10-602: Statistical Learning

Problem Set 2
Date: February 1, 2001 **Due: start of class February 8, 2001**

---

1. This problem reviews transformations of random variables. The random variable x has density

$$p(x = \hat{x}) = \begin{cases} 0.5/\hat{x} & 1/e \le \hat{x} \le e \\ 0 & \text{otherwise} \end{cases}$$

   (a) What is the density of $z = \ln(x)$?

   (b) What is the density of $y = 1/x$? What is the mode of $p(y = \hat{y})$, compared to the mode of $p(x = \hat{x})$? (The mode is the value with maximum density.)

   (c) For discrete random variables, the mode of the probability distribution can be interpreted as the most likely value. Does this interpretation apply equally well to continuous random variables? Explain.

2. This problem reviews linear algebra. Let $\mathbf{x}$ be a random vector with mean $\mathbf{m}$ and covariance matrix Σ.

   (a) What is the covariance matrix of $\mathbf{A}\mathbf{x}$?

   (b) Show that $\mathrm{tr}(\mathbf{AB}) = \mathrm{tr}(\mathbf{BA})$.

   (c) What is $E[\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}]$? (Hint: Use the last part.)

3. This problem compares the frequentist and Bayesian notion of confidence interval. Suppose samples arise from a uniform distribution $U(\theta - 1, \theta + 1)$ with unknown center θ. Given two sample points y_1 and y_2, we want a confidence interval on θ.

   (a) Number the points so that $y_2 > y_1$. Show that (y_1, y_2) is a frequentist 50% confidence interval for θ, that is

$$Pr(y_1 \le \theta \le y_2 | \theta) = 1/2 \tag{1}$$

   Note that y_1 and y_2 are the random variables in this equation.

   (b) With a uniform prior on θ, what is the Bayesian posterior for θ? Using the data, find a minimal interval (a, b) such that

$$Pr(a \le \theta \le b | y_1, y_2) = 1/2 \tag{2}$$

   Here θ is the random variable.

   (c) Compare the two confidence intervals when y_1 and y_2 are close together vs. far apart. Think about which behavior you would want in your applications.

4. We have two variables, x and y, and want to know which has larger expected value, as well as a measure of confidence. Assume the variables are normally distributed:

$$p(x|m_1, v_1) = \mathcal{N}(m_1, v_1) \tag{3}$$
$$p(y|m_2, v_2) = \mathcal{N}(m_2, v_2) \tag{4}$$

To make our judgement, we only have data $D_1 = \{x_1, ..., x_n\}$ and $D_2 = \{y_1, ..., y_n\}$. Thus m_1 and m_2 should have a vague prior distribution, such as

$$p(m_1, m_2) = \lim_{v_0 \to \infty} \mathcal{N}(m_1; 0, v_0)\mathcal{N}(m_2; 0, v_0) \tag{5}$$

(a) Suppose we know v_1 and v_2. What is the posterior distribution of m_1 and m_2, given the data? Are m_1 and m_2 independent in this posterior?

(b) What is the distribution of $b = m_1 - m_2$, given the data? Use this to compute the probability that $m_1 < m_2$. You can use the cumulative distribution of the normal, defined by

$$\phi(z) = \int_{-\infty}^{z} \mathcal{N}(w; 0, 1)dw \tag{6}$$

(c) Suppose v_1 and v_2 are unknown, with posterior distribution $p(v_1, v_2|D_1, D_2)$. What is the probability that $m_1 < m_2$? Leave your answer in the form of a succinct integral.

5. This problem addresses the issue of pooling data from multiple sources. You are trying to measure a quantity θ which can be considered a random variable whose prior $p(\theta)$ is essentially uniform over the real line. You have two different measuring devices at your disposal, which produce measurements a and b respectively. Unfortunately, the devices can only make noisy measurements, and the noise variances are different:

$$p(a|\theta) \sim \mathcal{N}(\theta, v_a) \qquad p(b|\theta) \sim \mathcal{N}(\theta, v_b) \tag{7}$$

You make two independent measurements with each device, which turn out to be

$$a_1 = 1.1 \qquad b_1 = 2.9$$
$$a_2 = 1.9 \qquad b_2 = 4.1$$

(a) Suppose you estimate v_a and v_b using maximum-likelihood and then believe that these are the true variances. What is your posterior distribution on θ?

(b) Suppose you instead treat v_a and v_b as random variables and assign them noninformative priors independent of each other and of θ:

$$p(v_a|\theta) = 1/v_a$$
$$p(v_b|\theta) = 1/v_b$$

What is your posterior distribution on θ? (Hint: It involves integrals that were solved in class. Also see Tanner, p18.) You can ignore the normalizing constant—it cannot be expressed with elementary functions.

(c) Plot the posterior you computed in part (b). Explain as best you can why it looks the way it does. How does this exact posterior compare to the approximate one in part (a)?