

ESE 524 - Homework 2

Assigned date: 02/5/19

Due date: 02/19/19

Total Points: 100 + 20 E.C.

1) MLE and CRB

Suppose we have N i.i.d. samples X_1, X_2, \dots, X_N , drawn from a Poisson Distribution, i.e.,

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad x = 0, 1, 2, \dots$$

where the rate λ is the parameter that we want to estimate.

- (5 pts)** Find the Maximum likelihood estimator (MLE) of λ . In addition, check whether the MLE is unbiased or not.
- (5 pts)** Compute the Cramer-Rao lower bound for the estimation of λ .
- (10 pts)** Check if the MLE is efficient.

2) Fisher Information

As learned from the course and slides, we know that under some regularity conditions, an unbiased estimator $T(\mathbf{x})$ is efficient for the parameter θ if and only if

$$I(\theta)[T(\mathbf{x}) - \theta] = \frac{\partial \log p(\mathbf{x}; \theta)}{\partial \theta},$$

where $I(\theta)$ is the Fisher information.

- (10 pts)** Use the definition of Fisher information to prove that the Fisher information $I(\theta)$ is non-negative.
- (10 pts)** Suppose that Fisher information is positive, and the maximum likelihood estimator can be found by $\frac{\partial \log p(\mathbf{x}; \theta)}{\partial \theta} = 0$, then prove that if an unbiased efficient estimator $T(\mathbf{x})$ exists, then it must be the MLE.

3) Exponential Family

Given a probability measure η (see the note below), the exponential family is defined to have the following general form:

$$p(\mathbf{x}|\eta) = h(\mathbf{x}) \exp\{\eta^T T(\mathbf{x}) - A(\eta)\}$$

for a parameter vector η , often referred to as the canonical parameter. The function $A(\eta)$ is known as the cumulant function.

- (10 pts)** Show that with respect to the measure ν ,

$$A(\eta) = \log \int h(\mathbf{x}) \exp\{\eta^T T(\mathbf{x})\} \nu(d\mathbf{x})$$

Note: A probability measure is a real-valued function defined on a set of events in a probability space that satisfies measure properties such as countable additivity. For this particular problem, you may just need the fact about ν :

$$\int p \nu(d\mathbf{x}) = 1$$

where p is the corresponding probability density.

- (10 pts)** Given a distribution in the exponential family:

$$p(\mathbf{x}; \eta) = h(\mathbf{x}) \exp [\eta^T T(\mathbf{x}) - A(\eta)],$$

Prove that

$$\mathbb{E}_{p(\mathbf{x}; \eta)} [T(\mathbf{X})] = \frac{\partial A(\eta)}{\partial \eta}, \quad \text{var}_{p(\mathbf{x}; \eta)} [T(\mathbf{X})] = \frac{\partial^2 A(\eta)}{\partial \eta \partial \eta^T}.$$

- 4) (Slides, I3.pdf) **Cramer-Rao Bound:** The measurements $\mathbf{X} = [X_1, X_2, \dots, X_n]^T$ are independent and identically distributed as Gaussian random variables with unknown mean μ and variance σ^2 .
- (a) (10 pts) Show that the sample mean $\bar{X} = \frac{1}{n} \sum_{k=1}^n X_k$ and sample variance $S^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X})^2$ are unbiased estimators and that they are uncorrelated and independent random variables. [Hint: Show that Gaussian variables $X_i - \bar{X}$ and \bar{X} are uncorrelated for $i = 1, \dots, n$].
- (b) (5 pts) Using the results in (i), derive the covariance matrix for the estimator $\hat{\theta} = [\bar{X}, S^2]^T$.
- (c) (5 pts) Derive the Cramer-Rao bound on the covariance matrix of any unbiased estimator $\hat{\theta}$ of $\theta = [\mu, \sigma^2]^T$. Compare to the results obtained in part (ii).

- 5) **Logistic Regression (MATLAB Problem):** In cases involving binary data (e.g. heads/tails coin flip, passing/failing a test) the assumptions required for linear models do not always hold. This problem will use data from space shuttle o-ring tests to provide a short introduction to Generalized Linear Models and Logistic Regression. Download the file named “challenger.mat” from the blackboard Assignments-4, which contains the variables “Temperature” and “Failure”.

- (i) (5 pts) Let $\mathbf{Y} \in \{0, 1\}$ be a Bernoulli random variable with mean p representing the outcome of interest (in our case whether the o-ring fails), $\mathbf{x} = [x_1 \dots x_n]$ be a vector of “explanatory” variables, and $\theta = [\theta_0 \dots \theta_n]$ be a set of unknown parameters. Logistic regression defines the following relationship between the probability of the outcome and the explanatory variables:

$$\ln \frac{\Pr(\mathbf{Y} = 1 | \mathbf{x}, \theta)}{1 - \Pr(\mathbf{Y} = 1 | \mathbf{x}, \theta)} = \theta_0 + \theta_1 x_1 + \dots + \theta_n x_n$$

Derive the expression for the mean $p = \Pr(\mathbf{Y} = 1 | \mathbf{x}, \theta)$ in terms of \mathbf{x} and θ .

- (ii) (5 pts) Let \mathbf{Y}^i be independent observations of \mathbf{Y} and \mathbf{x}^i be the corresponding explanatory vectors. Derive the gradient of the log likelihood function with respect to $[\theta_0, \dots, \theta_n]$:

$$\ln(L(\theta)) = \ln(\prod_{i=1}^n p^{Y^i} (1-p)^{1-Y^i})$$

(Hint: the function $f(x) = \frac{1}{1+\exp(-x)}$ is called the logistic function and its derivative is given by $f(x)(1-f(x))$.)

- (iii) (5 pts) Maximum Likelihood is usually done numerically. In the case of the o-rings data \mathbf{Y}^i is the vector “Failure” and the samples X^i are found in the “Temperature” vector. Construct the likelihood function and its gradient in MATLAB as a function of θ_0 and θ_1 and find the optimal weights using simple gradient descent (Note: 0.0001 works as a step size for gradient descent. You will need to multiply the log-likelihood by -1 to minimize using gradient descent). Compare your values to the coefficients generated using the `glmfit` function in MATLAB with “binomial” as the distribution.
- (iv) (5 pts) Plot $(X^i, P(Y^i = 1 | \hat{\theta}_1, \hat{\theta}_0))$ using the estimators you found in part (iii). On the same plot create a scatter plot of the actual values (X^i, Y^i) . Do the failed o-rings have a high probability of failure?

- 6) (20 pts) **Extra Credit:** Come up with an example and solution illustrating one or more concepts from class so far. This example should be something you believe would be good to present in class to help other students understand a concept from the lectures. Matlab (or other software) simulations are encouraged. Problems can be inspired by or explore applications from literature, but should not just copy the results of a paper.