Mathematical statistics
Homework 1

Dima Orekhov

October 19, 2018

1 Problem 1

$$f_i(x, y) = f_i(x) f_i(y), \tag{1}$$

Where

$$f_i(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp -\frac{(x-\mu_i)^2}{2\sigma^2} \tag{2}$$

Thus

$$f_i(x, y) = \frac{1}{2\sigma^2\pi} \exp -\frac{(x-\mu_i)^2 + (y-\mu_i)^2}{2\sigma^2} \tag{3}$$

Because x_i and y_i are two independent random values.

Maximum likelihood estimation of σ^2:

$$L = \prod_{i=1}^{n} f_i(x_i, y_i) \tag{4}$$

$$l = \ln \prod_{i=1}^{n} f_i(x_i, y_i) = \sum_{i=1}^{n} \ln f_i(x_i, y_i) = \tag{5}$$

$$= \sum_{i=1}^{n} \ln \frac{1}{2\pi\sigma^2} - \frac{(x_i-\mu_i)^2 + (y_i-\mu_i)^2}{2\sigma^2} \tag{6}$$

$$\frac{\partial l}{\partial \sigma^2} = \sum_{i=1}^{n} [\ln \frac{1}{2\pi\sigma^2}]' - ((x_i - \mu_i)^2 + (y_i - \mu_i)^2)[\frac{1}{2\sigma^2}]' \tag{7}$$

$$[\ln \frac{1}{2\pi\sigma^2}]' = 2\pi\sigma^2 [\frac{1}{2\pi\sigma^2}]' = -\frac{2\pi\sigma^2}{(2\pi\sigma^2)^2}[2\pi\sigma^2]' = -\frac{2\pi}{2\pi\sigma^2} = -\frac{1}{\sigma^2} \tag{8}$$

$$[\frac{1}{2\sigma^2}]' = -\frac{1}{4(\sigma^2)^2}[2\sigma^2]' = -\frac{2}{4(\sigma^2)^2} = -\frac{1}{2(\sigma^2)^2} \tag{9}$$

$$\frac{\partial l}{\partial \sigma^2} = \sum_{i=1}^{n} -\frac{1}{\sigma^2} + \frac{1}{2(\sigma^2)^2}((x_i - \mu_i)^2 + (y_i - \mu_i)^2) = \tag{10}$$

$$= -\frac{n}{\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}((x_i - \mu_i)^2 + (y_i - \mu_i)^2) \tag{11}$$

Equating partial derivative to zero to obtain MLE of σ^2:

$$-\frac{n}{\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}((x_i - \mu_i)^2 + (y_i - \mu_i)^2) = 0 \tag{12}$$

$$\frac{n}{\sigma^2} = \frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}((x_i - \mu_i)^2 + (y_i - \mu_i)^2) \tag{13}$$

$$\frac{2n\sigma^2}{2(\sigma^2)^2} = \frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}((x_i - \mu_i)^2 + (y_i - \mu_i)^2) \tag{14}$$

$$2n\sigma^2 = \sum_{i=1}^{n}((x_i - \mu_i)^2 + (y_i - \mu_i)^2) \tag{15}$$

$$\sigma^2 = \frac{\sum_{i=1}^{n}((x_i - \mu_i)^2 + (y_i - \mu_i)^2)}{2n} \tag{16}$$

Taking partial derivative of l with respect to μ_i

$$\frac{\partial l}{\partial \mu_i} = -\frac{1}{2\sigma^2}[(x_i - \mu_i)^2 + (y_i - \mu_i)^2]' \tag{17}$$

$$[(x_i - \mu_i)^2 + (y_i - \mu_i)^2]' = -2(x_i - \mu_i) - 2(y_i - \mu_i) = -2(x_i + y_i - 2\mu_i) \tag{18}$$

$$\frac{\partial l}{\partial \mu_i} = \frac{1}{\sigma^2}(x_i + y_i - 2\mu_i) \tag{19}$$

Obtaining the system of equations:

$$\begin{cases} \sigma^2 = \frac{\sum_{i=1}^{n}((x_i-\mu_i)^2+(y_i-\mu_i)^2)}{2n} \\ \frac{1}{\sigma^2}(x_i + y_i - 2\mu_i) = 0 \end{cases} \tag{20}$$

$$\begin{cases} \sigma^2 = \frac{\sum_{i=1}^{n}((x_i-\mu_i)^2+(y_i-\mu_i)^2)}{2n} \\ x_i + y_i - 2\mu_i = 0 \end{cases} \tag{21}$$

$$\begin{cases} \sigma^2 = \frac{\sum_{i=1}^{n}((x_i-\mu_i)^2+(y_i-\mu_i)^2)}{2n} \\ \mu_i = \frac{x_i+y_i}{2} \end{cases} \tag{22}$$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}((x_i - \frac{x_i+y_i}{2})^2 + (y_i - \frac{x_i+y_i}{2})^2)}{2n} = \tag{23}$$

$$= \frac{\sum_{i=1}^{n}((\frac{x_i-y_i}{2})^2 + (\frac{-x_i+y_i}{2})^2)}{2n} = \frac{\sum_{i=1}^{n}(\frac{1}{4}(x_i-y_i)^2 + \frac{1}{4}(-x_i+y_i)^2)}{2n} = \tag{24}$$

$$= \frac{\sum_{i=1}^{n}\frac{1}{4}2(x_i-y_i)^2}{2n} = \frac{\sum_{i=1}^{n}(x_i-y_i)^2}{4n} \tag{25}$$

Proof of the unconsistency:
Let $\xi_i = (x_i - y_i)$ for $i = 1, ..., n$, let's check if our estimation is unbiased:

$$E\frac{1}{4n}\sum_{i=1}^{n}\xi_i^2 = \frac{1}{4n}E\sum_{i=1}^{n}\xi_i^2 = \frac{1}{4n}\sum_{i=1}^{n}E\xi_i^2 \tag{26}$$

$$E\xi_i^2 = D\xi_i + [E\xi_i]^2 \tag{27}$$

$$D\xi_i = D(x_i - y_i) = Dx_i + (-1)^2Dy_i = 2\sigma^2 \tag{28}$$

$$[E\xi_i]^2 = [E(x_i - y_i)]^2 = [Ex_i - Ey_i]^2 = [\mu_i - \mu_i]^2 = 0 \tag{29}$$

$$E\xi_i^2 = 2\sigma^2 + 0 = 2\sigma^2 \tag{30}$$

$$\frac{1}{4n}\sum_{i=1}^{n}E\xi_i^2 = \frac{1}{4n}\sum_{i=1}^{n}2\sigma^2 = \frac{2n\sigma^2}{4n} = \frac{\sigma^2}{2} \tag{31}$$

Our estimation of σ^2 is biased. It is also asymptotically biased:

$$\lim_{n\to\infty}\frac{\sigma^2}{2} = \frac{\sigma^2}{2} \tag{32}$$

Consistent estimations can't be asymptotically biased. Thus out estimation is not consistent.

Q.E.D.

2 Problem 2

Let's make transition from Binomial distribution to Bernoulli distribution. Let:

$$a_i = \begin{cases} 1 & \text{In case } X_i \text{ contains at least one 1} \\ 0 & \text{otherwise} \end{cases} \tag{33}$$

γ will be the parameter of such Bernoulli distribution, it is the probability of X_i containing at least one 1. Let's rewrite γ as a function of p and n_i, parameters of initial Binomial distribution:

$$\gamma = 1 - \binom{n_i}{0}p^0(1-p)^{n_i} = 1 - (1-p)^{n_i} \tag{34}$$

Likelihood function:

$$L(\gamma) = \prod_{i=1}^{n} \gamma^{a_i} (1-\gamma)^{1-a_i} \tag{35}$$

$$l(\gamma) = \sum_{i=1}^{n} \ln \gamma^{a_i} (1-\gamma)^{1-a_i} = \sum_{i=1}^{n} a_i \ln \gamma + (1-a_i) \ln 1 - \gamma \tag{36}$$

Substitute γ with the result of equation 34:

$$l(p) = \sum_{i=1}^{n} a_i \ln (1 - (1-p)^{n_i}) + (1-a_i) \ln (1 - 1 + (1-p)^{n_i}) = \tag{37}$$

$$= \sum_{i=1}^{n} a_i \ln (1 - (1-p)^{n_i}) + (1-a_i) \ln (1-p)^{n_i} \tag{38}$$

Taking derivative:

$$\frac{dl(p)}{dp} = \sum_{i=1}^{n} \frac{a_i n_i (1-p)^{n_i - 1}}{(1-(1-p)^{n_i})} - \frac{(1-a_i) n_i (1-p)^{n_i - 1}}{(1-p)^{n_i}} = \tag{39}$$

$$= \sum_{i=1}^{n} \frac{a_i n_i (1-p)^{n_i - 1}}{(1-(1-p)^{n_i})} - \frac{(1-a_i) n_i}{(1-p)} \tag{40}$$

Equating to zero:

$$\sum_{i=1}^{n} \frac{a_i n_i (1-p)^{n_i - 1}}{(1-(1-p)^{n_i})} - \frac{(1-a_i) n_i}{(1-p)} = 0 \tag{41}$$

$$\sum_{i=1}^{n} \frac{a_i n_i (1-p)^{n_i - 1}}{(1-(1-p)^{n_i})} = \sum_{i=1}^{n} \frac{(1-a_i) n_i}{(1-p)} \tag{42}$$

$$(1-p) \sum_{i=1}^{n} \frac{a_i n_i (1-p)^{n_i - 1}}{(1-(1-p)^{n_i})} = \sum_{i=1}^{n} (1-a_i) n_i \tag{43}$$

$$\sum_{i=1}^{n} \frac{a_i n_i (1-p)^{n_i}}{(1-(1-p)^{n_i})} = \sum_{i=1}^{n} (1-a_i) n_i \tag{44}$$

Proving that likelihood function has 1 root:

In the left part of equation 43 we have a positive constant value or zero. To investigate right part's behaviour lets take it's derivative with respect to p:

$$\frac{df(p)}{dp} = \sum_{i=1}^{n} a_i n_i \frac{-n_i(1-p)^{n_i-1}(1-(1-p)^{n_i}) - (1-p)^{n_i} n_i (1-p)^{n_i-1}}{(1-(1-p)^{n_i})^2} = \tag{45}$$

$$= \sum_{i=1}^{n} a_i n_i \frac{-n_i((1-p)^{n_i-1} - (1-p)^{2n_i-1} + (1-p)^{2n_i-1})}{(1-(1-p)^{n_i})^2} = \tag{46}$$

$$= \sum_{i=1}^{n} a_i n_i \frac{-n_i(1-p)^{n_i-1}}{(1-(1-p)^{n_i})^2} = -\sum_{i=1}^{n} a_i n_i^2 \frac{(1-p)^{n_i-1}}{(1-(1-p)^{n_i})^2} \tag{47}$$

Since $p \in [0,1]$ the expression we obtained has negative value for every p except for $p = 1$ where it is equal zero. Negative derivative means that our function is monotonically decreasing.
Now let's find the range of values of the right part:

$$\lim_{p \to 0} \sum_{i=1}^{n} \frac{a_i n_i (1-p)^{n_i}}{(1-(1-p)^{n_i})} = \infty \tag{48}$$

$$\lim_{p \to 1} \sum_{i=1}^{n} \frac{a_i n_i (1-p)^{n_i}}{(1-(1-p)^{n_i})} = 0 \tag{49}$$

Earlier we've already established that left part of equation 44 is a positive constant value or zero. And since right part is monotonically decreasing function on $[0,1]$ with the range of values $[0,\infty]$ it the equation has one root.

Q.E.D.