Statistics 520

Exam 1 – Example Questions

Fall 2025

Short Answer Questions

(5 pts. each)

1. Given a measurement or observational operation O that has possible outcomes in a sample space $\mathcal{S}$, how does one compute the probability of an event $E \subset \mathcal{S}$ under a Laplacian concept of probability? Use $|A|$ to denote the size of a set A.

2. With iterative optimization algorithms used to locate maximum likelihood estimates, each type of algorithm is designed to achieve a particular type of convergence. Despite this, we would like an algorithm to result in three types of convergence. What are they?

3. We are given a data model $f(\boldsymbol{y}|\theta)$ for $\boldsymbol{y} \in \Omega$ and $\theta \in \Theta$, and a prior $\pi(\theta|\boldsymbol{\lambda})$ for $\theta \in \Theta$. If the data model and prior are conjugate, that means the posterior distribution of θ has the form,

$$p(\theta|\boldsymbol{y}) =$$

4. For a random variable Y, consider the assignment of an exponential family distribution written with the following parameterization, for $y \in \Omega$ and $\boldsymbol{\theta} \in \Theta$,

$$f(y|\boldsymbol{\theta}) = \exp\left[\sum_{j=1}^{p} \theta_j T_j(y) - B(\boldsymbol{\theta}) + c(y)\right].$$

What is given by the derivatives of $B(\boldsymbol{\theta})$. That is, for $j = 1, \ldots, p$,

$$\frac{\partial}{\partial \theta_j} B(\boldsymbol{\theta}) =$$

5. For random variables $Y_i;\ i = 1, \ldots, n$ that are independent and identically distributed according to a distribution having probability density function $f(y|\boldsymbol{\theta})$ with $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)^T$, the expected information in a single random variable is **defined** as a $p \times p$ matrix with uv^{th} element,

$$I_{u,v}(\boldsymbol{\theta}) =$$

6. Although we are not taking data model parameters to be random variables in a Bayesian analysis, in specifying priors and deriving posteriors it looks like that is exactly what we are doing. Why is this?

Multiple Choice

1. (15 pts.)

   Select the **one best** answer. Suppose we use least squares estimation with each of the following models – we are assuming the proper version of least squares is used, that's not the question. In what follows any quantities represented as x_i or $x_{i,j}$ are considered to be non-random covariates. For which models are small-sample or exact results available?

   (a) $Y_i = \sum_{j=1}^{p} x_{i,j}\beta_j + \sigma\epsilon_i$,
   where $\epsilon_i \sim$ iid $\mathrm{N}(0, 1)$.

   (b) $Y_i = \beta_0 + \beta_1 x_i + w_i\sigma\epsilon_i$,
   where $\epsilon_i \sim$ iid $\mathrm{N}(0, 1)$ and w_i are known constants for $i = 1, \ldots, n$.

   (c) $Y_i = g(\beta_0 + \beta_1 x_i) + \sigma\epsilon_i$,
   where $g(\cdot)$ is a known nonlinear function, and $\epsilon_i \sim$ iid $\mathrm{N}(0, 1)$.

   (d) $Y_i = \sum_{j=1}^{p} x_i^j\beta_j + \sigma\epsilon_i$,
   where $\epsilon_i \sim$ iid $\mathrm{N}(0, 1)$.

(e) $Y_i = \mu_i + \mu_i^2 \sigma \epsilon_i$,

where $\mu_i = \beta_0 + \beta_1 x_i$ and $\epsilon_i \sim$ iid $\mathrm{N}(0,1)$.

(f) Answers (a) and (d) only.

(g) Answers (a), (b) and (d) only.

(h) Answers (a) and (c) only.

(i) Answers (a), (c), and (d) only.

(j) Answers (a), (b), (c) and (d) only.

(k) Answers (a), (b), (c), (d) and (e).

Additional Questions

1. (25 pts.)

   Consider a one sample normal model for random variables $Y_1, \ldots, Y_n$ under which these random variables are independent and identically distributed with common probability density function, for some $-\infty < \mu < \infty$ and $\sigma^2 > 0$,

$$f(y|\mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left[-\frac{1}{2\sigma^2}(y-\mu)^2\right]; \quad -\infty < y < \infty.$$

   Suppose that σ^2 is considered known, and μ has been assigned a prior distribution having density with selected values of λ and τ^2,

$$\pi(\mu) = \frac{1}{(2\pi\tau^2)^{1/2}} \exp\left[-\frac{1}{2\tau^2}(\mu-\lambda)^2\right]; \quad -\infty < \mu < \infty.$$

   Derive the posterior distribution of μ.

2. (30 pts.)

   One property of exponential family distributions is that if $f(y|\boldsymbol{\theta})$ is an exponential family with $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)^T$, then for an integrable function $h(y)$,

$k = 1, \ldots, p$, and $d = 1, 2 \ldots$,

$$\frac{\partial^d}{\partial \theta_k} \int h(y) f(y|\boldsymbol{\theta})\, dy = \int \frac{\partial^d}{\partial \theta_k^d} h(y) f(y|\boldsymbol{\theta})\, dy. \tag{1}$$

Exponential family distributions are a subclass of exponential family distributions and so this property must apply to them as well. A random variable Y can be said to have an exponential dispersion family distribution if its probability mass or probability density function can be written in the form,

$$f(y|\theta, \phi) = \exp[\phi\{\theta y - b(\theta)\} + c(y, \phi)]. \tag{2}$$

Demonstrate that for an exponential dispersion family (2) the property (1) implies that,

$$E(Y) = \frac{d}{d\theta} b(\theta) = b'(\theta),$$
$$var(Y) = \frac{1}{\phi} \frac{d^2}{d\theta^2} b(\theta) = \frac{1}{\phi} b''(\theta).$$

Hint: start with,

$$\frac{d}{d\theta} \int \exp[\phi\{\theta y - b(\theta)\} + c(y, \phi)]\, dy = 0.$$