STAT 336: "MAJOR" HOMEWORK 2

SOLUTIONS

For problems 1-3, let $X_1, \ldots, X_n \sim \mathcal{U}(0, \theta)$, where θ is unknown.

(1) Let $\hat{\theta} = 2\bar{X}$. Find the bias, variance, and MSE for this estimator (as functions of θ).

(2) Let $\hat{\theta}$ be the MLE for θ: $\hat{\theta} = \max_{1 \le i \le n} X_i$.

  (a) Show that $\hat{\theta}$ is consistent. Recall that this entails showing that $\mathbb{P}(\left|\hat{\theta} - \theta\right| > \varepsilon) \to 0$ for an arbitrarily small ε. Hint: $\hat{\theta}$ cannot be above θ.

  (b) Show that $\mathbb{P}(\hat{\theta} = \theta) = 0$.

  (c) Suppose we simulate the sampling distribution of $\hat{\theta}$ via the nonparametric bootstrap procedure, and let $\hat{\theta}^*$ be one such simulated value of $\hat{\theta}$. Show that $\mathbb{P}(\hat{\theta}^* = \hat{\theta}) = 1 - (1 - \frac{1}{n})^n$, and hence, $\lim_{n\to\infty} \mathbb{P}(\hat{\theta}^* = \hat{\theta})$ is $1 - e^{-1} \approx 0.632$.

  (d) Set $\theta = 1$ and simulate M "oracle" samples of size 50 from the true distribution (using the software of your choice, and a large value of M, say 10000). Denote these oracle samples by $X^{(1)}, \ldots, X^{(M)}$, where $X^{(m)} = (X_1^{(m)}, \ldots, X_{50}^{(m)})$, the corresponding estimates by $\hat{\theta}^{(m)}$ Calculate the variance of the $\hat{\theta}^{(m)}$ using your simulated "oracle" samples. Then construct a bootstrap distribution for $\hat{\theta}$ from each oracle sample to calculate $\mathbb{V}[\hat{\theta}^* \,|\, X^{(m)}], m = 1, \ldots, M$ and plot these variances. Where does the variance you computed from the oracle samples themselves fall in the distribution? Do the bootstrap variance estimates tend to be near the oracle variance?

(3) Suppose we place a continuous prior on θ which is supported on $[0, \infty)$. What is the support of the posterior (in terms of $X_1, \ldots, X_n$)?

(4) Let $X_1, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$. Let τ be the 0.95 quantile of the unknown Normal distribution. Find the MLE and an expression for an approximate 95% confidence interval for τ. Hint: represent τ as a function of μ and σ^2, and use the fact that $\bar{X}$ and S^2 are independent statistics (which you can verify by factoring the likelihood function).

(5) Suppose $X_1, \dots, X_n \sim$ Exponential(θ); i.e.

$$f(x_i \,|\, \theta) = \theta e^{-\theta x_i}$$

Find a sufficient statistic for θ.

(6) Consider the family of categorical distributions supported on $\{a_1, \dots, a_D\}$, with density given by

$$f(x \,|\, \theta) = \prod_{d=1}^{D} \theta_d^{I(x=a_d)}$$

where $\theta = (\theta_1, \dots, \theta_D)$ is the parameter vector satisfying $\sum_d \theta_d = 1$, the $\{a_d\}$ are known, and $I(A)$ is the indicator function equal to 1 when the condition A holds, and 0 otherwise.

(a) Suppose $X_1, \dots, X_n \sim f(x \,|\, \theta)$. Show that the family of joint densities is an exponential family, and find the natural parameters and corresponding sufficient statistics.

(b) Find the MLE for θ given the sample $X_1, \dots, X_n$.

(c) Find a conjugate prior for θ up to a normalizing constant and show that the posterior parameter updates use only the sufficient statistics you found in (b).

(7) Let $f(\theta)$ be an arbitrary prior distribution. Show that $f(\theta \,|\, x) = f(\theta)$ as functions (i.e., for every θ) if and only if $\mathcal{L}(\theta; x) = h(x)$, for some $h(x)$ that does not depend on θ.

(8) Let θ be a scalar parameter and $f(\theta \,|\, x)$ be a posterior distribution. Suppose we wish to choose $\hat{\theta}$ so as to minimize $\mathbb{E}[(\hat{\theta} - \theta)^2 \,|\, X]$, where the expectation is with respect to $f(\theta \,|\, x)$. Show that the estimator

$$\hat{\theta} = \mathbb{E}[\theta \,|\, X]$$

is optimal in this sense.

(9) Show that a statistic $T(X)$ is sufficient for θ if and only if for every prior $f(\theta)$ we have $f(\theta \,|\, T(x)) = f(\theta \,|\, x)$. (Recall that we defined $T(X)$ to be a sufficient statistic for θ if $\mathcal{L}(\theta \,|\, X) = h(X) g(\theta, T(X))$, where $h(X)$ does not depend on θ, and g depends on X only through $T(X)$.)