STAT 830

Problems: Assignment 3

1. I have talked about non-parametric estimation of $\mathrm{E}(Y|X = x)$ in the videos. In this problem I want you to work on the somewhat simpler case of density estimation. The point of the problem is to discover roughly how wide the histogram bins should be to make the MSE as small as possible. Imagine we want to estimate the density of the heights of the "Mothers" in Pearson-Lee data. We might use

$$\hat{f}(x,\epsilon) = \frac{\sum_{i=1}^n 1(x-\epsilon \le X_i \le x+\epsilon)}{2n\epsilon}$$

where the data, the heights of the mothers, are $X_1, \ldots, X_n$.

   (a) Give a formula for the mean and variance of this estimator. The answer will depend on n, ϵ, the true cdf F and its density f.

   (b) If f is the $N(\mu, \sigma^2)$ density with say $\mu = 65$ and $\sigma = 2.5$ plot the bias of this estimator over the range 58 to 72 for $\epsilon = 1$ and for $\epsilon = 0.1$.

   (c) In the rest of this question stick to the example in part b. Imagine that you want to get approximate behaviour of this estimator by choosing a sequence $\epsilon_n = n^{-\alpha}$. Use Maple or similar program to compute a Taylor expansion of the mean in part a in terms of ϵ. Use this expansion to determine all values of α for which the bias converges to 0.

   (d) Simlarly use a Taylor expansion of the variance and discover those α for which the variance converges to 0.

   (e) Are there any α for which both converge to 0?

   (f) Is there a value of α for which the bias divided by the standard deviation has a limit? If so what is that limit and what is that α?

2. Suppose $\{X_{ij}; j = 1, \ldots, n_i; i = 1, \ldots, p\}$ are independent $N(\mu_i, \sigma^2)$ random variables. (This is the usual set-up for the one-way layout.)

   (a) Find the MLE's for μ_i, σ, and σ^2 .

(b) Find the expectations and variances of the estimators of μ_i and σ^2. You can use the facts about T_i given in the next question without proof and you can use the formulas for means and variances of chi-squared distributions without proof.

3. Let T_i be the error sum of squares in the ith cell in the previous question. In this question you may assume that

$$\frac{T_i}{\sigma^2} = \frac{\sum_{j=1}^{n_i}(X_{ij} - \bar{X}_{i\cdot})^2}{\sigma^2}$$

has a χ^2 distribution with $n_i - 1$ degrees of freedom. The notation is that $\bar{X}_{i\cdot} = \sum_j X_{ij}/n_i$ is the mean in cell i.

(a) Find the best estimate of σ^2 of the form $\sum_1^p a_i T_i$ in the sense of mean squared error. That is, find the constants a_i which minimize the mean squared error.

(b) Do the same under the condition that the estimator must be unbiased.

(c) Find the joint density of the T_i.

(d) If only $T_1, \ldots T_p$ are observed what is the MLE of σ? It is a fact which you may assume that the T_i are independent.

4. In question 2 take $n_i = 2$ for all i and let $p \to \infty$. What is the limit of the MLE of σ? Hint: use the mean and variance of the mle of σ^2; you found these two problems ago. Use Chebyshev's inequality or the law of large numbers; if you do the latter do it carefully.

5. Suppose that $Y_1, \ldots, Y_n$ are independent random variables and that $x_1, \ldots, x_n$ are the corresponding values of some covariate. Suppose that the density of Y_i is

$$f(y_i) = \exp\left(-y_i \exp(-\alpha - \beta x_i) - \alpha - \beta x_i\right) 1(y_i > 0)$$

where α, and β are unknown parameters.

(a) Find the log-likelihood, the score function and the Fisher information.

(b) For the following data set fit the model and produce a contour plot of the log-likelihood surface, the profile likelihood for β and an approximate 95% confidence interval for β.

x_i	Y_i	x_i	Y_i
-0.59	0.694	0.31	0.009
0.28	0.408	-0.45	0.032
-1.08	1.330	1.23	0.057
0.79	0.266	-1.71	1.530
-0.15	0.111	0.43	0.152
0.50	0.382	-0.61	0.358
0.54	0.165	0.60	0.214
-1.05	2.766	0.68	0.375
0.66	0.103	0.92	0.193
0.40	0.309	1.50	0.245

Due October 27, 2020.