statistics **Exam: ?, Time: ?**
Lund University **Bayesian Methods**
. **code**

Examiner: Jonas Wallin, tel secret

Allowed aids: None.

Correct, well motivated solution gives points indicated within the parentheses at each exercise.

---

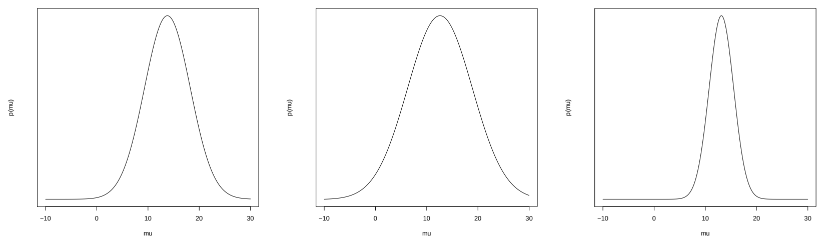
1. Assume in a population that three out of five people are male. Further suppose the probability of being above 190 cm is $\frac{4}{100}$ for males and $\frac{1}{100}$ for females. If a person is above 190 cm what is the probability that it is a female? (5p)

2. Suppose you can generate independent samples of the random variable X with posterior distribution, p. Describe how to calculate an estimate of the expected value $\mathbb{E}[X]$ and also describe how to check how uncertain you are of the estimate. (5p)

3. For the following model

$$
\begin{aligned}
y_i &\sim N(\mu, 20),\\
\mu &\sim N(0, 100),
\end{aligned}
$$

the three figures below show the posterior distribution for $p(\mu|y_1, \ldots y_{10})$, $p(\mu|y_1, \ldots y_{20})$ and $p(\mu|y_1, \ldots y_{70})$. Which figures represent which posterior distribution? (5p)

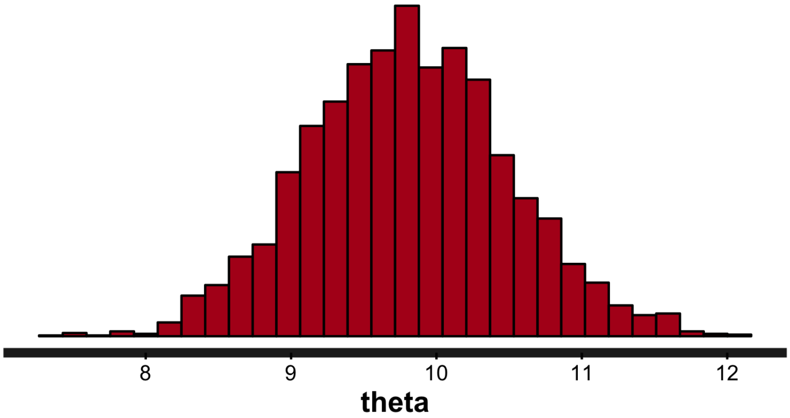


4. The WAIC is given by two quantities:

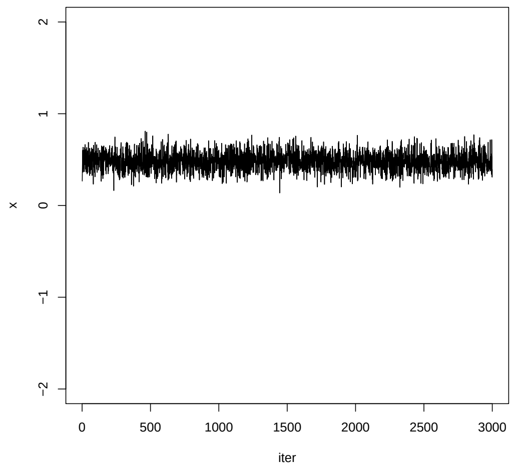
$$
WAIC = -\sum_{j=1}^{n} \log(\mathbb{E}_Y[p(Y = y_j|\sigma, y_1, \ldots, y_n)]) + 2\sum_{j=1}^{n} \mathbb{V}_Y[\log(p(Y = y_j|y_1, \ldots, y_n))]
$$

   a) Explain what WAIC is used for. (1p)
   b) Explain what the two terms measures. (2p)
   c) WAIC can be used for weighting, how is this done. (2p)

5. (a) In figure 1, we have output from posterior draws of θ. What approximately is the 90% PI of θ, and what is the map of the posterior distribution? (2.5p)



(b) In the figure below, we have a traceplot of x, can you use the samples from the MCMC chain to estimate the posterior mean of the posterior distribution? (2.5p)



6. Suppose that a drug company is testing the effect of a drug has for preventing diabetes at patients at risk. During a study one has monitored n_1 patients given the drug and also n_2 patient given placebo. Further one knows that the body weight has positive effect on the risk of diabetes. Write down a Bayesian model to test if the drug has a significant effect, after the observations, that includes that adjust for body weight. How would you evaluate if the drug had an effect? (5p)

7. Suppose you want to fit the following model:

$$\begin{aligned} y_i &\sim Po(\lambda_i) \\ log(\lambda_i) &= \alpha + \beta x_i, \\ \beta &\sim N^+(0, 100), \\ \alpha &\sim N(0, 100), \end{aligned}$$

where $N^+(0, 100)$ is a truncated (on the positive axis) Normal distribution. Write the Rstan code that would estimate such a model. (5p)

8. Assume that the data $y_1, \ldots, y_n$ comes from the following model:

$$
\begin{aligned}
y_i &\sim z_i N(y_i; \mu_1, \sigma_1) + (1 - z_i) N(y_i; \mu_2, \sigma_2) \\
z_i &\sim Bin(1, \theta), \\
\mu_1 &\sim N(0, 10), \\
\mu_2 &\sim N(0, 10), \\
\sigma_1 &\sim HC(0, 1), \\
\sigma_2 &\sim HC(0, 1), \\
\theta &\sim U[0, 1].
\end{aligned}
$$

Derive the posterior distribution of $p(z_1|\mu_1, \mu_2, \sigma_1, \sigma_2, \theta, y_1)$. Hint: z_1 can take two values: $0, 1$. (5p)