UC Berkeley
Department of Electrical Engineering and Computer Sciences

ELECTRICAL ENGINEERING 126: PROBABILITY AND RANDOM PROCESSES

Problem Set 8
Fall 2017

Self-Graded Scores Due: 5 PM, Monday, November 6, 2017
Submit your self-graded scores via the Google form:
https://goo.gl/forms/uXZIuKC75HftnnKi2.
Make sure you use your **SORTABLE NAME** on CalCentral.

---

1. **Estimating an Exponential Distribution**

You draw a sample $X_1, \dots, X_n$ (n is a positive integer) for the lifetime of a light bulb (assumed to be exponentially distributed).

(a) Frankly, you are beginning to think that your boss is unreasonable. She makes the following demands:

- You have exactly 3 samples, with $X_1 = 3$, $X_2 = 7$, $X_3 = 5$. That is, the first light bulb dies after 3 seconds, the second light bulb dies after 7 seconds, and the third light bulb dies after 5 seconds.
- Your confidence interval is specified to be the interval $(0, \varepsilon)$.
- You are no longer estimating the mean lifetime; now you are estimating the rate λ at which the light bulbs die.
- Your confidence interval must be *exact*: you may not use any inequalities or approximations.

What is your reported confidence, in terms of ε? [Please do not leave your answers in terms of an integral.]

(b) In the same setting as the previous part, can you give an exact confidence interval for the mean light bulb lifetime with the same confidence level as before? (*Hint*: You don't need to complete Part 1.a to answer this successfully.)

Solution:

(a) **Note**: The original wording of the question was confusing so please be lenient with the self-grades. The intended interpretation of the problem is as follows: find the confidence level $1 - \delta$ (where $\delta \in (0, 1)$) such that, if you had been asked to produce an exact $1 - \delta$ confidence interval $C(X_1, X_2, X_3)$ for λ using the samples (X_1, X_2, X_3), then upon observing $(X_1, X_2, X_3) = (3, 7, 5)$ you would have given the confidence interval $C(3, 7, 5) = (0, \varepsilon)$.

First, we make the observation that if X has the exponential distribution with parameter λ, then λX has the exponential distribution with parameter 1. Indeed,

$$\mathbb{P}(\lambda X \le x) = \mathbb{P}\Big(X \le \frac{x}{\lambda}\Big) = 1 - \mathrm{e}^{-\lambda(x/\lambda)} = 1 - \mathrm{e}^{-x}.$$

Noting that $X_1 + X_2 + X_3$ has the Erlang distribution with rate λ and order 3, we see that $\lambda(X_1 + X_2 + X_3)$ has the Erlang distribution with rate 1 and order 3.

Our goal is a $1 - \delta$ confidence interval of the form

$$C(X_1, X_2, X_3) = \big(0, c(X_1, X_2, X_3)\big),$$

where $c(X_1, X_2, X_3) > 0$, the upper limit for our confidence interval, is a random function of the observed data (X_1, X_2, X_3). This confidence interval should have the guarantee that for all $\lambda > 0$,

$$\mathbb{P}_\lambda\big(\lambda \in C(X_1, X_2, X_3)\big) = \mathbb{P}_\lambda\big(\lambda \le c(X_1, X_2, X_3)\big) \ge 1 - \delta \tag{1}$$

(where the notation $\mathbb{P}_\lambda$ denotes the probability when $X_1, X_2, X_3 \overset{\text{i.i.d.}}{\sim} \text{Exponential}(\lambda)$). We will take $c(X_1, X_2, X_3) = \alpha(X_1 + X_2 + X_3)^{-1}$ for some α that will be chosen later. With this choice of $c(X_1, X_2, X_3)$, then

$$\mathbb{P}_\lambda\Big(\lambda \le \frac{\alpha}{X_1 + X_2 + X_3}\Big) = \mathbb{P}_\lambda\big(\lambda(X_1 + X_2 + X_3) \le \alpha\big)$$

and the latter expression depends only on α, not on the parameter λ (since we observed earlier that $\lambda(X_1 + X_2 + X_3)$ is Erlang with rate 1, order 3). This is why we can achieve an *exact* confidence interval: we can now adjust α so that (1) is satisfied, regardless of what λ is.

To compute $\mathbb{P}_\lambda(\lambda(X_1 + X_2 + X_3) \le \alpha)$, we interpret this in terms of the Poisson process with rate 1. We need the third arrival to arrive before time α, which is the complement of the event that we have at most two arrivals in the time interval $(0, \alpha)$. The number of arrivals has the Poisson distribution with mean α, so

$$\mathbb{P}_\lambda\big(\lambda \in C(X_1, X_2, X_3)\big) = 1 - \mathrm{e}^{-\alpha}\Big(1 + \alpha + \frac{\alpha^2}{2}\Big).$$

In this question, we are asked to find the confidence level $1 - \delta$ such that the interval is $(0, \varepsilon)$. Since our interval is

$$\big(0, c(X_1, X_2, X_3)\big) = \big(0, \alpha(X_1 + X_2 + X_3)^{-1}\big),$$

we then have $\alpha = \varepsilon(X_1 + X_2 + X_3) = 15\varepsilon$. Hence, our confidence level is $1 - \mathrm{e}^{-15\varepsilon}(1 + 15\varepsilon + 225\varepsilon^2/2)$.

(b) If $\lambda \in (0, \varepsilon)$ with probability C (where C is your confidence level), then $\lambda^{-1} \in (\varepsilon^{-1}, \infty)$ with the same probability C, so our confidence interval is $(\varepsilon^{-1}, \infty)$.

2. **Chernoff Bound Application: Load Balancing**

   Here, we will give an application for the Chernoff bound which is instrumental for calculating confidence intervals. However, we will need a slightly more general version of the bound that works for any Bernoulli random variables. For any positive integer n, if $X_1, \ldots, X_n$ are i.i.d. Bernoulli, with $\mathbb{P}(X_i = 1) = p$, and $S_n = \sum_{i=1}^{n} X_i$, then the following bound holds for $0 \le \varepsilon \le 1$:

   $$\mathbb{P}\big(S_n > (1+\varepsilon)np\big) \le \exp\Big(-\frac{\varepsilon^2 np}{3}\Big). \tag{2}$$

   You may take (2) as a fact (or try to prove it on your own if you want!).

   Here is the setting: there are k (k a positive integer) servers and n users. The simplest load balancing scheme is simply to assign each user to a server chosen uniformly at random (think of the users as "balls" and we are tossing them into server "bins"). By using the union bound, show that with probability at least $1 - 1/k^2$, the maximum load of any server is at most $n/k + 3\sqrt{\ln k}\sqrt{n/k}$.

   Solution:

   Take $\varepsilon = 3\sqrt{k \ln k}/\sqrt{n}$. Let A_i denote the event that the load of the ith server is $> n/k + 3\sqrt{\ln k}\sqrt{n/k}$, for $i = 1, \ldots, n$. Then,

   $$\mathbb{P}(A_i) \le \exp\Big(-\frac{9k \ln k}{n} \cdot \frac{n}{3k}\Big) = \frac{1}{k^3},$$

   so taking the union bound gives

   $$\mathbb{P}(A_1 \cup \cdots \cup A_k) \le \sum_{i=1}^{k} \mathbb{P}(A_i) = k \cdot \frac{1}{k^3} = \frac{1}{k^2}.$$

   The union bound gives a pretty good result, but only because the initial Chernoff bound was quite strong! If we had tried to achieve the same result using Chebyshev's inequality, we would obtain a bound on $\mathbb{P}(A_i)$ of the order $\mathcal{O}(1/\ln k)$, so the union bound wouldn't give us anything useful at all.

   From this result, we can see that the naïve load balancing scheme actually performs well: the deviation from optimal performance is on the order of $\widetilde{\mathcal{O}}(\sqrt{n/k})$.

3. **Basic Properties of Jointly Gaussian Random Variables**

   Prove that a collection of jointly Gaussian random variables $X_1, \ldots, X_n$ (n is a positive integer) are independent if and only if they are uncorrelated. Also show that any linear combination of these random variables will be a Gaussian random variable. [*Hint*: For first part, use the characteristic function definition, and look at the covariance matrix for uncorrelated RVs.]

   Solution:

   Jointly Gaussian random variables can be defined in the following way: $X = (X_1, \ldots, X_n)$ is jointly Gaussian with mean $\mu = \begin{bmatrix}\mu_1 & \cdots & \mu_n\end{bmatrix}^{\mathsf{T}}$ and covariance matrix $C = \begin{bmatrix}C_{i,j}\end{bmatrix}_{i,j=1}^{n}$ where $C_{i,j} = \mathbb{E}[(X_i - \mu_i)(X_j - \mu_j)]$ if and only if the joint characteristic function is given by

   $$\phi_X(u) = \mathbb{E}\big(\exp(\mathrm{i}\langle u, X\rangle)\big) = \exp\Big(\mathrm{i}\langle u, \mu\rangle - \frac{1}{2}u^{\mathsf{T}} C u\Big)$$

where $u = \begin{bmatrix} u_1 & \cdots & u_n \end{bmatrix}^\mathsf{T}$ and $\langle \cdot, \cdot \rangle$ is the inner product.

If the components of X are uncorrelated, $C_{i,j} = \mathbb{E}[(X_i - \mu_i)(X_j - \mu_j)] = 0$ for $i \neq j$. In other words, C is diagonal. The characteristic function then can be written as

$$\phi_X(u) = \prod_{i=1}^{n} \phi_{X_i}(u_i)$$

where $\phi_{X_i}(u_i)$ is the characteristic function of the random variable X_i, i.e., in other words the characteristic function decomposes in product form, which is an alternate definition of independence. Hence, $X_1, \ldots, X_n$ are independent if they are uncorrelated.

If $X_1, \ldots, X_n$ are independent, then, $\mathbb{E}[(X_i - \mu_i)(X_j - \mu_j)] = \mathbb{E}(X_i - \mu_i)\mathbb{E}(X_j - \mu_j) = 0$. Hence, by definition, they are uncorrelated.

Consider, $Y = \sum_{i=1}^n a_i X_i$ and define $a = \begin{bmatrix} a_1 & \cdots & a_n \end{bmatrix}^\mathsf{T}$. One needs to show that Y is Gaussian. The characteristic function of Y is

$$\begin{aligned} \phi_Y(u) &= \mathbb{E}\Big[\exp\Big(\mathrm{i}u \sum_{i=1}^{n} a_i X_i\Big)\Big] = \phi_X(ua_1, \ldots, ua_n) \\ &= \exp\Big(\mathrm{i}u\langle a, \mu\rangle - \frac{1}{2}u^2 a^\mathsf{T} C a\Big). \end{aligned}$$

Therefore, $Y \sim \mathcal{N}(\langle a, \mu\rangle, a^\mathsf{T} C a)$.

4. **Gaussian Hypothesis Testing**

Consider a hypothesis testing problem that if $X = 0$, you observe a sample of $\mathcal{N}(\mu_0, \sigma^2)$, and if $X = 1$, you observe a sample of $\mathcal{N}(\mu_1, \sigma^2)$, where $\mu_0, \mu_1 \in \mathbb{R}$, $\sigma^2 > 0$. Find the Neyman-Pearson test for false alarm $\alpha \in (0, 1)$, that is, $\mathbb{P}(\hat{X} = 1 \mid X = 0) \leq \alpha$.

Solution:

Let y be the observation. We know that

$$f_i(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-(y-\mu_i)^2/(2\sigma^2)}, \qquad i = 0, 1.$$

Thus, the likelihood ratio is

$$\frac{f_1(y)}{f_0(y)} = \mathrm{e}^{-((y-\mu_1)^2 - (y-\mu_0)^2)/(2\sigma^2)}.$$

Without loss of generality suppose that $\mu_1 > \mu_0$. Then, solving

$$\frac{f_1(y)}{f_0(y)} > \lambda$$

and taking the logarithm of both sides we have

$$y > \frac{\sigma^2}{\mu_1 - \mu_0} \ln \lambda + \frac{\mu_1 + \mu_0}{2} = t.$$

We define the left hand side of the above equation as some threshold t. Now we want to find t such that the false alarm is α.

$$\mathbb{P}(\hat{X} = 1 \mid X = 0) = \int_t^{\infty} f_0(y)\,\mathrm{d}y = Q\Big(\frac{t-\mu_0}{\sigma}\Big) = \alpha$$

Thus, $t = \sigma Q^{-1}(\alpha) + \mu_0$. Here $Q = 1 - \Phi$, where Φ is the CDF of Gaussian distribution.

5. **Hypothesis Test for Uniform Distribution**

If $X = 0$, $Y \sim \text{Uniform}[-1, 1]$ and if $X = 1$, $Y \sim \text{Uniform}[0, 2]$. Solve a hypothesis testing problem so that the probability of false alarm is less than or equal $\beta \in (0, 1)$.

Solution:

Here, the likelihood ratio is

$$\frac{f_{Y|X}(y \mid 1)}{f_{Y|X}(y \mid 0)} = \frac{\mathbb{1}\{0 \le y \le 2\}}{\mathbb{1}\{-1 \le y \le 1\}}.$$

Thus, $\hat{X} = 1$ if $Y > 1$ and $\hat{X} = 0$ if $Y < 0$. If $Y \in [0, 1]$ we need randomization, so $\hat{X} = 1$ with some probability γ. We choose γ such that

$$\mathbb{P}(\hat{X} = 1 \mid X = 0) = \beta.$$

That is,

$$\gamma \mathbb{P}(Y \in [0, 1] \mid X = 0) = \frac{\gamma}{2} = \beta.$$

Thus, $\gamma = 2\beta$.

6. **BSC Hypothesis Testing**

You are testing a digital link that corresponds to a BSC with some error probability $\epsilon \in [0, 0.5)$. You observe n inputs and outputs of the BSC, where n is a positive integer. You want to solve a hypothesis problem to detect that $\epsilon > 0.1$ with a probability of false alarm at most equal to 0.05. Assume that n is very large and use the CLT.

Solution:

We observe $x_1, \dots, x_n$ and $y_1, \dots, y_n$. Let x and y be the vectors of these observations. Then, the likelihood is

$$\mathbb{P}(Y = y \mid X = x) = \epsilon^{\sum_{i=1}^n \mathbb{1}\{y_i \ne x_i\}} (1-\epsilon)^{\sum_{i=1}^n \mathbb{1}\{y_i = x_i\}}.$$

What matters for estimating ϵ is $t := \sum_{i=1}^n \mathbb{1}\{x_i \ne y_i\}$. For the hypothesis testing, we define random variable H, where $H = \mathbb{1}\{\epsilon > 0.1\}$. If t is large, ϵ is more likely to be larger, so we set the hypothesis testing to be $\hat{H} = \mathbb{1}\{t > \lambda\}$, for some λ to be chosen such that probability of false alarm is 0.05. Since n is very large, $T = \sum_{i=1}^n \mathbb{1}\{X_i \ne Y_i\}$ is approximately a normal random variable. Note that without the approximation T is a binomial since input-output pairs of the channel are independent. Now we calculate the following.

$$\mathbb{P}(X_1 \ne Y_1) = \mathbb{P}(Y_1 = 1 \mid X_1 = 0)\mathbb{P}(X_1 = 0) + \mathbb{P}(Y_1 = 0 \mid X_1 = 1)\mathbb{P}(X_1 = 1)$$

$$= \epsilon$$

Thus, $T \sim \mathcal{N}(n\epsilon, n\epsilon(1-\epsilon))$. Now we set the false alarm.

$$\mathbb{P}(\hat{H} = 1 \mid H = 0) = \mathbb{P}(T > \lambda \mid \epsilon < 0.1) \leq 0.05.$$

To find λ, we solve a stronger equation which is $\mathbb{P}(T > \lambda \mid \epsilon = 0.1) \leq 0.05$. Note that

$$\mathbb{P}(T > \lambda \mid \epsilon < 0.1) \leq \mathbb{P}(T > \lambda \mid \epsilon = 0.1) = 0.05.$$

Thus,

$$\mathbb{P}\big(\mathcal{N}(0.1n, 0.09n) > \lambda\big) = Q\Big(\frac{\lambda - 0.1n}{\sqrt{0.09n}}\Big) = Q(1.67) = 0.05.$$

Thus, $\lambda = 0.1n + 1.67\sqrt{0.09n}$.