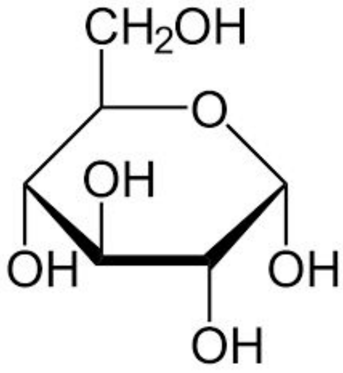
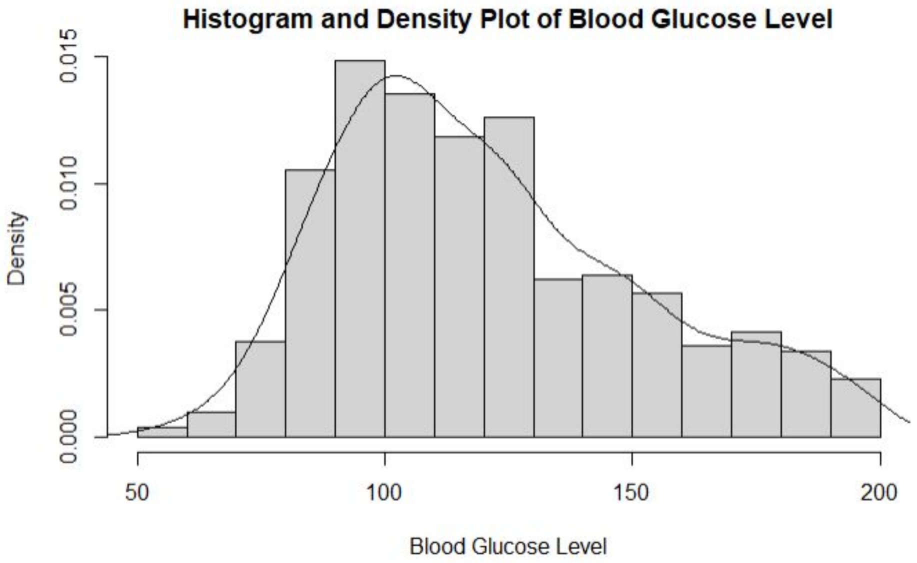
Bayesian Modeling of Blood Glucose

Alan Bouwman
MA 5770, Michigan Technological University

Introduction

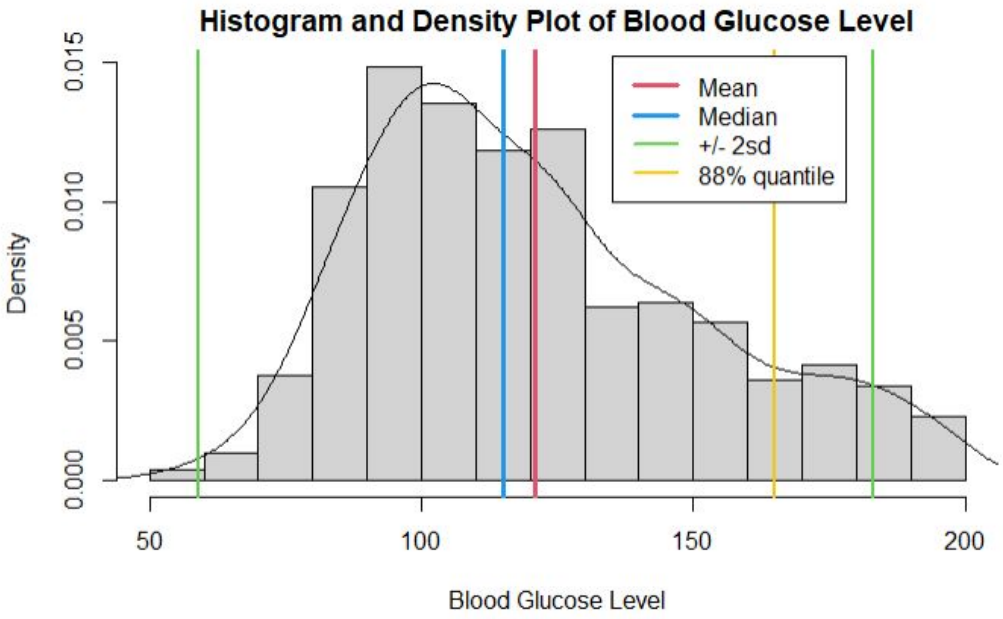


- A population of 532 women living near Phoenix, Arizona were tested for diabetes.
- Many factors were recorded, including blood glucose (sugar) level.
- 200 mg/dL or higher after two hours (after eating) suggests diabetes [MC].
- 126 mg/dL or higher on two separate tests (fasted) is diagnosed as diabetes [MC].
- Average blood glucose is 99 mg/dL or lower (fasted) [CDC].



Exploratory Data Analysis

- Data mean ≅ 121 mg/dL.
- Data Median = 115 mg/dL.
- Data sd ≅ 31 mg/dL.
- 200 of the 523 (38%) observations have glucose of 126 mg/dL or higher.
- About 12% of adult women in the U.S. have diabetes [CDC].
- Using a cutoff of 165 mg/dL, we get 63 of the 532 women (about 12%).



The Model: Sampling Distribution

We choose to use a mixture model for the data. For each of the $n = 532$ study participants, we assign a group membership variable X_i such that

$$X_i = \begin{cases} 1 & \text{with probability } \pi \\ 2 & \text{with probability } 1 - \pi \end{cases}$$

Then the observed data Y_i is given the follwing density:

$$p(y_i|x_i) = \begin{cases} dnorm(y_i; \theta_1, \sigma_1^2) & x_i = 1 \\ dnorm(y_i; \theta_2, \sigma_2^2) & x_i = 2 \end{cases}$$

Note that the X_i are independent and the Y_i are independent given the X_i.

The Model: Prior Distribution

We use the following prior distribution for the model:

$$
\begin{aligned}
p(\pi, \theta_1, \theta_2, \sigma_1^2, \sigma_2^2) &= p(\pi)p(\theta_1)p(\theta_2)p(\sigma_1^2)p(\sigma_2^2) \\
pi &\sim beta(\alpha, beta) \\
\theta_j &\sim normal(\mu_0, \tau_0^2) \text{ for both } j = 1, 2 \\
\sigma_j^2 &\sim inverse - gamma(\nu_0/2, \sigma_0^2\nu_0/2) \text{ for both } j = 1, 2
\end{aligned}
$$

Full Conditional Distribution

The full conditional distribution can be written as:

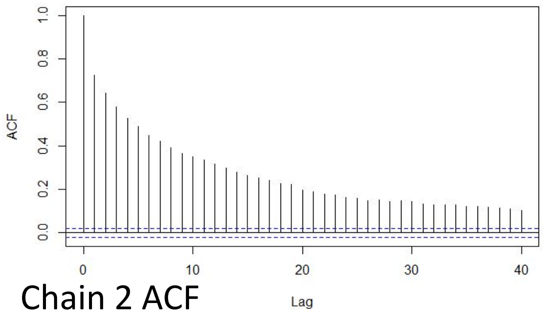
$$
\begin{aligned}
&p(X_i = x_i|\pi, \theta_1, \theta_2, \sigma_1^2, \sigma_2^2, \boldsymbol{Y}, \boldsymbol{X_{-i}}) \propto dbinom(x_i; n, 1, p_2/(p_1+p_2)) + 1 \\
&p(\pi|\theta_1, \theta_2, \sigma_1^2, \sigma_2^2, \boldsymbol{Y}, \boldsymbol{X}) \propto dbeta(\pi; \alpha + n_1, \beta + n_2) \\
&p(\theta_1|\pi, \theta_2, \sigma_1^2, \sigma_2^2, \boldsymbol{Y}, \boldsymbol{X}) \propto dnorm(\theta_1; \mu_{n,1}, \sigma_{n,1}^2) \\
&p(\sigma_1^2|\pi, \theta_1, \theta_2, \sigma_2^2, \boldsymbol{Y}, \boldsymbol{X}) \propto dinverse-gamma(\sigma_1^2; \nu_{n,1}/2, \tau_{n,1}^2 \nu_{n,1}/2) \\
&p(\sigma_2^2|\pi, \theta_1, \theta_2, \sigma_2^2, \boldsymbol{Y}, \boldsymbol{X}) \propto dinverse-gamma(\sigma_2^2; \nu_{n,2}/2, \tau_{n,2}^2 \nu_{n,2}/2)
\end{aligned}
$$

Where

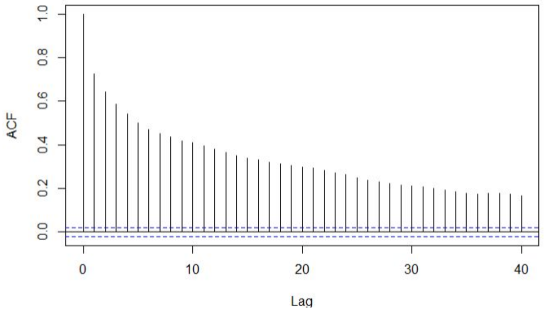
$$
\begin{aligned}
p_1 &= dnorm(y_i; \theta_1, \sigma_1^2) \\
p_2 &= dnorm(y_i; \theta_2, \sigma_2^2) \\
\mu_{n,1} &= \frac{\frac{1}{\tau_0}\mu_0 + \frac{n_1}{\sigma_1^2}\bar{y}_1}{\frac{1}{\tau_0} + \frac{n_1}{\sigma_1^2}} \\
\sigma_{n,1}^2 &= \frac{1}{\frac{1}{\tau_0} + \frac{n_1}{\sigma_1^2}} \\
\mu_{n,2} &= \frac{\frac{1}{\tau_0}\mu_0 + \frac{n_2}{\sigma_2^2}\bar{y}_2}{\frac{1}{\tau_0} + \frac{n_2}{\sigma_2^2}} \\
\sigma_{n,2}^2 &= \frac{1}{\frac{1}{\tau_0} + \frac{n_2}{\sigma_2^2}} \\
\nu_{n,1} &= \nu_0 + n_1 \\
\tau_{n,1}^2 &= \frac{\nu_0\sigma_0^2 + \sum_{i,x_i=1}(y_i - \theta)^2}{\nu_{n,1}} \\
\nu_{n,2} &= \nu_0 + n_2 \\
\tau_{n,2}^2 &= \frac{\nu_0\sigma_0^2 + \sum_{i,x_i=2}(y_i - \theta)^2}{\nu_{n,2}}
\end{aligned}
$$

Model Diagnostics

Chain 1 ACF



Chain 2 ACF

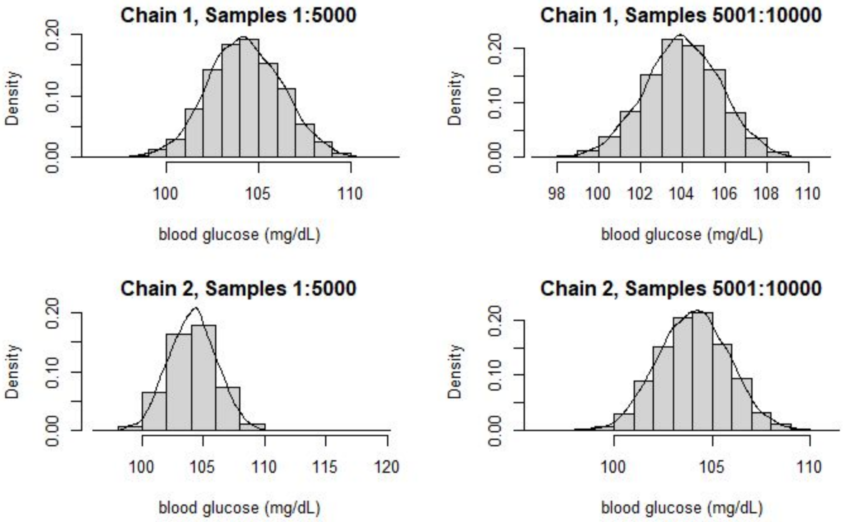


Gibbs Sampler Effective Sizes

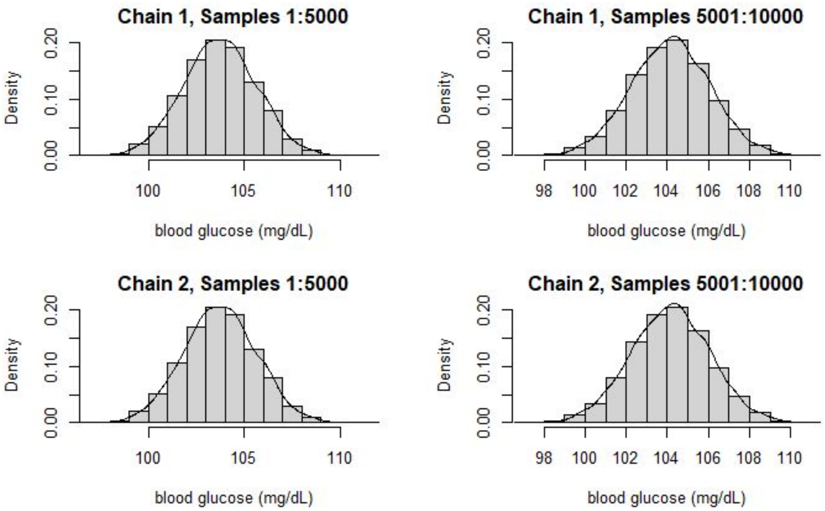
parameter	chain1 effective size	chain 2 effective size
theta1	464	225
theta2	227	425
pi	233	216

Model Diagnostics - JAGS Comparison

Gibbs Sampler low blood glucose group mean sample splits



JAGS low blood glucose group mean sample splits



Posterior Statistics

Posterior Statistics for Chain 1

variable <chr>	mean <dbl>	2.5% quantile <dbl>	97.5% quantile <dbl>
group 1 blood glucose (mg/dL)	104.13	100.36	107.94
group 2 blood glucose (mg/dL)	149.25	137.93	160.93
proportion group 1	0.62	0.48	0.75
min of groups	104.13	100.36	107.94

Posterior Statistics for Chain 2

variable <chr>	mean <dbl>	2.5% quantile <dbl>	97.5% quantile <dbl>
group 1 blood glucose (mg/dL)	148.80	137.47	160.67
group 2 blood glucose (mg/dL)	104.10	100.46	107.85
proportion group 1	0.38	0.25	0.53
min of groups	104.10	100.46	107.85

Conclusion

- The posterior means for group 1 and 2 blood glucose indicate the model is able to separate groups into diabetics and nondiabetics (“low glucose” and “high glucose”).
- The average blood glucose for a non-diabetic is very close to the posterior mean for the “low glucose” group.
- The ADA recommends diabetics to keep their blood sugar below 130 when fasted, or below 180 two hours after eating. 150 is a happy medium.
- The posterior percent of participants in the “high glucose” group (40%) is much higher than percent of women in the U.S. with diabetes (12%).

References

[CDC] CDC Diabetes tests: https://www.cdc.gov/diabetes/basics/getting-tested.html#:~:text=A%20fasting%20blood%20sugar%20level,higher%20indicates%20you%20have%20diabetes.

[MC] Mayo Clinic Diabetes: https://www.mayoclinic.org/diseases-conditions/diabetes/diagnosis-treatment/drc-20371451

[ADA] Americans with Diabetes Associates (Diabetes.org): https://diabetes.org/diabetes/a1c/diagnosis