

Homework 6 DS-GA 1002

Yuhao Zhao
Yz3085

November 12, 2015

Problem 1.

a) Given the observation x_1, x_2, \dots, x_n , the value that u could be is $[max(x_i), \infty)$

b) The density function of a $X \sim \text{Uniform}(0, u)$ is $\frac{1}{u} \times I(x \in [0, u])$

Therefore the likelihood function given observation x_1, x_2, \dots, x_n is:

$$l(x_1, x_2, \dots, x_n | u) = \frac{1}{u^n} \prod_{i=1}^n I(x_i \in [0, u])$$

The log likelihood is :

$$L(x_1, x_2, \dots, x_n | u) = -n \log(u) + \sum_{i=1}^n I(x_i \in [0, u])$$

To maximize the log likelihood, we notice that $-n \log(u)$ decreases with larger u , and $\sum_{i=1}^n I(x_i \in [0, u])$ is maximized at n , with $u > max(x_i)$. Therefore, the Maximum of log likelihood is attained at $\hat{u}_{MLE} = max(x_i) = x_{(n)}$.

c) We need to find out the pdf of \hat{U}_{ML} that is the pdf of order statistics $x_{(n)}$.

$$F(x_{(n)}) = P(x_{(n)} < t) = P(x_1 < t, x_2 < t, \dots, x_n < t)$$

$$\text{Since } x_i \text{'s are iid. RHS} = \prod P(x_i < t) = p(x_i < t)^n = \left(\frac{t}{u}\right)^n$$

$$\text{Therefore, the pdf is } dF/dt = \frac{n}{u} \left(\frac{t}{u}\right)^{n-1} = \frac{n}{u} \left(\frac{\hat{u}_{ML}}{u}\right)^{n-1}$$

$$\text{d) } E(\hat{U}_{ML}) = \int_0^u t \frac{n}{u} \left(\frac{t}{u}\right)^{n-1} dt, \text{ let } s = \frac{t}{u}, ds = \frac{1}{u} dt$$

$$\text{RHS} = \frac{n}{u} \int_0^1 u^2 s (s^{n-1}) ds = nu \frac{s^{n+1}}{n+1} \Big|_0^1 = \frac{nu}{n+1}$$

$E(\hat{U}_{ML}) \neq u$, The estimator is biased.

$$\text{e) } \lim_{n \rightarrow \infty} E((\hat{U}_{ML} - u)^2) = \lim_{n \rightarrow \infty} E((\hat{U}_{ML})^2) + u^2 - 2uE(\hat{U}_{ML})^2$$

$$E((\hat{U}_{ML})^2) = \int_0^u t^2 \frac{n}{u} \left(\frac{t}{u}\right)^{n-1} dt = \frac{n}{u} \int_0^1 s^2 u^2 s^{n-1} ds = u^2 n \int_0^1 s^{n+1} ds = \frac{nu^2}{n+2}$$

$$\lim_{n \rightarrow \infty} E((\hat{U}_{ML} - u)^2) = \lim_{n \rightarrow \infty} \frac{nu^2}{n+2} + u^2 - 2u \frac{nu}{n+1} = u^2 + u^2 - 2u^2 = 0$$

Therefore, \hat{U}_{ML} converges to u in Mean Square, which implies it converges to u in probability.

Problem 2.

a) The CDF of exponential distribution X is $1 - e^{-\lambda x}$. The half life is just the x that the CDF attains 0.5

Therefore, $1 - e^{-\lambda_i x_i} = \frac{1}{2}$, $\lambda_i = \frac{\log(2)}{x_i}$, where x_i is the half life of the isotope.

$$\lambda_{C15} = \frac{\log(2)}{2.45} \approx 0.283, \lambda_{C10} = \frac{\log(2)}{19.29} \approx 0.036, \lambda_{Sg266} = \frac{\log(2)}{30} \approx 0.023$$

b) Given the observed decay times: $l(x_1, x_2, \dots, x_n | \lambda) = \lambda^n e^{-\lambda \sum x_i}$

$$L(x_1, x_2, \dots, x_n | \lambda) = n \log(\lambda) - \lambda \sum x_i$$

we set $\partial L / \partial \lambda = 0$ and we have $\frac{n}{\lambda} = \sum x_i$, $\hat{\lambda}_{ML} = \frac{n}{\sum x_i}$

It make sense to restrict the value of λ to three values $[0.283, 0.036, 0.023]$, since we want to classify which isotope it is most likely to be.

c) Let the prior π_i be the probability $\lambda = \lambda_i$

$$P(\lambda_i | x_1, x_2, \dots, x_n) = \frac{\pi_i \times P(x_1, x_2, \dots, x_n | \lambda_i)}{\sum_{i=1,2,3} \pi_i \times P(x_1, x_2, \dots, x_n | \lambda_i)},$$

where $P(x_1, x_2, \dots, x_n | \lambda_i)$ is the likelihood of observations given $\lambda = \lambda_i$ which is $\lambda_i^n e^{-\lambda_i \sum x_i}$, $\sum_{i=1,2,3} \pi_i \times P(x_1, x_2, \dots, x_n | \lambda_i)$ is just a constant.

Therefore, the posterior is a constant times the prior times the likelihood.

d) Since the posterior mean minimize the Mean Square Error, i.e. $E(\lambda_{pos} | X) = \underset{\hat{\lambda}(X)}{\text{Argmin}} (\hat{\lambda}(X) - \lambda)^2$, it's a relatively good point estimate of the parameter of the exponential given the observations.

e) For the 10 observation data set, the posterior is $[2.54e-41, 0.92, 0.08]$

For the 100 observation data set, the posterior is $[0.0, 0.143, 0.856]$

For 10 observation data set, since the prior of Sg266 is very small and we don't have enough data, the probability of C10 is greater than that for Sg266. As we increase the sample size, even though the prior of Sg266 is small, the posterior of Sg266 is getting larger than other 2. This means that as more data are observed, the probability of $\lambda = \lambda_3$ will converge to 1.

Problem 3.

a) Let x'_i 's be the sample of ages in years, A reasonable non-parametric estimator of the pmf is $\hat{P}(X = i) = \frac{\sum_{j=0}^N I(X_j = i)}{N}$, which is the percentage of people with age = i from the sample.

b) let p_i be the true probability of age = i .

$$E(\hat{P}_i) = E\left(\frac{\sum_{j=0}^N I(X_j = i)}{N}\right) = \frac{\sum_{j=0}^N E(I(X_j = i))}{N} = \frac{\sum_{j=0}^N p_i}{N} = \frac{N p_i}{N} = p_i.$$

Therefore, the estimator is unbiased.

$$\begin{aligned} \text{c) } \lim_{n \rightarrow \infty} E\left(\left(\frac{\sum_{j=0}^N I(X_j=i)}{N} - p_i\right)^2\right) &= E\left(\left(\frac{\sum_{j=0}^N I(X_j=i)}{N}\right)^2\right) + p_i^2 - 2p_i E\left(\frac{\sum_{j=0}^N I(X_j=i)}{N}\right) \\ E\left(\left(\frac{\sum_{j=0}^N I(X_j=i)}{N}\right)^2\right) &= \frac{1}{N^2} E\left(\left(\sum_{j=0}^N I(X_j=i)\right)^2\right) = \frac{1}{N^2} E\left(\sum_{j=0}^N \sum_{k=0}^N I(X_j=i) I(X_k=i)\right) \\ &= \frac{1}{N^2} (E\left(\sum_{j=0}^N I^2(X_j=i)\right) + E\left(\sum_j \sum_{k \neq j} I(X_j=i) I(X_k=i)\right)) \end{aligned}$$

Since X_j, X_k are independent,

$$E\left(\sum_j \sum_{k \neq j} I(X_j=i) I(X_k=i)\right) = \sum_j \sum_{k \neq j} E(I(X_j=i) I(X_k=i)) = \sum_j \sum_{k \neq j} p_i^2 = N(N-1)p_i^2$$

$$E(I^2(X_l=i)) = p_i, \text{ since } I^2(\cdot) \equiv I(\cdot)$$

$$\text{Hence, } \lim_{n \rightarrow \infty} E\left(\left(\frac{\sum_{j=0}^N I(X_j=i)}{N} - p_i\right)^2\right) = \lim_{n \rightarrow \infty} \frac{Np_i + N(N-1)p_i}{N^2} + p_i^2 - 2p_i p_i = 0$$

Since the i above is arbitrary, therefore the estimator converges to true parameter in Mean Square, therefore it's consistent.

d) If we record the age at a higher granularity (in minutes), for each minutes the number of observation would be relatively small. And there might be no observation for some time. Therefore the method we discuss above would be not accurate, since the empirical distribution needs large amount of data to converge to the true distribution.

Problem 4.

a) The likelihood function of Poisson distribution is $l(\vec{X}|\lambda) = \frac{\lambda^{\sum x_i} e^{-n\lambda}}{\prod x_i!}$

The log likelihood is $L(\vec{X}|\lambda) = \sum x_i \log(\lambda) - n\lambda - \sum \log(x_i)$, setting $\partial L / \partial \lambda = 0$
 $\frac{\sum x_i}{\lambda} - n = 0$, Therefore, $\hat{\lambda}_{ML} = \frac{\sum x_i}{n}$

b) ML fitting errors:

2 days: 0.432813611914
September: 0.506881457941

Nonparametric fitting errors:

2 days: 0.518518518519
September: 0.402777777778

c) The ML method is better for the 2 day data, and the Non-parametric method is better for the September data. For small amount of data available, the Non-parametric method are not precise, since the EPMF need more data to converge to the true PMF. Therefore we should consider Parametric method like MLE to model the data. For large amount data available, the Non-parametric model fits the data better.

Problem 5.

a) Let y_1, \dots, y_n be samples from a $Y \sim \text{Bernoulli}(p)$

MLE:

From the lecture notes, we know that MLE for Bernoulli is $\frac{n_1}{n_1+n_2} = \frac{\sum_i(y_i)}{n}$, where n_1, n_2 are number of 1 and 0's

Moment Estimator :

First moment: $E(Y) = p = \sum y_i/n$, therefore $\hat{p}_1 = \frac{\sum y_i}{n}$

Second moment: $\text{Var}(Y) = p(1-p) = E(Y^2) - E(Y)^2 = \frac{\sum y_i^2}{n} - (\frac{\sum y_i}{n})^2$

Therefore, $(p - 0.5)^2 = \frac{1}{4} - \frac{\sum y_i^2}{n} + (\frac{\sum y_i}{n})^2$, $\hat{p}_2 = (\frac{1}{4} - \frac{\sum y_i^2}{n} + (\frac{\sum y_i}{n})^2)^{\frac{1}{2}} + 0.5$

b) Let y_1, \dots, y_n be samples from a $Y \sim \text{Geometric}(p)$

MLE:

The log likelihood given y_i 's is $L(y_i's|p) = ((\sum y_i) - n)\log(1-p) + n\log(p)$

Maximizing the log likelihood yields: $-\frac{\sum y_i - n}{1-p} + \frac{n}{p} = 0$, $\hat{p}_{ML} = \frac{n}{\sum y_i} = \frac{1}{\bar{y}}$

Moment Estimator :

First moment: $E(Y) = \frac{1}{p} = \sum y_i/n$ therefore $\hat{p}_1 = \frac{n}{\sum y_i}$

Second moment: $\text{Var}(Y) = \frac{1-p}{p^2} = E(Y^2) - E(Y)^2 = \frac{\sum y_i^2}{n} - (\frac{\sum y_i}{n})^2 = C$

Therefore $\hat{p}_2 = \frac{\sqrt{4C+1}-1}{2C}$

c) Let y_1, \dots, y_n be samples from a $Y \sim \text{Poisson}(\lambda)$

MLE:

From Problem 4 a), we know the MLE is $\hat{\lambda}_{ML} = \frac{\sum y_i}{n}$

Moment Estimator :

First moment: $E(Y) = \lambda = \sum y_i/n$ therefore $\hat{\lambda}_1 = \frac{\sum y_i}{n}$

Second moment: $\text{Var}(Y) = \lambda = E(Y^2) - E(Y)^2 = \frac{\sum y_i^2}{n} - (\frac{\sum y_i}{n})^2$

Therefore $\hat{\lambda}_2 = \frac{\sum y_i^2}{n} - (\frac{\sum y_i}{n})^2$

d) Let y_1, \dots, y_n be samples from a $Y \sim \text{Exponential}(\lambda)$

MLE:

From Problem 2 b), we know the MLE is $\hat{\lambda}_{ML} = \frac{n}{\sum y_i}$

Moment Estimator :

First moment: $E(Y) = \lambda^{-1} = \sum y_i/n$ therefore $\hat{\lambda}_1 = \frac{n}{\sum y_i}$

Second moment: $\text{Var}(Y) = \lambda^{-2} = E(Y^2) - E(Y)^2 = \frac{\sum y_i^2}{n} - (\frac{\sum y_i}{n})^2$

Therefore $\hat{\lambda}_2 = (\frac{\sum y_i^2}{n} - (\frac{\sum y_i}{n})^2)^{-\frac{1}{2}}$

