Data Science HW1

2021.10.07

Submission

- Deadline: 10/19 Tue. 23:59
  - Submission delay: will get no points
- Upload: Ceiba homework section
- File format: PDF
  - Format error: -10%

Problem 1 Random Number Transformation

Min() and Max() appears frequently in applications of data science.

Let X and Y be two independent random variables with identical probability density function given by

$$f(x) = \begin{cases} e^{-x} & for\ x > 0 \\ 0 & elsewhere. \end{cases}$$

(1) What is the probability density function of Z=max(X,Y) ? (10%)
(2) What is the probability density function of W=min(X,Y) ? (10%)

(Hint: max(X, Y) <= a, mean “X<=a and Y <= a”.)

Problem 2 Statistical Distances

(1) Let f() be a function mapping from X × X to non-negative real numbers. What are the four conditions that f must satisfy for f() to be considered a metric? (5%)
(2) Let x, y be two points (two vectors) in space. The function d(x, y) = ||x-y||^2 is called the squared Euclidean distance. Prove that the squared Euclidean distance is a Bregman divergence (10%) (hint: what is the F() in this Bregman divergence?)
(3) You are given an artificial neural network. The network implements a function that takes an input x and produce an output y. That is, it implements y = F(x). Prove that the entropy of the output of this neural network will always be equal or less that of the input. (10%)

Problem 3 Point Estimation

We have a population X whose distribution is uniform over the interval $(0,\theta)$. The prior distribution of θ is uniform over the interval $(0,1)$. Please derive the estimator of θ based on a sample of size $n >= 2$, using:

(1) The moment method (5%)
(2) The MAP method (10%)
(3) The bayesian method using squared error loss function (10%)

Problem 4 Goodness of Estimation

Let $\theta 1$ and $\theta 2$ be two unbiased estimators of θ, with $Var(\theta 1)=1$, $Var(\theta 2)=2$ and $Cov(\theta 1,\theta 2)=1/4$.

Let $\theta 3$ and $\theta 4$ be two unbiased estimators of θ, with $Var(\theta 3)=1$, $Var(\theta 4)=2$ and $Cov(\theta 3,\theta 4)=3/4$.

(1) What is the unbiased estimator with the lowest variance that you can construct from a linear combination of $\theta 1$ and $\theta 2$, and what's its variance? (5%)

(2) Also answer the same question for $\theta 3$ and $\theta 4$ (5%)

(3) Observe which combination can produce a new estimator with lower variance (for your own pleasure, no additional points awarded)

Problem 4 Interval Estimation

Let X1, X2, ..., Xn be a random sample from a population with density function

$$f(x;\theta) = e^{-(x-\theta)},\ if\, \theta < x < \infty$$

where θ∈R is an unknown parameter.

(1) Show that Q = X(1) - θ is a pivotal quantity. (10%).

(2) Use this pivotal quantity find a 100(1-α)% confidence interval for θ. (10%)