Sample Exam

1. Suppose that $X_1, \ldots, X_n \stackrel{iid}{\sim} N(\mu, 1)$, and we want to test $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$. Consider a test that rejects H_0 if either $\bar{X}_n < c_1$ or $\bar{X}_n > c_2$.

   (a) What is the power function $\beta(\mu)$ of the test? In your answer, use the notation $\Phi(z)$ to denote $P(Z \leq z)$ where $Z \sim N(0, 1)$.

   (b) Determine the values of the constants c_1 and c_2 such that $\beta(\mu_0) = 0.1$ and the function $\beta(\mu)$ is symmetric with respect to the point $\mu = \mu_0$. Again, write your answer in terms of Φ.

   (c) Show that both the size 0.1 Wald test and the size 0.1 likelihood ratio test have the form in (b).

2. Suppose we take a random sample of size n from a population of people. Let X_1 denote the number of individuals with a particular genotype AA, X_2 denote the number with Aa, and X_3 denote the number with aa. Assuming the gene frequencies are in equilibrium, the Hardy-Weinberg law says that the genotypes AA, Aa, and aa occur with probability $p_1 = (1-\theta)^2$, $p_2 = 2\theta(1-\theta)$, and $p_3 = \theta^2$, respectively.

   (a) What is the likelihood function for θ, treating $X = (X_1, X_2, X_3)$ as a sample from the multinomial distribution with size n and $p = (p_1, p_2, p_3)$?

   (b) Find the maximum likelihood estimator for θ under this model.

   (c) Find the asymptotic distribution for the maximum likelihood estimator.

3. Let $X_1, \ldots, X_n$ be an *iid* sample from the distribution with PDF

$$f(x; \theta) = (\theta + 1)x^\theta$$

   for $0 \leq x \leq 1$ and $\theta > -1$.

   (a) Find the method of moments estimate of θ.

   (b) Find the MLE of θ.

   (c) Find an estimate of the standard error of the MLE, based on an asymptotic argument.

   (d) Suppose instead of using the estimate in (c), we use the bootstrap to estimate the standard error of the MLE. Give the algorithm for doing this.

4. Consider a Bayesian model in which, conditional on unknown parameter θ, $X_1, \ldots, X_n$ are *iid* Bernoulli random variables, with $P(X_i = 1) = \theta$, and the prior distribution is $Beta(a, b)$.

   (a) Find the posterior distribution for θ, conditioning on $X_1, \ldots, X_n$. It is fine to write the family and specify its parameters; you do not need to write out the PDF or CDF.

   (b) Show that the posterior mean can be written as a weighted average of the prior mean and the MLE for θ, which is $\bar{X}_n$.

   (c) Suppose we use squared error loss, and let $\hat{\theta}_n$ be the Bayes estimator based on observing $X_1, \ldots, X_n$. Show that the sequence of Bayes estimators is consistent for θ, i.e. that $\hat{\theta}_n \xrightarrow{P} \theta$ as $n \to \infty$.

5. Let $X_1, \ldots, X_n$ be *iid* with PDF $f(x)$ on $[0, 1]$. Let h_n be a positive constant and define

$$\hat{f}(0) = \frac{Y_n}{nh_n}$$

   where Y_n is the number of observations in the interval $[0, h_n]$. Show that if $h_n \to 0$ and $nh_n \to \infty$ then $\hat{f}(0) \xrightarrow{P} f(0)$. Hints:

   - You may assume that $f(x) \approx f(0 + xf'(0)$ and that for all x, $|f'(x)| < C$ for some finite C.
   - Note that $Y_n \sim Binomial(n, p_n)$ where $p_n = \int_0^{h_n} f(x)dx$.
   - Since $Y_n \sim Binomial(n, p_n)$, $Var(Y_n) = np_n(1 - p_n)$. But since p_n is small, you can use the approximation $Var(Y_n) \approx np_n$.

6. Consider the simple linear regression model

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \epsilon_i$$

   for $i = 1, \ldots, n$, with $\epsilon_1, \ldots, \epsilon_n$ *iid* and $\epsilon_i \sim Normal(0, \sigma^2)$.

   Note: you do not need to carry out any matrix computations below, only need to identify the matrices in the computations.

   (a) Find an unbiased estimator for σ^2 (we consider $X_1, \ldots, X_n$ as given).

   (b) Find the MLE for σ^2 (we consider $X_1, \ldots, X_n$ as given).

   (c) Carry out a test for testing $H_0 : \beta_2 = 0$ vs. $H_1 : \beta_2 \neq 0$.

   (d) Describe a cross-validation method, with detailed procedures, for making a model selection between $Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \epsilon_i$ and $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$.