Statistical Inference and Learning- Ex-2

Yiqing Wang and Boaz Nadler

due by Wed. Dec. 16, 2015

Q1. You are given n independent samples from a one dimensional Gaussian distribution $x_1, \ldots, x_n \sim N(\theta, \sigma^2)$ whose mean θ is unknown but is assumed to come from a distribution $\theta \sim N(\mu_0, \sigma_0^2)$.

1. Find the ML and MAP estimates for θ.
2. Find the bias, variance and mean squared error of these estimators.
3. What is the Cramer-Rao lower bound on the mean squared error? Is it achieved by the above estimators? If a gap between the bound and the MSE of an estimator exists, explain why.

Q2 In many practical scenarios the goal is to estimate a random quantity x, but unfortunately we can only observe it with some additive noise y (which may not necessarily have zero mean !). Consider such a setting: Let x, y_1 be two independent Gaussian random variables with distributions $x \sim N(\mu, \sigma^2)$, $y_1 \sim N(\mu_1, \sigma_1^2)$, where $\mu, \mu_1, \sigma, \sigma_1$ are known. Suppose we observe $z = x + y_1$.

1. What is the distribution of z ?

2. Find $\mathbb{E}[x|z]$.

3. Suppose that given z, your goal is to construct an estimate $\hat{x} = \hat{x}(z)$ of x, with minimal mean squared error (MSE) $\mathbb{E}[(\hat{x} - x)^2]$. Find the optimal such estimate under the family of linear transformations, $\hat{x}(z) = a + bz$. What are the optimal a, b ? Is $\mathbb{E}[\hat{x}(z)|x] = x$? What happens when $\sigma_1 \to \infty$, can you explain this ?

4. In some cases one can measure the same quantity with different measurement devices, hence leading to multiple (noisy) measurements of the same quantity. Suppose here we observe both $z_1 = x + y_1$, and $z_2 = x + y_2$, where $y_1 \sim N(0, \sigma_1^2)$ and $y_2 \sim N(0, \sigma_2^2)$ are independent of each other. Assume both σ_1, σ_2 are known. How would you form a single estimate of x in this case ?

Q3

For many statistical problems, if we observe i.i.d. data x_i from some density $p_\theta(x)$ and estimate θ by the Maximum Likelihood principle, we obtain a consistent estimator that as $n \to \infty$ converges to the true value θ. However, this is not always the case !

Consider the following setting. The unknown density $p_\theta(x)$ is a mixture of two 1-D Gaussians, where one is known explicitly and the other is unknown. That is, $\theta = (\mu, \sigma)$ is of the form

$$p_{\mu,\sigma}(x) = \frac{1}{2}\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) + \frac{1}{2}\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$$

1. Let $x_1, \ldots, x_n$ denote n observed data samples. Show that the log-likelihood grows unbounded for certain values of μ, as $\sigma \to 0$. [Hint: What happens if $\mu = x_j$, one of the observed data points ?]

2. One approach to circumvent the problem above is to set $\sigma > \sigma_0$. However, if one of the Gaussians has width $\sigma < \sigma_0$ it will not be recovered even in the limit $n \to \infty$. Suppose that instead we put a sample-size dependent constraint $\sigma > \sigma(n)$. How fast can $\sigma(n)$ tend to zero while still ensuring the consistency of σ, μ as $n \to \infty$?