Massachusetts Institute of Technology
Department of Electrical Engineering and Computer Science
6.437 Inference and Information
Spring 2015

Problem Set 4

Issued: Tuesday, March 3, 2015 **Due:** Tuesday, March 10, 2015

Problem 4.1

(a) Let $p_{\mathsf{y}}(y; \mathbf{x})$ be a member of the exponential family. Put $\mathsf{z} = \mathsf{y} + a$, where a is a known constant. Is $p_{\mathsf{z}}(z; \mathbf{x})$ in the exponential family, as well?

(b) Let y_1 and y_2 be two independent identically distributed continuous random variables with a distribution $p_{\mathsf{y}}(y; x)$ that is in the canonical (one-parameter) exponential family. That is, $p_{\mathsf{y}}(y; x) = \exp(xy - \alpha(x) + \beta(y))$ for some functions α and β. Put $\mathsf{z} = \mathsf{y}_1 + \mathsf{y}_2$.

Show that $p_{\mathsf{z}}(z; x)$ is also in the canonical exponential family. That is, show that $p_{\mathsf{z}}(z; x) = \exp(xz - \widetilde{\alpha}(x) + \widetilde{\beta}(z))$ for some functions $\widetilde{\alpha}$ and $\widetilde{\beta}$. Express $\widetilde{\alpha}$ in terms of α, and $\widetilde{\beta}$ in terms of β.

This result means that the canonical exponential family possesses a kind of stability characteristic.

Problem 4.2

(a) Let u be a random variable whose density $p_{\mathsf{u}}(\mathbf{u}; \mathbf{a})$ is parameterized by $\mathbf{a}$ and is in the exponential family. Let v be a random variable whose density $p_{\mathsf{v}}(\mathbf{v}; \mathbf{b})$ is parameterized by $\mathbf{b}$ and is in the exponential family. Let $\mathsf{y} = (\mathsf{u}, \mathsf{v})$. Show that if u and v are independent, then $p_{\mathsf{y}}(\mathbf{y}; \mathbf{a}, \mathbf{b})$ is also in the exponential family.

(b) Let z be a random variable whose density $p_{\mathsf{z}}(z; x)$ is in the exponential family. Let y be another random variable that we observe in an attempt to infer something about z. Assume that the observation model $p_{\mathsf{y}|\mathsf{z}}$ does not depend on x. Suppose we observe $\mathsf{y} = y_0$. Is it true that the conditional density $p_{\mathsf{z}|\mathsf{y}}(z \mid y_0; x)$ is in the exponential family?

Problem 4.3

Consider a member of a canonical exponential family $p_{\mathsf{y}}(y; x) = e^{xy - \alpha(x) + \beta(y)}$ with a continuous parameter x.

(a) Express Fisher information $J_{\mathsf{y}}(x)$ in terms of $\alpha(\cdot)$ and $\beta(\cdot)$.

(b) Show that the maximum likelihood estimate of x based on observation y, $\hat{x}_{\mathrm{ML}}(y)$, must satisfy

$$y = F(\hat{x}_{\mathrm{ML}}(y)),$$

and find F in terms of $\alpha(\cdot)$ and $\beta(\cdot)$.

(c) Consider a binary hypothesis test

$$
\begin{aligned}
H_0 &: \quad p_{\mathsf{y}|\mathsf{H}}(y|H_0) = e^{\beta(y)} \\
H_1 &: \quad p_{\mathsf{y}|\mathsf{H}}(y|H_1) = e^{xy-\alpha(x)+\beta(y)} \quad \text{for a known} \quad x > 0,
\end{aligned}
$$

where the two hypotheses are equally probable a priori. Show that the likelihood ratio test (LRT) that minimizes the probability of error can be expressed in the form

$$A(x)\, y + B(x) \underset{\hat{H}(y)=H_0}{\overset{\hat{H}(y)=H_1}{\gtrless}} 0$$

and find $A(x)$ and $B(x)$.

(d) (**practice**) Now let's assume we do not know the true value of x for H_1. Instead, we replace x in the LRT of part (c) with the ML estimate $\hat{x}_{\mathrm{ML}}(y)$ to obtain a new decision rule. For this question, assume that the estimate $\hat{x}_{\mathrm{ML}}(y)$ is sufficiently close to the true value x, but $\hat{x}_{\mathrm{ML}}(y) \neq x$. In addition, denote $\mathcal{Y}$ as the alphabet of the values that y can take, and assume that its cardinality satisfies $|\mathcal{Y}| \geq 2$. Our goal is to show that this new decision rule favors hypothesis H_1 more than the true likelihood ratio test.

(i) Show that

$$[A(\hat{x}_{\mathrm{ML}}(y))\, y + B(\hat{x}_{\mathrm{ML}}(y))] - [A(x)\, y + B(x)] = \gamma(\hat{x}_{\mathrm{ML}}(y) - x)^2 + o((\hat{x}_{\mathrm{ML}}(y) - x)^2),$$

where γ is a constant, independent of x and y. $o(z)$ denotes terms that decrease faster than z as $z \to 0$. Find γ in terms of $\alpha(\cdot)$, $\beta(\cdot)$, and $\hat{x}_{\mathrm{ML}}$.

(ii) Prove that $\gamma > 0$.

(iii) Let the decision regions for H_1 corresponding to using x and $\hat{x}_{\mathrm{ML}}(y)$ be $\mathcal{Y}_1^x$ and $\mathcal{Y}_1^{\hat{x}_{\mathrm{ML}}}$, respectively. Show that for $\hat{x}_{\mathrm{ML}}(y)$ sufficiently close to x, $\mathcal{Y}_1^x \subseteq \mathcal{Y}_1^{\hat{x}_{\mathrm{ML}}}$.

Problem 4.4 (practice)

In this problem, let's recall Problem 3.4 in Problem Set 3 and associate it with exponential families. Let y be an exponentially distributed random variable with parameter x, where x is in turn an exponentially distributed random variable with parameter μ.

In Problem 3.4, we have obtained that the distribution of y is

$$p_{\mathsf{y}}(y;\mu) = \frac{\mu}{(\mu+y)^2}, \quad \text{for } y > 0.$$

In addition, we have established that the maximum likelihood (ML) estimator of μ given observation $\mathsf{y} = y$ is not unbiased and thus not efficient.

In this problem, determine whether $p_{\mathsf{y}}(y;\mu)$ is a member of an exponential family. If it is, then find the parameters of the family. If not, explain.

Problem 4.5 (practice)
Let x be a Bernoulli random variable with parameter p, i.e.,

$$p_{\mathsf{x}}(x;p) = p^x(1-p)^{1-x}.$$

Let y be a scalar random variable whose distribution is in an exponential family $\mathtt{E}(\lambda, t(\cdot), \beta(\cdot))$:

$$p_{\mathsf{y}|\mathsf{x}}(y|x;\lambda_0,\lambda_1) = \exp\left\{\lambda_x t(y) - \alpha(\lambda_x) + \beta(y)\right\} \quad \text{for } x = 0, 1.$$

(a) Let $\mathbf{z} = (\mathsf{y}, \mathsf{x})$ be the random vector obtained by combining the observation y and the binary label x. Show that $p_{\mathbf{z}}(\mathbf{z};\lambda_0,\lambda_1,p)$ is a member of a 3-parameter exponential family, i.e.,

$$p_{\mathbf{z}}(\mathbf{z} = (y,x);\lambda_0,\lambda_1,p) = \exp\left\{\sum_{i=1}^{3} \eta_i(\lambda_0,\lambda_1,p)u_i(y,x) - \alpha_{\mathbf{z}}(\lambda_0,\lambda_1,p) + \beta_{\mathbf{z}}(y,x)\right\},$$

and determine the natural parameters $\eta_1(\cdot)$, $\eta_2(\cdot)$, $\eta_3(\cdot)$, the natural statistics $u_1(\cdot)$, $u_2(\cdot)$, $u_3(\cdot)$, the log base distribution $\beta_{\mathbf{z}}(\cdot)$ and the log partition function $\alpha_{\mathbf{z}}(\cdot)$. Your answer may depend on $t(\cdot)$, $\beta(\cdot)$, λ_0, λ_1, and p.

Hint: The likelihood of y can also be written as

$$p_{\mathsf{y}|\mathsf{x}}(y|x;\lambda_0,\lambda_1) = \left[p_{\mathsf{y}|\mathsf{x}}(y|1;\lambda_0,\lambda_1)\right]^x \left[p_{\mathsf{y}|\mathsf{x}}(y|0;\lambda_0,\lambda_1)\right]^{1-x}.$$

(b) Let $\mathbf{z}^{(N)} = \{(x_1,y_1),\ldots,(x_N,y_N)\}$ be a sequence of N independent, identically distributed samples generated from the distribution $p_{\mathbf{z}}(\mathbf{z};\lambda_0,\lambda_1,p)$. Let $\hat{\lambda}_{0_{\mathrm{ML}}}$ be the maximum likelihood estimate of the parameter λ_0 from the observations $\mathbf{z}^{(N)}$. Let $\hat{q}$ be the corresponding member of the exponential family, i.e., $\hat{q}(y) = \exp\left\{\hat{\lambda}_{0_{\mathrm{ML}}} t(y) - \alpha(\hat{\lambda}_{0_{\mathrm{ML}}}) + \beta(y)\right\}$. Show that

$$\mathbb{E}_{\hat{q}}\left[t(\mathsf{y})\right] = C \sum_{i \in \mathcal{S}} t(y_i),$$

where $\mathbb{E}_{p(\cdot)}\left[\mathsf{w}\right]$ denotes the expectation of a random variable w computed with respect to a probability distribution $p(\cdot)$, C is a constant, and $\mathcal{S}$ is a set of indices, $\mathcal{S} \subseteq \{1,\ldots,N\}$. Determine the constant C and the set $\mathcal{S}$ in terms of observations $\mathbf{z}^{(N)} = \{(x_1,y_1),\ldots,(x_N,y_N)\}$.

This result implies that for exponential families, performing maximum likelihood estimation can be viewed as matching the *expected value* of the natural statistic to a particular function of the *observed values* of natural statistic.

(c) For most models, the marginal distribution $p_{\mathsf{y}}(\cdot;\lambda_0,\lambda_1,p)$ of the random variable y cannot be represented as an exponential family, but showing it is quite challenging. Here we consider a limited example.

Let the distribution of y conditioned on $\mathsf{x}=x$ be a unit-variance Gaussian distribution with mean λ_x, i.e.,

$$p_{\mathsf{y}|\mathsf{x}}(y|x;\lambda_0,\lambda_1)=\frac{1}{\sqrt{2\pi}}e^{-\frac{(y-\lambda_x)^2}{2}}\qquad\text{for } x=0,1,$$

and, further, let $p=1/2$ and $\lambda_1=-\lambda_0$. You are to prove by contradiction that $p_{\mathsf{y}}(\cdot;\lambda_0,-\lambda_0,1/2)$ is not a full 1-parameter exponential family of distributions. In particular,

i) First, suppose that $p_{\mathsf{y}}(\cdot;\lambda_0,-\lambda_0,1/2)$ is a full 1-parameter exponential family $\mathtt{E}(\theta(\cdot),\tilde{t}(\cdot),\tilde{\beta}(\cdot))$, i.e., for all λ_0,

$$p_{\mathsf{y}}(y;\lambda_0,-\lambda_0,\tfrac{1}{2})=\exp\left\{\theta(\lambda_0)\tilde{t}(y)-\tilde{\alpha}(\lambda_0)+\tilde{\beta}(y)\right\}.$$

Show that this implies that for all λ_0 and all y

$$\theta'(\lambda_0)\,\tilde{t}(y)-\tilde{\alpha}'(\lambda_0)+\lambda_0=yF(y\lambda_0).$$

for some function $F(\cdot)$. Determine $F(\cdot)$.

ii) Use the result of part i) above to prove that $p_{\mathsf{y}}(\cdot;\lambda_0,-\lambda_0,1/2)$ cannot be a full 1-parameter exponential family.

Problem 4.6

Suppose that y_1 and y_2 are independent random variables each uniformly distributed between x and $x+1$. Let $\mathsf{s}=\max(\mathsf{y}_1,\mathsf{y}_2)$ and $\mathsf{r}=\mathsf{y}_1-\mathsf{y}_2$.

(a) Show that s is not a sufficient statistic for $p_{\mathsf{y}_1,\mathsf{y}_2}(y_1,y_2;x)$.

(b) An ancillary statistic is one whose distribution does not depend on the parameters of the model. Show that r is an ancillary statistic for $p_{\mathsf{y}_1,\mathsf{y}_2}(y_1,y_2;x)$.

(c) Is $\mathbf{u}=\begin{bmatrix}\mathsf{s}\\ \mathsf{r}\end{bmatrix}$ a sufficient statistic for $p_{\mathsf{y}_1,\mathsf{y}_2}(y_1,y_2;x)$?

Now consider a general model $p_{\mathsf{y}}(y;x)$.

(d) Suppose $\mathsf{t}=t(\mathsf{y})$ is a complete sufficient statistic for $p_{\mathsf{y}}(y;x)$, and that $\mathsf{r}=r(\mathsf{y})$ is an ancillary statistic for $p_{\mathsf{y}}(y;x)$. Show that t and r are independent. Clearly indicate where you use each of the facts that 1) t is sufficient; 2) t is complete; and 3) r is ancillary.

Hint: Consider an arbitrary function g, and let

$$\mu(x) = \mathbb{E}[g(\mathsf{r})] = \int g(r(y)) p_{\mathsf{y}}(y; x)\, \mathrm{d}y.$$

Moreover, let $\phi(\mathsf{t}; x) = \mathbb{E}[g(\mathsf{r}) - \mu(x) \mid \mathsf{t}]$.

Now we consider a different model. Instead of uniformly distributed over the interval $[x, x+1]$, the independent random variables y_1 and y_2 in the new model are each uniformly distributed on the interval $[0, x]$ where $x \geq 0$. In addition, $\mathsf{s} = \max(\mathsf{y}_1, \mathsf{y}_2)$.

(e) Determine whether s is a sufficient statistic for $p_{\mathsf{y}_1,\mathsf{y}_2}(y_1, y_2; x)$ in the new model.

Problem 4.7

Let $\mathbf{y} = [\mathsf{y}_1\ \mathsf{y}_2]^{\mathrm{T}}$ be a vector random variable whose components are i.i.d. Bernoulli random variables with parameter x, $0 < x < 1$, i.e., $\mathbb{P}(\mathsf{y}_i = 1) = x$, $i = 1, 2$.

(a) Show that $t(\mathbf{y}) = y_1 + y_2$ is a sufficient statistic.

(b) Let $\hat{x}(\mathbf{y}) = y_1$ be an estimator of the parameter x from the observation $\mathbf{y}$. Find $\mathrm{MSE}_{\hat{x}}(x)$, the mean-square error of this estimator.

(c) Let $\hat{x}'(t) = \mathbb{E}[\hat{x}(\mathbf{y}) | \mathsf{t} = t]$ be an estimator of the parameter x that uses the sufficient statistic t instead of the observations $\mathbf{y}$.

  (i) Show that $\hat{x}'(t)$ is a valid estimator, i.e., it does not depend on x.

  (ii) Show that $\mathrm{MSE}_{\hat{x}'}(x) = \gamma\, \mathrm{MSE}_{\hat{x}}(x)$ and find the constant γ.

(d) We now consider a generalization of this problem. Let $\mathbf{y}$ be a random variable generated by a distribution $p_{\mathbf{y}}(\cdot; x)$ and $\mathbf{t}(\mathbf{y})$ be a sufficient statistic. Let $\hat{x}(\mathbf{y})$ be an estimator of the parameter x based on the observation $\mathbf{y}$. We define an alternate estimator $\hat{x}'(\mathbf{t}) = \mathbb{E}[\hat{x}(\mathbf{y}) | \mathbf{t} = \mathbf{t}]$.

  (i) Show that $\hat{x}'(\mathbf{t})$ is a valid estimator, i.e., it does not depend on x.

  (ii) Show that for any cost function $C(x, \hat{x})$ that is convex in $\hat{x}$, the following inequality holds:

$$\mathbb{E}[C(x, \hat{x}'(\mathbf{t}))] \leq \mathbb{E}[C(x, \hat{x}(\mathbf{y}))].$$

  Hint: You may find Jensen's inequality useful: If $\phi(\cdot)$ is a convex function and v is a random variable, then

$$\mathbb{E}[\phi(\mathsf{v})] \geq \phi(\mathbb{E}[\mathsf{v}]).$$

Problem 4.8
Prove that in binary hypothesis testing, the likelihood ratio is a sufficient statistic.

Problem 4.9 (practice)

Determine the mildest conditions you can think of under which the natural statistics for the general k-parameter exponential family are minimal sufficient statistics.
Hint: Consider the conditions on $\boldsymbol{\lambda}$ and $\mathbf{t}$ for complete sufficient statistics (and think Laplace transforms).

Problem 4.10 (practice)
Let x be a deterministic unknown parameter, and denote y as a random variable with distribution $p_{\mathsf{y}}(y;x)$. Let $\mathsf{t}(y)$ be a sufficient statistic for x given y. Determine whether the following statement is true or false:

$J_{\mathsf{y}}(x) = J_{\mathsf{t}}(x)$, where $J_{\mathsf{y}}(x)$ and $J_{\mathsf{t}}(x)$ are the corresponding Fisher information.

Prove if it is true and give a counter example if it is not necessarily true.