10-602: Statistical Approaches to Learning and Discovery
Final Examination

Your solutions for this examination are due by 5 p.m. Thursday May 3. You are allowed to consult class notes, any material posted on the course website or distributed in class, the textbook (Tanner), and a basic mathematical statistics textbook such as DeGroot which is used in 36-325/326. Your work is to be your own (no collaborations).

1. We have two variables, y_1 and y_2, and want to know which has larger variance. Assume the variables are normally distributed:

$$p(y_1|m_1, v_1) = \mathcal{N}(m_1, v_1) \tag{1}$$
$$p(y_2|m_2, v_2) = \mathcal{N}(m_2, v_2) \tag{2}$$

Give the parameters a conjugate prior distribution:

$$p(m_j|v_j) \sim \mathcal{N}(\mu_j, \alpha_j^{-1} v_j) \tag{3}$$
$$p(v_j) \sim IG(\nu_j, \lambda_j) \qquad j = 1, 2 \tag{4}$$

The data you will analyze come from a experiment in which samples of yarn were taken at random from spinning machines in a certain production area. The first sample represents the observed breaking strength (in grams) before the machines were modified and the second sample comes after the modification of the machines. The summary data for this problem are as follows: $D = (n_1 = 20, n_2 = 12, \bar{y}_1 = 50, \bar{y}_2 = 55, S_1 = 208, S_2 = 440)$, where the $\{\bar{y}_j\}$ are sample means and $S_j = \sum_i (y_{ji} - \bar{y}_j)^2$ are sample scatters.

   (a) Find the posterior distribution for the variances alone $p(v_1, v_2|D)$. Write it in terms of n_j, $\bar{y}_j$, S_j, and the hyperparameters μ, α, ν, and λ. You can reach the answer fastest by spotting the form of the distribution.

   (b) Let $\gamma_i = \log(v_i)$. Use the delta method to get a normal approximation for the probability that $v_1 < v_2$. Write it in terms of n_j, $\bar{y}_j$, S_j and the hyperparameters.

   (c) Use a second-order Laplace approximation to get another normal approximation for $Pr(v_1 < v_2|D)$.

   (d) If $\mu_j = 0$, $\alpha_j = 1$, and $\nu_j = \lambda_j = 1$, what are your estimates of $Pr(v_1 < v_2|D)$ for the above data?

2. Consider the following model for a time series of observations $D = (y_1, ..., y_n)$ (this is a sequence, not a set):

$$y_0 = 0 \tag{5}$$
$$y_t = a y_{t-1} + \epsilon \tag{6}$$
$$\epsilon \sim \mathcal{N}(0, v) \tag{7}$$

This model describes a Markov chain—specifically an auto-regressive model similar to a random walk. The likelihood function is

$$
\begin{aligned}
p(D|a, v) &= \prod_{t=1}^{n} \mathcal{N}(y_t; a y_{t-1}, v) &(8)\\
&= \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi v}} \exp\left(-\frac{(y_t - a y_{t-1})^2}{2v}\right) &(9)
\end{aligned}
$$

(a) What are the maximum-likelihood estimates for a and v given D? This should tell you why it is called auto-regression.

(b) Let a and v have the following prior distributions:

$$
\begin{aligned}
p(a) &\sim \mathcal{N}(m_a, v_a) &(10)\\
p(v) &\sim IG(c, d) = \frac{d^c}{\Gamma(c) v^{c+1}} \exp(-d/v) &(11)
\end{aligned}
$$

where m_a, v_a, c, and d are known constants. What is $p(a|v, D)$? What is $p(v|a, D)$?

3. In this problem you will do changepoint analysis of time series data using the auto-regressive model. On the course web page you will find a time series where the Markov parameters changed abruptly from (a_1, v_1) to (a_2, v_2). Your goal is to make inferences about where the change occurred. If the changepoint was k, the assumption is that $y_1, ..., y_k$ were generated according to (a_1, v_1) and that $y_{k+1}, ..., y_n$ were generated according to (a_2, v_2). Note that y_{k+1} is statistically dependent on y_k, the last value before the change.

Use the priors in problem 2, where m_a, v_a, c, and d are common to both sets of parameters (a_j, v_j). Let $p(k)$ be uniform and give m_a and v_a prior distributions

$$
\begin{aligned}
p(m_a) &\sim \mathcal{N}(0, 1) &(12)\\
p(v_a) &\sim IG(1, 1) &(13)
\end{aligned}
$$

Set $c = 2$ and $d = 0.1$ (to skip sampling them).

(a) Plot the data as a curve. Is a changepoint evident in the series?

(b) What is the joint MLE for (k, a_1, v_1, a_2, v_2)? Does the estimated changepoint seem reasonable?

(c) Find a Monte Carlo formula for estimating $p(k|D)$ given samples from the full posterior, $p(a_1, v_1, a_2, v_2, m_a, v_a, k|D)$.

(d) Find the conditionals needed for Gibbs sampling from the full posterior.

(e) Run the Gibbs sampler and plot the estimate of $p(k|D)$. Make the estimate as accurate as you can.