E0 232 Probability and Statistics *Due:* Nov 15, 2012

Assignment 6

1. A company produces light bulbs whose lifetimes (say in months) are independent random variables all distributed as $\text{Exp}(\lambda)$ for some unknown $\lambda > 0$. You buy 10 light bulbs and observe their lifetimes (in months): 10.2, 8.1, 4.3, 12.7, 11.3, 11.9, 8.7, 9.3, 1.7, 13.4. What is the maximum likelihood estimate of λ?

2. Let $X_1, \ldots, X_n \sim \text{Poisson}(\lambda)$ where λ is an unknown parameter. Find the ML estimator of λ.

3. Let $X_1, \ldots, X_n \sim \text{Unif}([a, b])$ where a, b are unknown parameters with $a < b$. Find the ML estimators of a, b and show that they are consistent.

4. A random variable X has the Gamma distribution with parameters $\alpha, \beta > 0$, written $X \sim \text{Gamma}(\alpha, \beta)$, if it has PDF given by
$$f_X(x) = \begin{cases} \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} & \text{if } x \geq 0 \\ 0 & \text{otherwise,} \end{cases}$$
where $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$.
   (a) Show that $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$. [*Hint: Use integration by parts.*]
   (b) Find $\Gamma(1)$. Use this together with the result of part (a) to find $\Gamma(n)$ for any positive integer n.
   (c) Let $Y \sim \text{Exp}(\lambda)$ for some $\lambda > 0$. Find $\alpha, \beta > 0$ such that Y can be written as $\text{Gamma}(\alpha, \beta)$.
   (d) Let Z be an Erlang random variable of order k with parameter $\lambda > 0$ (recall that Z is a sum of k independent $\text{Exp}(\lambda)$ random variables). Find $\alpha, \beta > 0$ such that Z can be written as $\text{Gamma}(\alpha, \beta)$.
   (e) Let $X \sim \text{Gamma}(\alpha, \beta)$. Find $E[X]$ and $\text{var}(X)$.
   (f) Let $X_1, \ldots, X_n \sim \text{Gamma}(2, \beta)$ where $\beta > 0$ in unknown. Find the ML estimator of β.
   (g) Consider a Poisson process with unknown rate parameter $\lambda > 0$. Your friend observes the arrival times in this process, and tells you the values of the arrival times of the first 10 even-numbered arrivals (i.e. the 2nd, 4th, 6th, and so on up to the 20th arrival times): 2.3, 3.8, 4.7, 6.3, 9.7, 11.2, 12.1, 12.9, 13.8, 15.3. What is the ML estimate of λ?

5. Your friend has a real number a that he is trying to communicate to you over a noisy channel. Due to the noise in the channel, each time he sends you the number, you observe a corrupted version of a.
   (a) Suppose you know that on each transmission, the channel adds noise distributed as $\mathcal{N}(0, 1)$ to the value sent, independent of previous transmissions.
      i. Your friend transmits his real number n times. Let $X_1, \ldots, X_n$ denote the values you observe. What is the ML estimator of a?
      ii. Suppose your friend transmits his real number 5 times, and based on the 5 values you observe, the ML estimate is $\widehat{a} = 1.2$. Give a 95% confidence interval for a.
      iii. Now suppose your friend transmits his real number 100 times, and based on the 100 values you observe, the ML estimate is again $\widehat{a} = 1.2$. How does your 95% confidence interval for a change?
   (b) Suppose you know that on each transmission, the channel adds noise distributed as $\mathcal{N}(0, 2)$ to the value sent, independent of previous transmissions. Repeat parts i-iii above.
   (c) Suppose you know that on each transmission, the channel adds noise distributed as $\text{Unif}([-1, 1])$ to the value sent, independent of previous transmissions. Repeat parts i-iii above.

6. Consider again the setting of Problem 5, where your friend has a real number a that he is trying to communicate to you over a noisy channel.

   (a) Suppose you know that on each transmission, the channel adds noise distributed as $\mathcal{N}(0, 1)$ to the value sent, independent of previous transmissions.

      i. Your friend transmits his real number n times. Let $X_1, \ldots, X_n$ denote the values you observe. Design a procedure to test the following hypotheses at a 5% significance level (i.e. with false rejection probability at most 5%):

$$\begin{aligned} H_0 : &\qquad a = 1 \\ H_1 : &\qquad a \neq 1 \end{aligned}$$

      ii. Suppose your friend transmits his real number 5 times, and you observe the values 0.9, 1.5, 1.1, 1.7, 1.6. Based on these observations, will the test you have designed allow you to reject the null hypothesis H_0 at the 5% significance level?

      iii. Suppose again your friend transmits his real number 5 times, and you observe the values 0.9, 1.5, 1.1, 1.7, 1.6. Find the p-value of your test given these observations (i.e. find the smallest value α for which the test you have designed would allow choosing a critical value such that, based on these observations, H_0 is rejected at a $(1 - \alpha)$ significance level.)

   (b) Suppose you know that on each transmission, the channel adds noise distributed as $\mathcal{N}(0, 2)$ to the value sent, independent of previous transmissions. Repeat parts i-iii above.

   (c) Suppose you know that on each transmission, the channel adds noise distributed as $\text{Unif}([-1, 1])$ to the value sent, independent of previous transmissions. Repeat parts i-iii above.

7. Consider a setting in which n_1 patients are given treatment 1 and n_2 patients are given treatment 2. Under treatment 1, each patient responds favorably with an unknown probability p_1, independent of other patients. Under treatment 2, each patient responds favorably with an unknown probability p_2, independent of other patients. Let X_1 and X_2 denote the numbers of patients who respond favorably to treatment 1 and treatment 2, respectively.

   (a) Find the ML estimators of p_1, p_2.

   (b) Let $\psi = p_1 - p_2$. Find the ML estimator of ψ.

   (c) Design a procedure to test the following hypotheses at a 1% significance level (i.e. with false rejection probability at most 1%):

$$\begin{aligned} H_0 : &\qquad p_1 = p_2 \\ H_1 : &\qquad p_1 \neq p_2 \end{aligned}$$

8. A shipment of n light bulbs is received from a supplier. Unfortunately, the supplier cannot trace whether the shipment came from company A or company B. It is known that light bulbs from company A have independent, exponentially distributed lifetimes with mean 10 months, and that light bulbs from company B have independent, exponentially distributed lifetimes with mean 12 months. Let $X_1, \ldots, X_n$ be the observed lifetimes of the n bulbs in the shipment. It is of interest to test the following hypotheses:

$$\begin{aligned} H_0 : &\qquad \text{Shipment came from company A} \\ H_1 : &\qquad \text{Shipment came from company B} \end{aligned}$$

   (a) Show that the likelihood ratio test gives a rejection region of the form

$$R = \big\{(x_1, \ldots, x_n) : x_1 + \ldots + x_n < r\big\}$$

   for some $r \in \mathbb{R}$.

   (b) Explain how you would find r corresponding to a test at the 5% significance level, i.e. such that the resulting test has false rejection probability at most 5%. Express your answer in terms of the CDF of an appropriate random variable.

9. As discussed in class, a random variable X has the Beta distribution with parameters $\alpha, \beta > 0$, written $X \sim \text{Beta}(\alpha, \beta)$, if it has PDF given by

$$f_X(x) = \begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} & \text{if } 0 \le x \le 1 \\ 0 & \text{otherwise.} \end{cases}$$

   (a) Write a simple program (e.g. in MATLAB) to plot the above PDF for the following values of α and β (you can use MATLAB's `gamma` function to compute $\Gamma(\alpha), \Gamma(\beta)$, etc):
      i. $\alpha = \beta = 1$
      ii. $\alpha = \beta = 2$
      iii. $\alpha = \beta = 3$
      iv. $\alpha = \beta = \frac{1}{2}$
      v. $\alpha = 1,\ \beta = 2$
      vi. $\alpha = 1,\ \beta = 3$
      vii. $\alpha = 1,\ \beta = \frac{1}{2}$
      viii. $\alpha = 2,\ \beta = 1$
      ix. $\alpha = 3,\ \beta = 1$
      x. $\alpha = \frac{1}{2},\ \beta = 1$
      xi. $\alpha = 2,\ \beta = 5$
      xii. $\alpha = 5,\ \beta = 2$

   Include a copy of both the code for your program and the above plots in your assignment solutions.

   (b) Suppose you assume a uniform prior on the bias of a coin, and on tossing it 10 times, obtain 8 heads. What does the posterior distribution look like? Plot this posterior density. Why does this make intuitive sense? What are the MAP and LMS estimates of the bias? What happens if you obtain 3 heads? Again, plot the posterior PDF and find the MAP and LMS estimates of the bias.

   (c) Repeat part (b) assuming a Beta(2,2) prior.

   (d) Repeat part (b) assuming a Beta(2,5) prior.

   (e) Repeat part (b) assuming a Beta(2,20) prior.

10. Let $X_1, \ldots, X_n | \Lambda = \lambda \sim \text{Exp}(\lambda)$. Assuming a $\text{Gamma}(\alpha, \beta)$ prior on Λ, find the MAP and LMS estimators of λ. Compute the mean squared error of both estimators.

11. Let $X_1, \ldots, X_n | \Lambda = \lambda \sim \text{Poisson}(\lambda)$. Design a conjugate prior for Λ and find the posterior distribution of Λ under this prior. Also find the MAP and LMS estimators of λ under this prior and compute their mean squared errors.

12. Consider again the setting in Problem 8. Suppose you have prior knowledge that 90% of the supplier's shipments come from company A and 10% from company B. Construct the MAP test in this setting and find its probability of error.