of chain match end of chain?
- Gelman: `gelman.diag()` Do the chains get to the same place?
- Heidel-Berg Welch: `heidel.diag()` Chain stationary?
- Auto Correlation `autocorr.plot()` Independent samples (markov property)?
- Effective Sample Size: `effectiveSize()` Is the parameter space well traversed?

MCMC Diagnostics

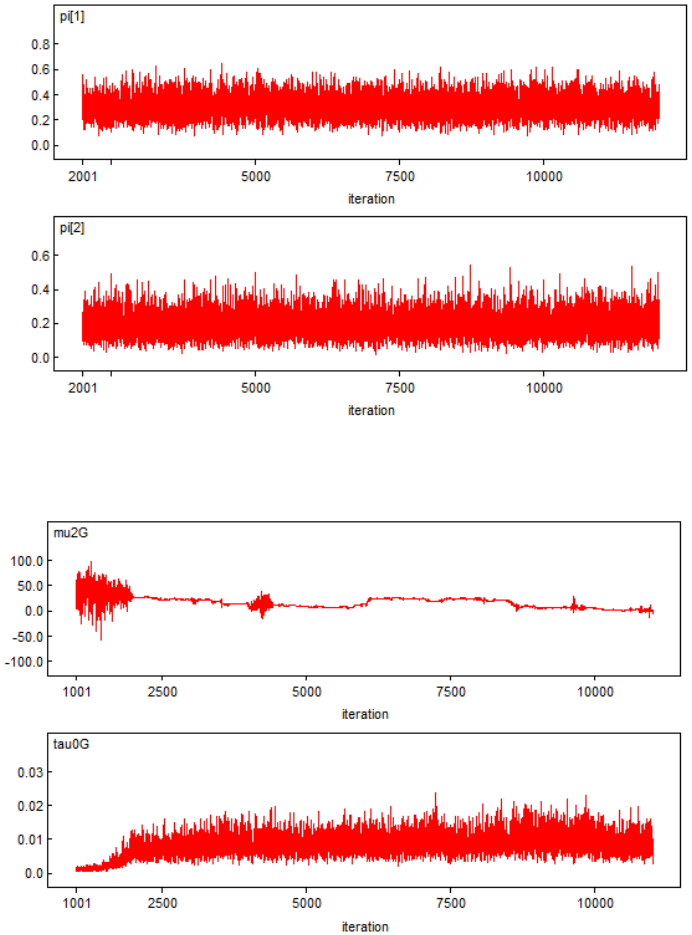
Here are a few terms that will be helpful when discussing MCMC diagnostics

- Thinning-utilizing fewer of posterior samples for analysis in a systematic way.
- Chain length-the number of posterior samples requested for MCMC.
- Burn-in-The walk the MCMC chain takes prior to arriving at the true posterior.

Occasionally, we will use these, separately or in combination, to "fix" a markov chain obtained through Gibbs sampling.

MCMC Diagnostics

Trace Plots: `plot()`
Patterns are bad.



- Increase burn-in period and lengthen chain.

MCMC Diagnostics Formal

Raftery-Lewis: `raftery.diag()` Posterior tails explored?

```
> raftery.diag(twoprop.out)
[[1]]

Quantile (q) = 0.025
Accuracy (r) = +/- 0.005
Probability (s) = 0.95

           Burn-in  Total Lower bound  Dependence
           (M)      (N)   (Nmin)       factor (I)
 deviance  2        4168  3746         1.110
 pi[1]     2        3646  3746         0.973
 pi[2]     2        3680  3746         0.982
```

- Dependence factor > 5 means strong autocorrelations present.
- May need to choose a different initial starting point to explore the tails better.

MCMC Diagnostics Formal

Geweke: `geweke.diag()` Start of chain match end of chain?

```
> geweke.diag(twoprop.out)
[[1]]

Fraction in 1st window = 0.1
Fraction in 2nd window = 0.5

deviance    pi[1]    pi[2]
 -0.7078  -0.1223   1.1081
```

- Compares first 10% of chain to last 50% of chain.
- Do they have the same mean?
- Quantities reported are z-scores so $> |2|$ would be problematic.

MCMC Diagnostics Formal

Gelman: `gelman.diag()` Do the chains get to the same place?

```
> gelman.diag(twoprop.out)
Potential scale reduction factors:

         Point est. Upper C.I.
deviance          1          1
pi[1]             1          1
pi[2]             1          1

Multivariate psrf

1
```

- Compares the within-chain to between-chain variances similar to ANOVA.
- Potential scale recduction factor of 1 means chains are indistinguishable.
- `gelman.plot()` can be used to see where the chains converge.

MCMC Diagnostics Formal

Heidel-Berg Welch: `heidel.diag()` Chain stationary?

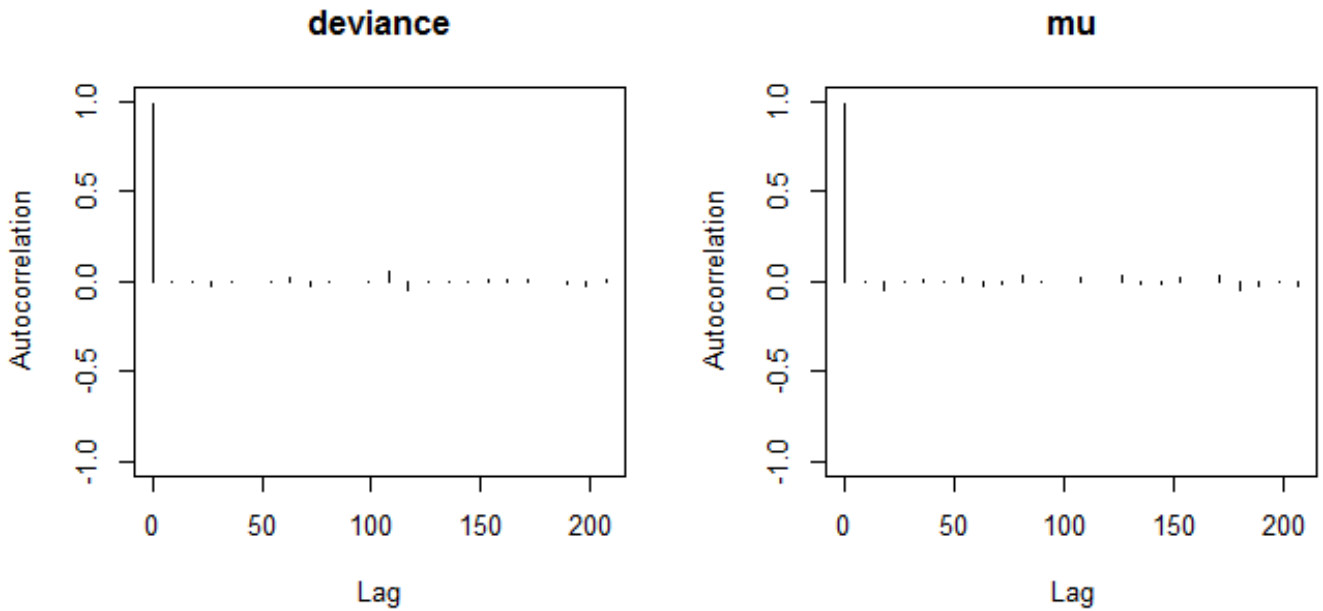
```
> heidel.diag(twoprop.out)
[[1]]

         Stationarity start     p-value
         test         iteration
deviance passed       1         0.804
pi[1]    passed       1         0.762
pi[2]    passed       1         0.247

         Halfwidth Mean   Halfwidth
         test
deviance passed    52.495 0.03631
pi[1]    passed     0.309 0.00175
pi[2]    passed     0.193 0.00159
```

- Performs frequentists tests for the markovian stationary property.
- If Halfwith mean test fails, the run should be extended by a factor $I > 1.5$

MCMC Diagnostics

Autocorrelation: `autocorr.plot()`

- Consecutive Gibbs samples will be correlated.
- Too much auto correlation is bad.
- MC standard error reflects accuracy of Monte Carlo process to estimate true posterior *mean* with dependent samples.



MCMC Diagnostics Formal

Effective Sample Size: `effectiveSize()` Autocorrelation?

```
> effectiveSize(twoprop.out)
deviance    pi[1]    pi[2]
19954.94 20000.00 20873.17
```

- Represents the # of indendent draws.
- If autocorrlation is high from M draws, $ESS << M$.
- May need to run longer chain and increase thinning.

MCMC Diagnostics for SBP Example

See file named "SBP Example Diagnostics.R"

Dealing with Missing Data the Bayesian Way

Missing data are common in practice and there are many alternatives for handling it.
A Bayesian perspective would view missing data in the same way it views unknown parameters.

- Just need to specify the joint model for the missing and observed data and model parameters.
- MCMC can be used to generate a predicted value for the missing data in the usual way.
- The reason for the missingness (mechanism) will dictate the appropriateness of the joint model.

Dealing with Missing Data the Bayesian Way

Three missing data mechanisms and how to handle them in JAGS are outlined below

- Missing Completely At Random (MCAR)-Probability of missingness does not depend on the observed or unobserved quantities.
  - Do nothing, just be sure the data value is NA.
  - WinBUGS will generate a predicted value from the posterior.
  - Missing data mechanism is assumed to be *ignorable*.
- Missing At Random (MAR)-Probability for the missingness depends only on the observed data.
  - Do nothing, just be sure the data value is NA.
  - WinBUGS will generate a predicted value from the posterior.
  - Missing data mechanism is assumed to be *ignorable*.
- Missing Not At Random (MNAR)-Neither MCAR or MAR hold.
  - Model the missing data from the observed and prior knowledge.
  - Need to specify additional likelihood and prior terms for missing data.
  - Missing data mechanism is assumed to be *informative*.

Missing Data for SBP Example

See file named "Systolic Blood Pressure Missing Data Example.odc"