Final Exam [30 pts]

Monday, May 13th, 2019 1:30–3:30 pm

PUBH 8442: Bayes Decision Theory and Data Analysis

Give your final answers in simplified, closed form wherever possible. However, partial credit will be awarded for incomplete solutions. Good luck!

1. *Linear Model with Measurement Error* [18 pts]
   In what follows, assume that all variables are independent unless otherwise specified. Consider a linear model with measurement error in the predictors. Data (X_i, Y_i) are observed, where

$$\begin{aligned} Y_i &= \beta \tilde{X}_i + \epsilon_i, \\ \epsilon_i &\sim N(0, \sigma^2), \text{ and} \\ \tilde{X}_i \mid X_i &\sim N(X_i, \tau^2) \quad \text{for } i = 1, \dots, n. \end{aligned}$$

   Assume σ^2 is known. We use an improper flat prior for $\beta : p(\beta) = 1\ \forall \beta \in \mathbb{R}$, and in parts (a-d) we use an inverse-gamma prior for τ^2, $\text{IG}(a, b)$:

$$p(\tau^2) = \frac{b^a}{\Gamma(a)} (\tau^2)^{-(a+1)} e^{-(b/\tau^2)}.$$

   Let $\mathbf{Y} = (Y_1, \dots, Y_n)$, $\mathbf{X} = (X_1, \dots, X_n)$, and $\tilde{\mathbf{X}} = (\tilde{X}_1, \dots, \tilde{X}_n)$.

   (a) (3 points) What is the marginal distribution of $\mathbf{Y}$ given $\mathbf{X}$, β and τ^2: $p(\mathbf{Y} \mid \beta, \tau^2, \mathbf{X})$?

   (b) (3 points) What is the conditional posterior distribution of β, $p(\beta \mid \tilde{\mathbf{X}}, \tau^2, \mathbf{X}, \mathbf{Y})$?

   (c) (3 points) What is the conditional posterior distribution of τ^2, $p(\tau^2 \mid \tilde{\mathbf{X}}, \beta, \mathbf{X}, \mathbf{Y})$?

   (d) (5 points) Describe explicitly a Gibbs sampling algorithm to simulate from the joint posterior distribution $p(\beta, \tau^2, \tilde{\mathbf{X}} \mid \mathbf{X}, \mathbf{Y})$.

   (e) (4 points) Instead of using an $\text{IG}(a, b)$ prior for τ^2, describe a reasonable empirical Bayes approach to estimate τ^2. Write your answer in closed form as a function of the observed data, $\hat{\tau}^2 = f(\mathbf{X}, \mathbf{Y}, \sigma^2)$.

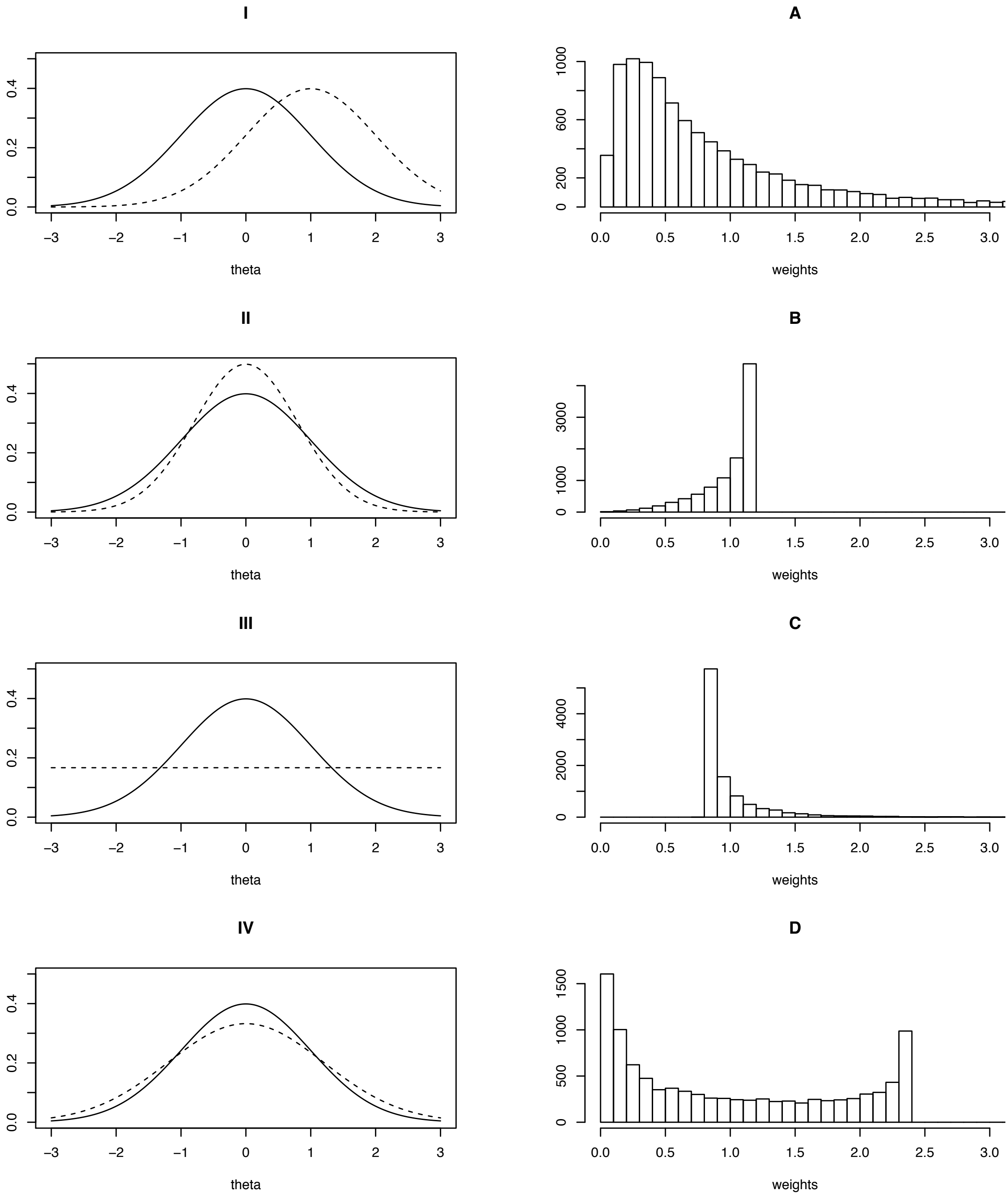

Figure 1

2. *Importance Sampling* [4 pts]
   Figures 1 illustrate importance sampling from a $N(0, 1)$ distribution. The left column gives four different importance densities (dashed line) that were used to draw importance samples from the target distribution (black line). The right column shows histograms of the sample weights generated from each scenario on the left. Match the weight histogram on the right (A,B,C, or D) that corresponds to each panel on the left (I, II, III, or IV).

3. *Genomic testing* [8 pts]
Consider genetic data for N_1 individuals that are affected by a disease, and N_0 unaffected individuals. A single neucleotide polymorphism (SNP) is recorded as present ($Y = 1$) or absent ($Y = 0$) for each individual. Thus, the data are of the form $Y_{ij} \in \{0, 1\}$ for affected/unaffected status $j = 0, 1$, and individuals $i = 1, \ldots, N_j$. Assume

$$Y_{ij} \overset{\text{indep}}{\sim} \text{Bernoulli}(\theta_j),$$

so θ_0 is the probability that the SNP is present for an unaffected individual and θ_1 is the probability that the SNP is present for an affected individual. Consider two models, one where disease status has no effect (M_0) and one where it does (M_a):

- M_0: $\theta_0 = \theta_1 = \theta$ and $\theta \sim \text{Beta}(1, 1)$,
- M_a: $\theta_0 \sim \text{Beta}(1, 1)$ and $\theta_1 \sim \text{Beta}(1, 1)$, where θ_0 and θ_1 are independent.

(a) (2 pts) What are the posterior distributions of θ_0 and θ_1 under M_a, $p(\theta_0 \mid \mathbf{Y}, M_a)$ and $p(\theta_1 \mid \mathbf{Y}, M_a)$?

(b) (2 pts) What is the posterior distribution of θ_0, under M_0, $p(\theta_0 \mid \mathbf{Y}, M_0)$?

(c) (4 pts) Assume S_0 unaffected individuals have the SNP and S_1 affected have the SNP. What is the Bayes factor for M_0 over M_a?

Have a good summer!