Math 106 Data-driven Uncertainty Quantification
Homework 02

Winter 2021

Instructor: Yoonsang Lee (Yoonsang.Lee@dartmouth.edu)

Due Feb 10, 2021 11:59 pm (EDT)

Starred problems are for math-major students. Non-math major students are welcome to try starred problems.

1. Let $X_1, X_2, ..., X_n$ is IID Uniform$(0, \theta)$. (a) Find the estimator using the method of moments, and check whether it is based or not. (b) Find the MLE of θ. Also check whether the MLE is biased or not.

2. A spectator saw racing cars numbered 15, 45, and 38. Assuming that cars are numbered sequentially, what is an unbiased estimate of the total number of cars in the racing? Justify your answer.

3. Derive the Nadaraya-Watson nonparametric regression.

4. Let $p_1(u) = \phi(u; u_1, \sigma_1^2)$ and $p_2(u) = \phi(u; u_2; \sigma_2^2)$ where $\phi(u; m, \sigma^2)$ is the normal distribution with mean m and variance σ^2. Show that $p(u) = cp_1(u)p_2(u)$ where c is a normalization constant is also normal. Find the mean and variance, say u_3 and σ_3^2, of $p(u)$.

5. Use the previous problem to show the following Kalman update formula

$$u_{posterior} = u_{prior} + K(v - u_{prior})$$

$$\sigma^2_{posterior} = (1 - K)\sigma^2_{prior}$$

where K is the Kalman gain

$$K = \frac{\sigma^2_{prior}}{\sigma^2_{prior} + \sigma^2_{obs}}$$

Specify the correspondence between $u_1, u_2, u_3, \sigma_1^2, \sigma_2^2, \sigma_3^2$ and $u_{posterior}, u_{prior}, v, \sigma^2_{posterior}, \sigma^2_{prior}, \sigma^2_{obs}$.

6. This problem is about the density estimation in 2D. Let X and Y follow the normal distributions with mean $(0, 0)$ and $(-5, 5)$ and covariance matrices $\begin{pmatrix} 5 & .5 \\ .5 & 1 \end{pmatrix}$ and $\begin{pmatrix} 5 & 2 \\ 2 & 1 \end{pmatrix}$ respectively. Let $Z = X + Y$. Draw a sample of size $n = 1000$ from Z and use KDE (using the Gaussian kernel) to estimate the density. Compute the relatively entropy. Use the relative entropy to find the optimal bandwidth. Can you improve your estimation using a matrix as a bandwidth (so that you can effectively has different bandwidth in different directions)?

7. - Let X be a random variable with a density $\frac{1}{3}\phi(x;0,1)+\frac{2}{3}\phi(x;1,4)$ where $\phi(x;m,\sigma^2)$ is a Gaussian density with mean m and variance σ^2.
   - Generate an IID sample of size $n = 10, 100, 500, 1000, 10000$, and 100000. Use seed = 0.
   - From the sample, $\{X_i\}$, estimate the density using (i) histogram, and (ii) KDE.
   - Compute the relative entropy using the estimated densities.
   - Plot the relative entropy as a function of the sample size n.

8. (cont'd)
   - Let $Y = X^2$ where the density of X is given as above. Find the density of Y numerically (like a histogram or KDE).
   - Estimate the mean of X, say $\hat{m}$.
   - Calculate the variance of $\hat{m}$ analytically.
   - Estimate the variance of $\hat{m}$ using the bootstrap.

9. * Prove that the MLE is asymptotically normal (Fill out the gap of the proof in the lecture slides).

10. * Prove the Glivenko-Cantelli lemma.