UC Berkeley
Department of Electrical Engineering and Computer Sciences

Electrical Engineering 126: Probability and Random Processes

Problem Set 9
Spring 2017

Self-Graded Scores Due: 5pm April 17, 2017
Submit your self-graded scores via the Google form:
https://goo.gl/forms/iLewm2KVRnGqsyN13.
Make sure you use your **SORTABLE NAME** on bCourses.

1. **Bayesian Estimation of Exponential Distribution**

   We have already learned about MLE (non-Bayesian perspective) and MAP (Bayesian perspective). In this problem, we will introduce the fully Bayesian approach to statistical estimation.

   Suppose that X is an exponential random variable with unknown rate Λ (Λ is a random variable). As a Bayesian practitioner, you have a prior belief that Λ is equally likely to be λ_1 or λ_2.

   You collect one sample X_1 from X.

   (a) Find the posterior distribution $\Pr(\Lambda = \lambda_1 \mid X_1 = x_1)$.

   (b) If we were using the MLE or MAP rule, then we would choose a single value λ for Λ; this is sometimes called a *point estimate*. This amounts to saying X has the exponential distribution with rate λ.

   In the Bayesian approach, we will not use a point estimate. Instead, we will keep the full information of the posterior distribution of Λ, and we compute the distribution of X as

$$f_X(x) = \sum_{\lambda \in \{\lambda_1, \lambda_2\}} f_{X|\Lambda}(x \mid \lambda) \Pr(\Lambda = \lambda \mid X_1 = x_1).$$

   Notice that in the Bayesian approach, we do not necessarily have an exponential distribution for X anymore. Compute $f_X(x)$ in closed-form.

   (c) You might guess from the previous part that the fully Bayesian approach is often computationally intractable. This is one of the main reasons why point estimates are common in practice.

   Compute the MAP estimate for Λ and calculate $f_X(x)$ again using the MAP rule.

   Solution:

(a) The likelihood of the data is

$$f_{X_1|\Lambda}(x_1 \mid \lambda) = \lambda e^{-\lambda x_1}.$$

The prior distribution for Λ is $\Pr(\Lambda = \lambda_1) = \Pr(\Lambda = \lambda_2) = 1/2$. Therefore, the posterior distribution is

$$\begin{aligned}\Pr(\Lambda = \lambda_1 \mid X_1 = x_1) &= \frac{(1/2)\lambda_1 e^{-\lambda_1 x_1}}{(1/2)\lambda_1 e^{-\lambda_1 x_1} + (1/2)\lambda_2 e^{-\lambda_2 x_1}} \\ &= \frac{\lambda_1 e^{-\lambda_1 x_1}}{\lambda_1 e^{-\lambda_1 x_1} + \lambda_2 e^{-\lambda_2 x_1}}.\end{aligned}$$

(b) We compute

$$\begin{aligned}f_X(x) &= \frac{\lambda_1 e^{-\lambda_1 x}\lambda_1 e^{-\lambda_1 x_1}}{\lambda_1 e^{-\lambda_1 x_1} + \lambda_2 e^{-\lambda_2 x_1}} + \frac{\lambda_2 e^{-\lambda_2 x}\lambda_2 e^{-\lambda_2 x_1}}{\lambda_1 e^{-\lambda_1 x_1} + \lambda_2 e^{-\lambda_2 x_1}} \\ &= \frac{\lambda_1^2 e^{-\lambda_1(x+x_1)} + \lambda_2^2 e^{-\lambda_2(x+x_1)}}{\lambda_1 e^{-\lambda_1 x_1} + \lambda_2 e^{-\lambda_2 x_1}}, \qquad x \ge 0.\end{aligned}$$

(c) We have already calculated $\Pr(\Lambda = \lambda \mid X_1 = x_1)$. The MAP rule says to choose the value of λ which maximizes this posterior probability, i.e. choose λ_1 if $\lambda_1 e^{-\lambda_1 x_1} > \lambda_2 e^{-\lambda_2 x_1}$. Assume WLOG that $\lambda_1 > \lambda_2$.

$$\text{MAP}[\Lambda \mid X_1] = \lambda_1 1\left\{X_1 < \frac{\ln(\lambda_1/\lambda_2)}{\lambda_1 - \lambda_2}\right\} + \lambda_2 1\left\{X_1 > \frac{\ln(\lambda_1/\lambda_2)}{\lambda_1 - \lambda_2}\right\}.$$

In the MAP rule, X is exponential with rate given by $\text{MAP}[\Lambda \mid X_1]$. Hence,

$$f_X(x) = \begin{cases}\lambda_1 e^{-\lambda_1 x}, & X_1 < \dfrac{\ln(\lambda_1/\lambda_2)}{\lambda_1 - \lambda_2}, \\ \lambda_2 e^{-\lambda_2 x}, & X_1 > \dfrac{\ln(\lambda_1/\lambda_2)}{\lambda_1 - \lambda_2},\end{cases} \qquad \text{for } x \ge 0.$$

2. Flipping Coins and Hypothesizing

You flip a coin until you see heads. Let $X = 0$ be the hypothesis that the bias of the coin (the probability of heads) is p, and $X = 1$ be the hypothesis that the bias of the coin is q, for $q > p$. Solve the hypothesis testing problem: maximize $\Pr[\hat{X} = 1 \mid X = 1]$ subject to $\Pr[\hat{X} = 1 \mid X = 0] \le \beta$ for $\beta \in [0, 1]$.

Solution:

Let Y be the number of flips until we see heads. Write the likelihood ratio.

$$L(y) = \frac{\Pr[Y = y \mid X = 1]}{\Pr[Y = y \mid X = 0]} = \frac{(1-q)^{y-1}q}{(1-p)^{y-1}p},$$

which is strictly decreasing in y since $q > p$. Hence, the hypothesis testing rule is of the form $\hat{X} = 1$ if $Y < \alpha$ for some α. Observe that

$$\Pr[Y < \alpha \mid X = 0] = \sum_{y=1}^{\alpha-1} p(1-p)^{y-1} = 1 - (1-p)^{\alpha-1}.$$

Therefore, we should choose α such that $1-(1-p)^{\alpha-1} \le \beta$, i.e.

$$\alpha \le 1 + \frac{\log(1-\beta)}{\log(1-p)}.$$

Therefore, take $\alpha = \lfloor 1 + \log(1-\beta)/\log(1-p) \rfloor$. For the randomization, let $\Pr[\hat{X} = 1 \mid Y = \alpha] = \gamma$. The probability of false detection is

$$\begin{aligned}\Pr[\hat{X} = 1 \mid X = 0] &= \Pr[Y < \alpha \mid X = 0] + \gamma \Pr[Y = \alpha \mid X = 0] \\ &= 1 - (1-p)^{\alpha-1} + \gamma p (1-p)^{\alpha-1} \le \beta,\end{aligned}$$

so we take

$$\gamma = \frac{\beta - 1 + (1-p)^{\alpha-1}}{p(1-p)^{\alpha-1}}.$$

Hence, for the values of α and γ described above,

$$\hat{X} = \begin{cases} 1, & Y < \alpha, \\ Z, & Y = \alpha, \\ 0, & Y > \alpha, \end{cases}$$

where Z is 1 with probability γ and 0 otherwise.

3. **Hypothesis Test for Uniform Distribution**

If $X = 0$, $Y \sim U[-1,1]$ and if $X = 1$, $Y \sim U[0,2]$. Solve a hypothesis testing problem so that the probability of false alarm is less than or equal β.

Solution:

Here, the likelihood ratio is

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

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

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

That is,

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

Thus, $\gamma = 2\beta$.

4. **Sufficient Statistics**

Suppose $X_1, \ldots, X_n$ are i.i.d. samples drawn from a probability distribution parameterized by θ (we are in the non-Bayesian setting, so θ is deterministic, but unknown). A statistic $T(X_1, \ldots, X_n)$ is a *sufficient statistic* for θ if for all t, the conditional distribution of $X_1, \ldots, X_n$ given $T = t$ does not depend on θ. Intuitively, $T(X_1, \ldots, X_n)$ "captures all that there is to know about θ from the sample $X_1, \ldots, X_n$".

(a) Let $X_1, \ldots, X_n$ be drawn from a Poisson distribution with mean μ. Show that $T = \sum_{i=1}^n X_i$ is a sufficient statistic for μ.

(b) Let T be a sufficient statistic for θ. Let $\hat{\theta}$ be an estimator for θ with $E[\hat{\theta}^2] < \infty$. Prove that for all θ,

$$E[(E[\hat{\theta} \mid T] - \theta)^2] \leq E[(\hat{\theta} - \theta)^2].$$

Remark: The above result states that $E[\hat{\theta} \mid T]$ is at least as good as $\hat{\theta}$ at estimating θ, in a mean-squared error sense. Since $E[\hat{\theta} \mid T]$ is a function of T, the result implies that we should be looking for estimators of θ that are functions of sufficient statistics.

Solution:

(a) We compute $\Pr(X_1 = x_1, \ldots, X_n = x_n \mid T = t)$. If $t \neq \sum_{i=1}^n x_i$, then the probability is 0; otherwise

$$\Pr(X_1 = x_1, \ldots, X_n = x_n, T = t) = \prod_{i=1}^n \frac{e^{-\mu} \mu^{x_i}}{x_i!} = \frac{e^{-n\mu} \mu^t}{\prod_{i=1}^n x_i!}$$

and

$$\Pr(T = t) = \frac{e^{-n\mu} (n\mu)^t}{t!},$$

since T is Poisson with mean $n\mu$. So:

$$\Pr(X_1 = x_1, \ldots, X_n = x_n \mid T = t) = \begin{cases} 0, & \sum_{i=1}^n x_i \neq t \\ t!/(n^t \prod_{i=1}^n x_i!), & \sum_{i=1}^n x_i = t \end{cases}$$

In either case, the conditional distribution has no dependence on μ.

(b) Observe that

$$\begin{aligned} E[(\hat{\theta} - \theta)^2] &= E[(\hat{\theta} - E[\hat{\theta}] + E[\hat{\theta}] - \theta)^2] \\ &= E[(\hat{\theta} - E[\hat{\theta}])^2] + E[(E[\hat{\theta}] - \theta)^2] = \operatorname{var}(\hat{\theta}) + E[(E[\hat{\theta}] - \theta)^2]. \end{aligned}$$

since $E[\hat{\theta} - E[\hat{\theta}]] = 0$. In machine learning contexts, this is commonly known as the bias-variance decomposition. Similarly, we have

$$\begin{aligned} E[(E[\hat{\theta} \mid T] - \theta)^2] &= \operatorname{var}(E[\hat{\theta} \mid T]) + E[(E[E[\hat{\theta} \mid T]] - \theta)^2] \\ &= \operatorname{var}(E[\hat{\theta} \mid T]) + E[(E[\hat{\theta}] - \theta)^2], \end{aligned}$$

by the law of iterated expectation. Hence, we see that it suffices to prove $\operatorname{var}(\hat{\theta}) \geq \operatorname{var}(E[\hat{\theta} \mid T])$, but this is immediate from the decomposition

$$\operatorname{var}(\hat{\theta}) = \operatorname{var}(E[\hat{\theta} \mid T]) + E[\operatorname{var}(\hat{\theta} \mid T)].$$

This result is known as the Rao-Blackwell Theorem.

5. **Gaussian LLSE**

The random variables X, Y, Z are i.i.d. $\mathcal{N}(0,1)$.

(a) Find $L[X^2 + Y^2 \mid X + Y]$.

(b) Find $L[X + 2Y \mid X + 3Y + 4Z]$.

(c) Find $L[(X + Y)^2 \mid X - Y]$.

Solution:

(a) We note that

$$E[(X^2 + Y^2)(X + Y)] = E[X^3 + X^2 Y + XY^2 + Y^3] = 0.$$

Hence,

$$\text{cov}(X^2 + Y^2, X + Y) = 0,$$

so that

$$L[X^2 + Y^2 \mid X + Y] = E[X^2 + Y^2] = 2.$$

(b) We find

$$\text{cov}(X + 2Y, X + 3Y + 4Z) = E[(X + 2Y)(X + 3Y + 4Z)] = 1 + 6 = 7$$

and

$$\text{var}(X + 3Y + 4Z) = 1 + 9 + 16 = 26.$$

Hence,

$$L[X + 2Y \mid X + 3Y + 4Z] = \frac{7}{26}(X + 3Y + 4Z).$$

(c) We observe that $\text{cov}(X + Y, X - Y) = 0$, so that $X + Y$ and $X - Y$ are independent. Hence,

$$L[(X + Y)^2 \mid X - Y] = E[(X + Y)^2] = 2.$$

6. **Photodetector LLSE**

Consider a photodetector in an optical communications system that counts the number of photons arriving during a certain interval. A user conveys information by switching a photon transmitter on or off. Assume that the probability of the transmitter being on is p. If the transmitter is on, the number of photons transmitted over the interval of interest is a Poisson random variable Θ with mean λ, and if it is off, the number of photons transmitted is 0. Unfortunately, regardless of whether or not the transmitter is on or off, photons may be detected due to "shot noise". The number N of detected shot noise photons is a Poisson random variable N with mean μ. Given the number of detected photons, find the LLSE of the number of transmitted photons.

Solution:

Let T be the number transmitted photons and D be the number of detected photons. We are looking for:

$$L[T \mid D] = E[T] + \frac{\text{cov}(T, D)}{\text{var}(D)}(D - E[D])$$

We find each of these terms separately. We can see by the tower property that $E[T] = p\lambda$. Now, we have:

$$\begin{aligned}
\text{cov}(T, D) &= E[(T - E[T])(D - E[D])] \\
&= E[(T - E[T])(T - E[T] + N - E[N])] \\
&= E[(T - E[T])^2] + E[(T - E[T])(N - E[N])] \\
&= \text{var}(T) \\
&= p(\lambda^2 + \lambda) - (p\lambda)^2
\end{aligned}$$

where the second to last equality follows since T and N are independent. We now find:

$$\begin{aligned}
\text{var}(D) &= \text{var}(T + N) \\
&= \text{var}(T) + \text{var}(N) \\
&= p(\lambda^2 + \lambda) - (p\lambda^2) + \mu
\end{aligned}$$

Finally, we have $E[D] = E[T] + E[N] = p\lambda + \mu$. Putting these together, we have the LLSE (no need to simplify the equation).