Problem Set III. Maximum Likelihood

Probability and Beyond
January 26, 2017

Due: January 27, 11.59 pm

Directions: It is strongly recommended to solve the problems on your own. If you end up consulting outside sources, cite your sources. Feel free to work with others, but the final write-up should be entirely your own and based on your own understanding.

Problem 3.1. Write down the density function and characteristic function of the normal $\mathcal{N}(\mu, \sigma^2)$. Compute its first four moments using
i) density function
ii) characteristic function.
Problem 3.2. Exercise 3.1 from lecture notes.
Problem 3.3. A probability distribution or density on a set $\mathcal{D}$, parametrized by $\theta \in \mathbb{R}^n$, is called an *exponential family* if it has the form

$$p_\theta(x) = a(\theta) \exp\{\theta^T c(x)\}, \tag{1}$$

for $x \in \mathcal{D}$, where $c : \mathcal{D} \to \mathbb{R}^n$, and $a(\theta)$ is a normalizing function. Here we interpret $p_\theta(x)$ as a density function when $\mathcal{D}$ is a continuous set, and a probability distribution when $\mathcal{D}$ is discrete. Thus we have

$$a(\theta) = \left(\int_{\mathcal{D}} \exp\{\theta^T c(x)\}\, dx \right)^{-1} \tag{2}$$

when p_θ is a density, and

$$a(\theta) = \left(\sum_{x \in \mathcal{D}} \exp\{\theta^T c(x)\} \right)^{-1} \tag{3}$$

when p_θ represents a distribution. We consider only values of θ for which the integral or sum above is finite, i.e. $< +\infty$. Many families of distributions have this form, for appropriate choice of parameter θ and function $c(\cdot)$.

(a) When $c(x) = x$ and $\mathcal{D} = \mathbb{R}^n_+$, what is the associated family of densities? What is the set of valid values of θ?

(b) Consider the case with $\mathcal{D} = \{0, 1\}$, with $c(0) = c(1) = 1$. What is the associated exponential family of distributions? What are the valid values of $\theta \in \mathbb{R}$?

(c) Show that for any $x \in \mathcal{D}$, the log-likelihood function $\log p_\theta(x)$ is concave in θ. This means that maximum-likelihood estimation for an exponential family leads to a convex optimization problem. It is OK to prove it for finite set $\mathcal{D}$ and explain how it can be extended for discrete but infinite $\mathcal{D}$ and continuous $\mathcal{D}$.

Problem 3.4. Let $\mathbf{X} = \{x_1, \ldots, x_n\}$, where x_i sampled from Poiss(λ). Find the estimator of parameter λ as i) an expected value of posterior distribution $p(\lambda \,|\, \mathbf{X})$; ii) maximum a posteriori. Use the Gamma distribution as prior distribution on λ, i.e.

$$p(\lambda) = \mathcal{G}(\lambda \,|\, \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} \exp(-\beta\lambda), \tag{4}$$

where $\alpha, \beta > 0$ are considered to be given. Is it the same as MLE estimator?

Problem 3.5. Prove that Normal, Bernoulli and Poisson distributions belong to exponential family. Give at least one example of a well known distribution which is not from exponential family.

Problem 3.6. Estimate the probability of getting a tail p of an unfair coin (maximizing the posterior probability) given n experiments using a prior for p, $p \sim \mathcal{B}\mathbf{eta}(\alpha, \beta)$ with density

$$p_{\alpha,\beta}(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)} \quad B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}, \tag{5}$$

which implies $p_{\alpha,\beta}(x) \propto x^{\alpha-1}(1-x)^{\beta-1}$. Comment on what you obtain.