UC Berkeley
Department of Electrical Engineering and Computer Sciences

Electrical Engineering 126: Probability and Random Processes

Discussion 10
Fall 2017

---

1. **Statistical Estimation**

Given $X \in \{0, 1\}$, the random variable Y is exponentially distributed with rate $3X + 1$.

(a) Assume $\mathbb{P}(X = 1) = p \in (0, 1)$ and $\mathbb{P}(X = 0) = 1 - p$. Find the MAP estimate of X given Y.

(b) Find the MLE of X given Y.

(c) Solve the hypothesis testing problem of X given Y with a probability of false alarm at most 0.1. That is, find $\hat{X}$ as a function of Y that maximizes $\mathbb{P}(\hat{X} = 1 \mid X = 1)$ subject to $\mathbb{P}(\hat{X} = 1 \mid X = 0) \leq 0.1$.

(d) For what value of p does one have the same solution for (a) and (c)?

Solution:

(a) We know that when $X = 0$, $f_{Y|X}(y \mid 0) = \exp(-y)\mathbb{1}\{y > 0\}$ and when $X = 1$, $f_{Y|X}(y \mid 1) = 4\exp(-y)\mathbb{1}\{y > 0\}$. The MAP maximizes $f_{X|Y}(x, y)$ over x for the given observation y, which is equivalent to maximizing $f_{X,Y}(x, y)$. Thus, $f_{X,Y}(0, y) = (1 - p)\exp(-y)\mathbb{1}\{y > 0\}$, $f_{X,Y}(1, y) = 4p\exp(-4y)$, and

$$\mathrm{MAP}[X \mid Y] = 1 \iff 4p\exp(-4Y) > (1 - p)\exp(-Y)$$

which gives

$$\mathrm{MAP}[X \mid Y] = \mathbb{1}\Big\{Y < \frac{1}{3}\ln\frac{4p}{1 - p}\Big\}.$$

(b)

$$\mathrm{MLE}[X \mid Y] = \mathbb{1}\Big\{Y < \frac{1}{3}\ln 4\Big\} = \mathbb{1}\{Y < 0.462\}.$$

(c) Declare 1 if $Y < -\ln 0.9$ and 0 otherwise.

(d) $p = 1/(1 + 4(0.9)^3) = 0.255$.

2. **Exponential MLE, MAP, Hypothesis Testing**

The random variable X is exponentially distributed with mean 1. Given X, the random variable Y is exponentially distributed with rate X.

(a) Find $\mathrm{MLE}[X \mid Y]$;

(b) Find $\text{MAP}[X \mid Y]$;

(c) Solve the following hypothesis testing problem:

$$\text{Maximize } \mathbb{P}(\hat{X} = 1 \mid X = 1) \text{ subject to } \mathbb{P}(\hat{X} = 1 \mid X = a) \le 5\%$$

where $a > 1$ is given.

Solution:

(a) The density of Y, given $X = x$, is $f(y) = x\exp(-xy)$ for $y > 0$, so $\ln f(y) = \ln x - xy$. To maximize this over x, we differentiate to obtain $1/x - y = 0$, so $x = 1/y$, that is, $\text{MLE}[X \mid Y] = 1/Y$.

(b) The posterior density of X is

$$\begin{aligned} f_{X|Y}(x \mid y) \propto f_{Y|X}(y \mid x) f_X(x) &= x\exp(-xy)\exp(-x) \\ &= x\exp\big(-x(1+y)\big) \end{aligned}$$

so we can maximize $\ln x - x(1+y)$ over x. Differentiating, we have $1/x - 1 - y = 0$, or $1/x = 1 + y$. Hence, $\text{MAP}[X \mid Y] = 1/(1+Y)$.

(c) The likelihood ratio is

$$\frac{f_{Y|X}(y \mid a)}{f_{Y|X}(y \mid 1)} = \frac{a\exp(-ay)}{\exp(-y)} = a\exp\big(-(a-1)y\big),$$

which is decreasing with y. Our decision rule is of the form:

$$\hat{X} = \begin{cases} 1, & Y \ge y^* \\ a, & Y \le y^* \end{cases}$$

So, given $Y = y^*$,

$$\mathbb{P}(\hat{X} = 1 \mid X = a) = \mathbb{P}(Y \ge y^* \mid X = a) = \exp(-ay^*) \le \frac{1}{20},$$

so $-ay^* = -\ln 20$, that is, $y^* = (\ln 20)/a$.

3. **Laplace Prior & ℓ^1-Regularization**

Suppose you draw n i.i.d. data points $(x_1, y_1), \dots, (x_n, y_n)$, where n is a positive integer and the true relationship is $Y = WX + \varepsilon$, $\varepsilon \sim \mathcal{N}(0, \sigma^2)$. (That is, Y has a linear dependence on X, with additive Gaussian noise.) Further suppose that W has a prior distribution with density

$$f_W(w) = \frac{1}{2\beta}\mathrm{e}^{-|w|/\beta}, \qquad \beta > 0.$$

(This is known as the **Laplace distribution**.) Show that finding the MAP estimate of W given the data points $\{(x_i, y_i) : i = 1, \dots, n\}$ is equivalent to minimizing the cost function

$$J(w) = \sum_{i=1}^{n}(y_i - wx_i)^2 + \lambda|w|$$

(you should determine what λ is). This is interpreted as a one-dimensional ℓ^1-regularized least-squares criterion, also known as LASSO.

Solution:

The likelihood for W is

$$\mathcal{L}\big(w \mid (x_1, y_1), \ldots, (x_n, y_n)\big) \propto \mathcal{L}\big((x_1, y_1), \ldots, (x_n, y_n) \mid W = w\big) f_W(w)$$

(technically, the expression on the right should be divided by the likelihood of the data, but this has no dependence on w, so we omit the denominator for simplicity)

$$= \prod_{i=1}^{n} \mathcal{L}\big((x_i, y_i) \mid W = w\big) f_W(w)$$

(the data points are conditionally independent given W)

$$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-(y_i - wx_i)^2/(2\sigma^2)} \cdot \frac{1}{2\beta} \mathrm{e}^{-|w|/\beta}$$

(here we say that the likelihood of (x_i, y_i) given W is the density of ε_i, which is $\mathcal{N}(0, \sigma^2)$, evaluated at $y_i - wx_i$)

$$\propto \prod_{i=1}^{n} \mathrm{e}^{-(y_i - wx_i)^2/(2\sigma^2)} \mathrm{e}^{-|w|/\beta}$$

(again, we throw out constant factors that do not depend on the data points or w).

We wish to maximize this expression w.r.t. w, but we will find it more convenient to take the log-likelihood instead.

$$\ell\big(w \mid (x_1, y_1), \ldots, (x_n, y_n)\big) = -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - wx_i)^2 - \frac{1}{\beta}|w|.$$

Since we want to *maximize* the log-likelihood, this is equivalent to *minimizing* the cost function

$$J(w) = \sum_{i=1}^{n} (y_i - wx_i)^2 + \lambda |w|,$$

where $\lambda = 2\sigma^2/\beta$.