STA 602 HW 8

Ryan Tang

October 28th 2022

1 Exercise 6.2

(1) Glucose KDE The empirical distribution follows a "somewhat" normal shape. But it is skewed to the right.

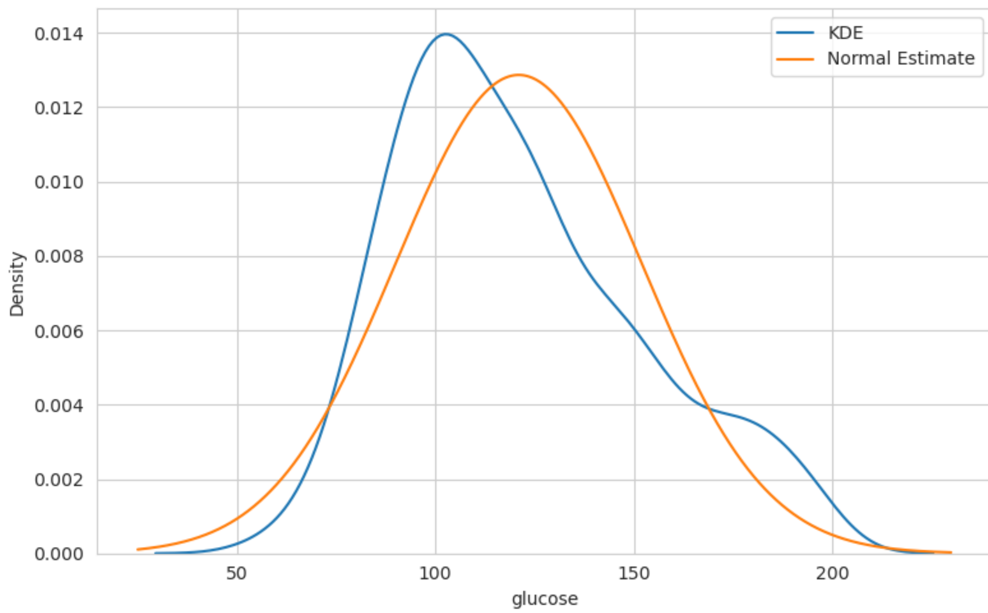

Figure 1: KDE vs Normal Assumption Comparison

(2) Full conditionals We are given the following priors and hierarchical model.

$$\begin{aligned} Y_i|X_i = k &\sim \mathcal{N}(\theta_k, \sigma_k^2) && k \in \{1,2\} \\ \theta_k &\sim \mathcal{N}(\mu_o, \tau_o^2) \\ 1/\sigma_k^2 = \gamma_k &\sim Gamma(\frac{\nu_o}{2}, \frac{\nu_o \sigma_o^2}{2}) \\ X_i &\sim Bernoulli(p) \\ p &\sim Beta(a,b) \end{aligned}$$

Then we can write the following full conditionals regarding the joint in proportionality. And the full conditionals are decomposable in our setup.

$$\begin{aligned} N_k &= \sum_i^N \mathbb{I}(X_i = k) \\ p(Y,\theta,\sigma^2,X,p) &= p(Y|\theta,\sigma^2,X,p)p(\theta|X,p)p(\sigma^2|X,p)p(X|p)p(p) \\ &= p(p)\prod_i^{N_1} p(Y_i|\theta_1,\sigma_1^2)p(\theta_1)p(\sigma_1^2)p(X_i=1|p)\prod_i^{N_2} p(Y_i|\theta_2,\sigma_2^2)p(\theta_2)p(\sigma_2^2)p(X_i=2|p) \end{aligned}$$

In their full conditionals, X_i and p are still Bernoulli and Beta distributions.

$$
\begin{aligned}
p(p|Y,\theta,\sigma^2,X) &\propto p(Y,\theta,\sigma^2,X,p) \\
&\propto p(X|p)p(p) \\
&\propto p^{a-1}(1-p)^{b-1}p^{N_1}(1-p)^{N_2} \\
&\propto p^{a+N_1-1}(1-p)^{b+N_2-1} \\
&\sim Beta(a+N_1, b+N_2)
\end{aligned}
$$

$$
\begin{aligned}
p(X|Y,\theta,\sigma^2,p) &\propto \prod_i^N p(Y_i|\theta,\sigma^2,X_i,p)p(X_i|p) \\
&\propto \prod_i^N \left[p(Y_i|\theta_1,\sigma_1^2)p(X_i=1|p)\right]^{\mathbb{I}(X_i=1)} \left[p(Y_i|\theta_2,\sigma_2^2)p(X_i=2|p)\right]^{\mathbb{I}(X_i=2)} \\
p(X_i=1|Y_i,\theta,\sigma^2,p) &\propto p(Y_i|\theta_1,\sigma_1^2)p \\
&\sim Bernoulli\left(p_1 = \frac{\mathcal{N}(Y_i|\theta_1,\sigma_1^2)p}{\mathcal{N}(Y_i|\theta_1,\sigma_1^2)p + \mathcal{N}(Y_i|\theta_2,\sigma_2^2)(1-p)}\right) \\
p(X_i=2|Y_i,\theta,\sigma^2,p) &\propto p(Y_i|\theta_2,\sigma_2^2)(1-p) \\
&\sim Bernoulli\left(p_2 = \frac{\mathcal{N}(Y_i|\theta_2,\sigma_2^2)(1-p)}{\mathcal{N}(Y_i|\theta_1,\sigma_1^2)p + \mathcal{N}(Y_i|\theta_2,\sigma_2^2)(1-p)}\right) \\
&= 1 - p(X_i=1|Y_i,\theta,\sigma^2,p)
\end{aligned}
$$

θ_k and σ_k are the normal semi-conjugate posterior forms in their full conditionals. However, the sufficient statistics are in terms of N_1 and N_2 that depend on the hidden X.

$$
N_k\bar{Y}_k = \sum_i^N \mathbb{I}(X_i=k)Y_i
$$

$$
\begin{aligned}
p(\theta_1|Y,\sigma^2,X,p) &\propto p(\theta_1)\prod_i^{N_1} p(Y_i|\theta_1,\sigma_1^2) \\
&\sim \mathcal{N}(\mu_{N_1} = \frac{\tilde{\tau}_o}{\tilde{\tau}_{N_1}}\mu_o + \frac{N_1\bar{Y}_1}{\tau_{\tilde{N}_1}}\gamma_1, \tilde{\tau}_{N_1} = \tilde{\tau}_o + N_1\gamma_1) \\
p(\theta_2|Y,\sigma^2,X,p) &\sim \mathcal{N}(\mu_{N_2} = \frac{\tilde{\tau}_o}{\tilde{\tau}_{N_2}}\mu_o + \frac{N_2\bar{Y}_2}{\tau_{\tilde{N}_2}}\gamma_2, \tilde{\tau}_{N_2} = \tilde{\tau}_o + N_2\gamma_2) \\
p(1/\sigma_1^2|Y,\theta,X,p) &= p(\gamma_1|Y,\theta_1,X,p) \\
&\sim Gamma(\frac{\nu_{N_1}}{2}, \frac{\nu_{N_1}\sigma_{N_1}^2}{2}) \\
\nu_{N_1} &= \nu_o + N_1,\ \sigma_{N_1}^2 = \frac{1}{\nu_{N_1}}(\nu_o\sigma_o^2 + \sum_{i\in N_1}(Y_i-\theta_1)^2) \\
p(1/\sigma_2^2|Y,\theta,X,p) &= p(\gamma_2|Y,\theta_2,X,p) \\
&\sim Gamma(\frac{\nu_{N_2}}{2}, \frac{\nu_{N_2}\sigma_{N_2}^2}{2}) \\
\nu_{N_2} &= \nu_o + N_2,\ \sigma_{N_2}^2 = \frac{1}{\nu_{N_2}}(\nu_o\sigma_o^2 + \sum_{i\in N_2}(Y_i-\theta_2)^2)
\end{aligned}
$$

(3) Gibbs Samples Here we plot the autocorrelation for the two θ samples, and we can see very strong AR between samples. The effective sample size (ESS) for $\theta_{(1)}^{(s)}$ is 971 out of 10,000 draws. And the ESS for $\theta_{(2)}^{(s)}$ is $1,202$. The chain is probably not that well mixed after all.

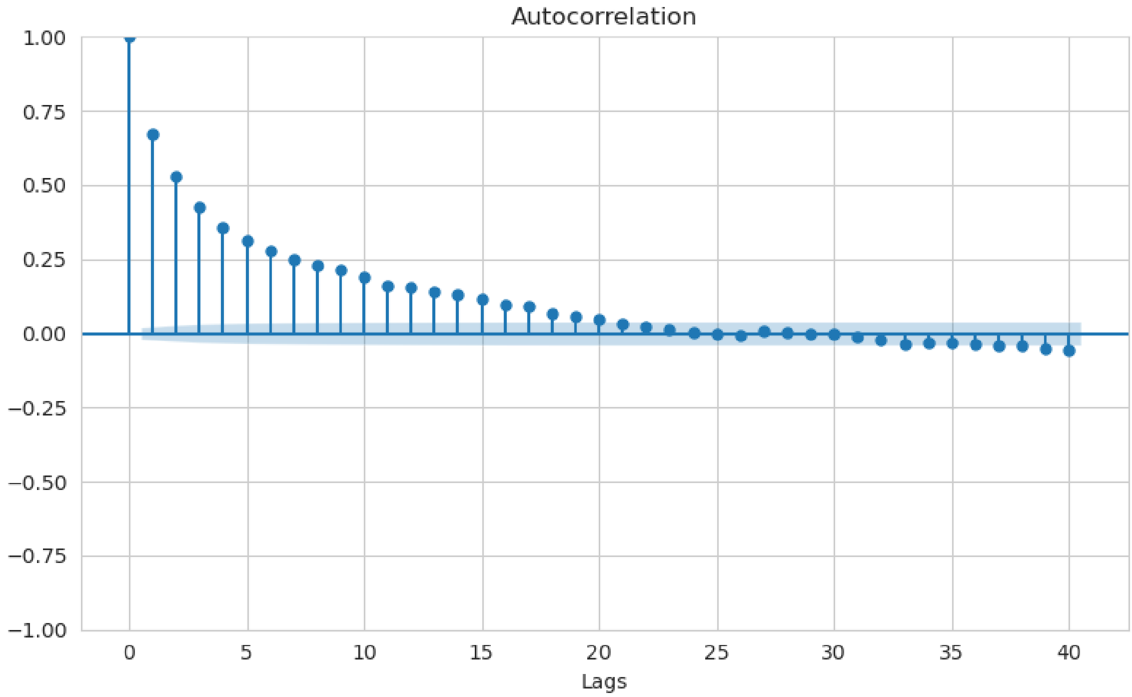

Figure 2: $\theta_{(1)}^{(s)}$ The min of two θ Autocorrelation Plot

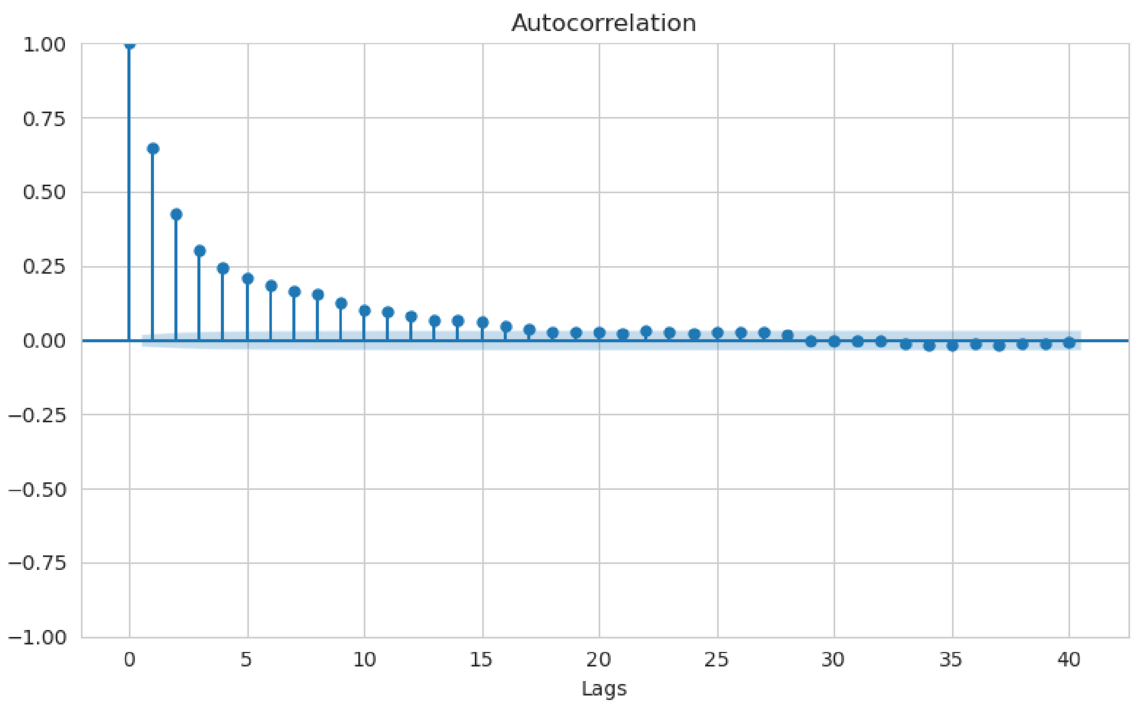

Figure 3: $\theta_{(2)}^{(s)}$ The max of two θ Autocorrelation Plot

(d) Posterior Predictive From the surface, the posterior distribution doesn't quite match the observations. However, I think it is not the mixture model's fault, but Gibbs sampler didn't do a good job exploring the entire sample space. If we use EM instead, the solution might well be better.

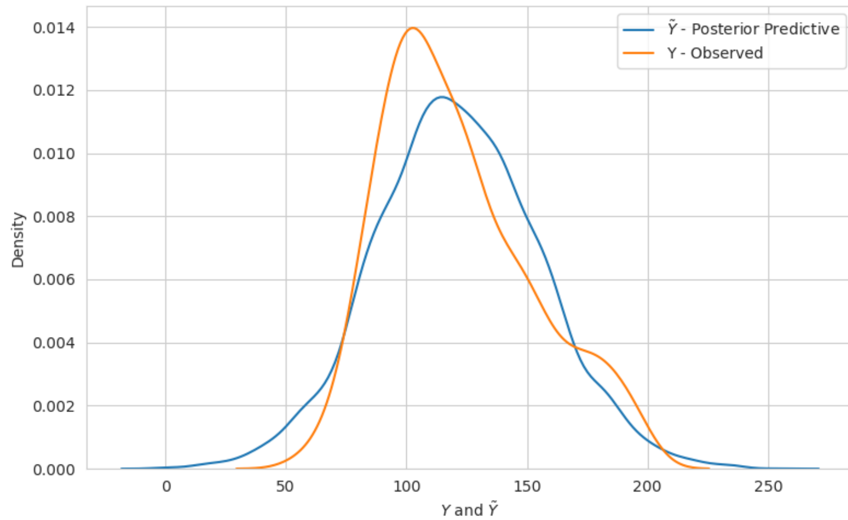

Figure 4: $\tilde{Y}$ Posterior Predictive Plot

2 Exercise 6.3

(a) Full Conditional when c is known We are given the following hierarchical model in the probit regression with a known c.

$$
\begin{aligned}
Z_i &= \beta X_i + \epsilon_i \\
Y_i &= \delta_{c,\infty}(Z_i) = \mathbb{I}(Z_i > c)1 + \mathbb{I}(Z_i \leq c)0 \\
\epsilon_i &\overset{iid}{\sim} \mathcal{N}(0,1) \\
\beta &\sim \mathcal{N}(0, \tau^2_{\beta,o})
\end{aligned}
$$

$$
\begin{aligned}
p(Y, X, Z, c, \beta) &= p(Y|Z, c)p(Z|X, \beta)p(\beta) \\
&= p(\beta)\prod_i^N p(Y_i|Z_i, c)p(Z_i|X_i, \beta)
\end{aligned}
$$

$$
\begin{aligned}
p(\beta|Y, X, Z, c) &\propto p(\beta)\prod_i^N p(Z_i|X_i, \beta) \\
&\propto \exp\left[-\frac{1}{2}[\beta^2\tilde{\tau}_{\beta,o} + \sum_i^N (Z_i - X_i\beta)^2]\right] && \frac{1}{\tau^2_{\beta,o}} = \tau_{\tilde{\beta},o} \\
&\propto \exp\left[-\frac{1}{2}[\beta^2(\tilde{\tau}_{\beta,o} + \sum_i^N X_i^2) - 2\beta\sum_i Z_iX_i]\right] \\
&\propto \exp\left[-\frac{1}{2}\tilde{\tau}_{\beta,N}[\beta^2 - 2\beta\frac{\sum_i Z_iX_i}{\tilde{\tau}_{\beta,N}}]\right] && \tilde{\tau}_{\beta,N} = \tilde{\tau}_{\beta,o} + \sum_i^N X_i^2 \\
&\sim \mathcal{N}\left(\frac{\sum_i Z_iX_i}{\tilde{\tau}_{\beta,N}}, \tau^2_{\beta,N}\right)
\end{aligned}
$$

(b) Full Conditional for c and Z_i We can see c is a constrained normal distribution under a interval.

$$
\begin{aligned}
p(c) &= \mathcal{N}(0, \tau_c^2) \\
p(c|Y, X, Z, \beta) &\propto p(Y|Z, c)p(c) \\
&\propto p(c) \prod_i p(Y_i|Z_i, c) \\
&\propto p(c) \prod_{i \in Y_i = 1} \delta_{(c,\infty)}(Z_i)^{Y_i} \prod_{j \in Y_i = 0} \delta_{(-\infty,c)}(Z_j)^{1-Y_j} \\
&\propto \mathcal{N}(0, \tau_c^2)\mathbb{I}(\max\{Z_i \in Y_i = 0\} \leq c < \min\{Z_i \in Y_i = 1\})
\end{aligned}
$$

And on the other hand, the full conditional of Z_i is a truncated normal. First, because we know the latent variable Z_i itself is given by a normal generative model with X_i and β. But the additional information from Y_i constrained the region that Z_i could be. Hence, the following formulations.

$$
p(Z_i|X_i, Y_i, \beta, c) \propto p(Y_i|Z_i, c)p(Z_i|X_i, \beta)
$$

$$
\begin{aligned}
p(Z_i|X_i, \beta, Y_i = 1, c) &\propto p(Y_i = 1|Z_i, c)p(Z_i|X_i, \beta) \\
&\propto \mathcal{N}(Z_i|X_i\beta, 1)\delta_{(c,\infty)}(Z_i) \\
p(Z_i|X_i, \beta, Y_i = 2, c) &\propto \mathcal{N}(Z_i|X_i\beta, 1)\delta_{(-\infty,c)}(Z_i)
\end{aligned}
$$

(c) Gibbs Samples Here we implemented the above formulation using Gibbs sampler. I had it ran for 2,000 burn-in epochs and additional 30,000 sampling epochs. The following chain is the result. We still see high autocorrelation for both c and β during the sampling. The effective sample size for c is 1,156 and it is 1,651 for β. The chains were very sticky.

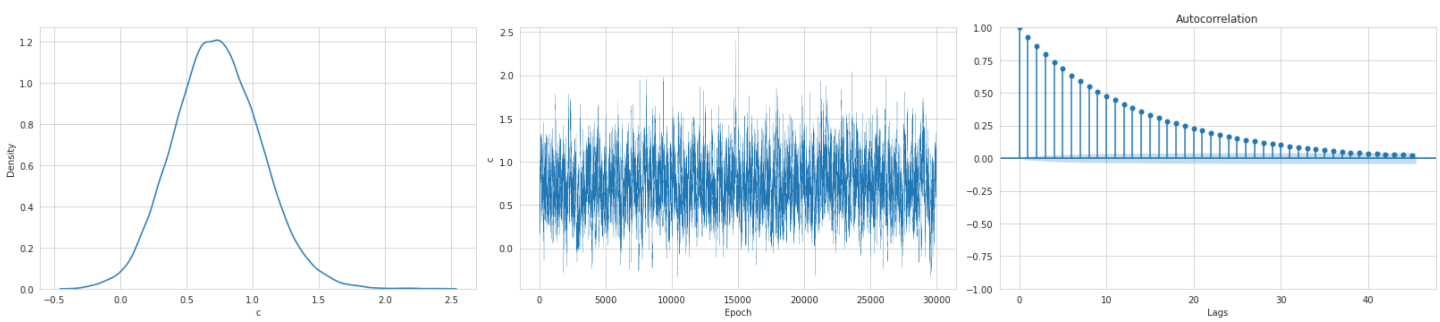

Figure 5: c Posterior Gibbs Samples

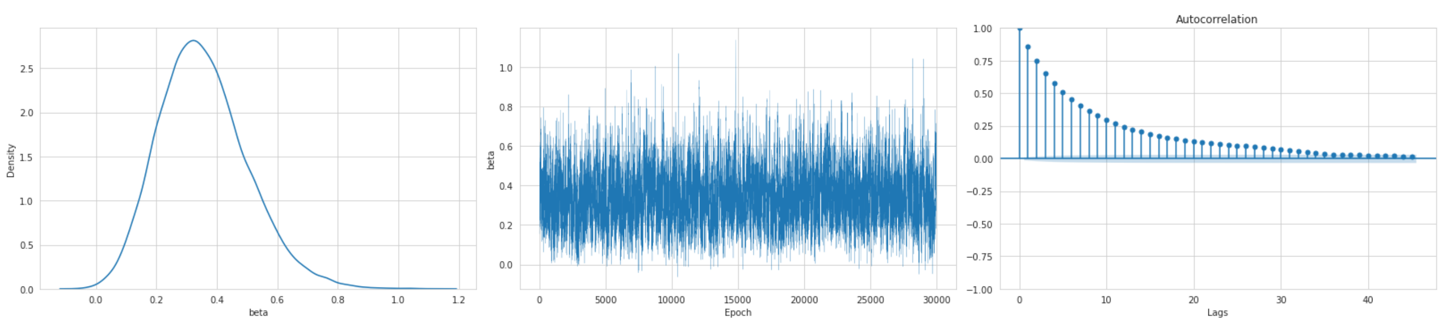

Figure 6: β Posterior Gibbs Samples

(d) β Inference Lastly, we like to know if the age differences between men and woman would lead to higher divorce rate. Base on the observation data, we have our $\beta|Y, X$ posterior's 95% confidence interval between $[0.104, 0.673]$. And, the $Pr[\beta > 0|Y, X] = 0.9991$