12/03/21 End Semester Examination


Kishlay Singh
2003421005


~~Que1~~ → Given:-

~~$p(s|\alpha) \propto s^{-\alpha}$~~

~~$p(s|\alpha) = a s^{-\alpha}$~~

~~where 'a' is a constant of proportionality~~

Que1 → ~~The model probability~~

(a) The model probability distribution for S is

$$p(s)\, ds = (\alpha-1)\, S_0^{\alpha-1}\, S^{-\alpha}\, ds$$

where the factor $\alpha-1$ in front of the terms arises from the normalization requirement.

$$\int_{S_0}^{\infty} ds\, p(s) = 1$$

So the likelihood function $L$ for n observed sources is:-

$$L = \prod_{i=1}^{n} (\alpha-1)\, S_0^{\alpha-1}\, S_i^{-\alpha}$$

Taking natural log both sides:-

$$\ln L = \sum_{i=1}^{n} \left[\ln(\alpha-1) + (\alpha-1)\ln S_0 - \alpha \ln S_i\right]$$

Maximizing $\ln L$ w.r.t. $\alpha$:-

$$\frac{\partial \ln L}{\partial \alpha} = \sum_{i=1}^{n}\left[\frac{1}{\alpha-1} + \ln S_0 - \ln S_i\right] = 0$$

We find the minimum when

$$\alpha = 1 + \frac{n}{\sum_{i=1}^{n} \ln(S_i/S_0)}$$

(b). Suppose we only observe one source with flux twice the cut-off, $S_1 = 2S_0$, then

$$\alpha = 1 + \frac{1}{\ln 2} = 2.44$$

but with a large uncertainity.

(c). To get a rough estimate of width of the credibility interval for $\alpha$; we can compute:

$$\sigma_\alpha^{-2} \simeq -\frac{\partial^2 \ln L}{\partial \alpha^2} = \sum_{i=1}^{n} \frac{1}{(\alpha-1)^2}$$

For single source, $\alpha = 2.44$ and $n=1$, so the estimate of the error is

$$\sigma_\alpha = 1.44$$

Que 2 →

(a) Detection of photons by telescope is an independent event than the previous one

The probability of detecting N out of M photons is :-

$$P(N) = \frac{M!}{N!\,(M-N)!}\, p^N (1-p)^{M-N}$$

Since these 2 are independent events ~~so the~~ so the joint probability is :-

$$P(M,N) = \frac{\mu^M}{M!} e^{-\mu} \frac{M!}{N!\,(M-N)!} p^N q^{(M-N)} \quad -①$$

$\mu = \lambda t$ , $q = 1-p$

(b) From the previous expression ①

$$P(M,N) = \frac{\mu^M}{M!} e^{-\mu} \frac{M!}{N!\,(M-N)!} p^N q^{(M-N)}$$

Marginalization over M

$$P(N) = \frac{p^N q^{-N} e^{-\mu}}{N!} \sum_{M=N}^{\infty} \frac{\mu^M \times q^M}{(M-N)!} \quad -②$$

The Sum starts from M=N as M less than N is not possible. Hence meaning that number of detected photons cannot be greater than number of emitted photon in absence of any ~~Systematic~~ Systemic error.

(c). The series is

$$\sum_{M=N}^{\infty} \frac{(\mu q)^M}{(M-N)!}$$

$$= \sum_{K=0}^{\infty} \frac{(\mu q)^{K+N}}{K!} \qquad (\text{Putting } M-N=K)$$

$$= (\mu q)^N \sum_{K=0}^{\infty} \frac{(\mu q)^K}{K!}$$

$$= (\mu q)^N e^{\mu q} \qquad -③$$

From ② and ③, we get

$$p(N) = \frac{p^N q^{-N} e^{-\mu}}{N!} (\mu q)^N e^{\mu q}$$

$$= \frac{\left(\frac{p \cdot \mu q}{q}\right)^N \times e^{\mu(q-1)}}{N!}$$

$$= \frac{(p\mu)^N e^{-p\mu}}{N!} \qquad (\because q-1=-p)$$

Que 3 $\longrightarrow$

(a). Baye's theorem is given by

$$P(\theta|D) = \frac{P(D|\theta)\ P(\theta)}{P(D)}$$

where

$P(\theta|D) \longrightarrow$ Posterior
$P(D|\theta) \rightarrow$ likelihood
$P(\theta) \longrightarrow$ Prior
$P(D) \longrightarrow$ Evidence

P(D) or Evidence essentially represents the probability of the data, given a model. It is usually neglected if only one model is assumed.

However, in ~~...~~ cases where multiple models are involved, the evidence can no longer be ignored and must be explicitly calculated along with the other terms.

(b). When we fit a set of data points to some hypothesized function, a sample value of $\chi^2$ becomes available if we happen to have made independent estimates of the y variances $\sigma_y^2$. The sum of the squares of residuals is equal to $\chi^2$ and its value can be used to test whether or not our data points are well described by the hypothesized function.

If M independent ~~...~~ variables $x_i$ are normally distributed, each with mean $\mu_i$ and variance $\sigma_i^2$, then the quantity $\chi^2$ is defined by :-

$$\chi^2 = \frac{(y_1 - \mu_1)^2}{\sigma_1^2} + \frac{(y_2 - \mu_2)^2}{\sigma_2^2} + \cdots \frac{(y_M - \mu_M)^2}{\sigma_M^2}$$

$$\Rightarrow \cancel{\chi^2 = \sum_{i=1}^{M} (y_i - \mu_i)^2}$$

$$\chi^2 = \sum_{i=1}^{M} \frac{(y_i - \mu_i)^2}{\sigma_i^2}$$

Where $M$ is the degree of freedom, or simply the number of data points.

(c) Goodness of fit of a statistical model describes how well the model fits the set of observations. The flow chart below shows how chi-square can be used to test goodness of fit :(Assuming that $\sigma_i^2$ are given)

Choose a function / Model

↓

Calculate expectation values $\mu_i$ for each $x_i$

↓

Find out $M$ (number of data points)

↓

Calculate $\chi^2$ using $\chi^2 = \sum_{k=1}^{M} \frac{(y_k - \mu_k)^2}{\sigma_k^2}$

↓

Calculate $\chi^2/M$

↓

Choose a value of significance

↓

Determine corresponding values of $\chi^2_{M,\alpha}/M$

↓

Compare $\chi^2_{M,\alpha}/M$ with $\chi^2/M$

↓

| | | |
|---|---|---|
| $\chi^2 > \chi^2_{M,\alpha}$ | $\chi^2/M \sim 1$ | $\chi^2 < \chi^2_{M,1-\alpha}$ |
| Model can be rejected with $[(1-\alpha)\times 100]$ % confidence | Model is a good fit | Over-estimated $\sigma_k$ or fraudulent data |

(d). We find the weighted mean $\overline{M}_Z$ and its standard error $s_{M_Z}$ using

$$\bar{x} = \frac{\sum w_k x_k}{\sum w_k}$$

For $w_k = \dfrac{1}{\sigma_k^2}$ (weight)

and $$s^2 = \left[\frac{\sum w_k}{(\sum w_k)^2 - (\sum w_k^2)}\right] \sum w_k (x_k - \bar{x})^2$$

$$\overline{M}_Z \pm s_{M_Z} = 91.177 \pm 0.006 \text{ GeV}/c^2$$

We then calculate our sample value of $\chi^2$ using the equation in part (b):-

$$\chi^2 = \sum_{k=1}^{4} \frac{(M_k - \overline{M}_Z)^2}{\sigma_k^2} \simeq 2.78$$

We expect this value of $\chi^2$ to be drawn from a ~~chi-shapo~~ Chi-Square distribution with 3 degrees of freedom. The number is 3 because we have used up one degree of freedom to calculate $\mu$, the true mass of Z° boson

$$\therefore \chi^2/\nu = 2.78/3 \simeq 0.93$$

Now from the graph of $\alpha$ vs $\chi^2/\nu$, we find that for 3 degrees of freedom, $\alpha \simeq 0.42$, meaning that if the experiment was to be repeated, we won't have a 42% chance of finding a $\chi^2$ larger than

2.78, assuming our hypothesis to be correct.

We therefore have no good reason to reject the hypothesis and conclude that the four measurements of the $z^0$ boson mass are consistent with each other.

Que 5 →

~~(a)~~

(b) The linear fit plot the ~~B~~ functional form of the best fit line is:

$$y = 3.404\,x + 40.3105\ .$$

This function is chosen because the deviations of the observed points form the ~~to the~~ line in the least

(c). A polynomial function would be a better fit as given in 5c (image).

(d). The given data is HeteroSkedastic. This ~~a~~ is because as we can see from plot of 5 b, the variance of the observed data pts are not ~~same~~ same with x.

Que 6 →

| n = 165 | | Predicted -ve | Predicted +ve |
|---|---|---|---|
| Actual Examples | -ve | 50 (a) | 10 (b) |
| | +ve | 5 (c) | 100 (d). |

There are n = 165 data points or examples. Out of which 50 are correct rejections (True Negative); 10 are False alarms (False positive), type 1 error; 5 are False negative misses (over looked danger), type 2 error; 100 are True posive hits. We have K = 2 Classes here

Now accuracy is the percent of correct classifications.

For this model,

$$\text{accuracy} = \frac{a+d}{a+b+c+d} = \frac{TN + TP}{Total}$$

$$= \frac{50+100}{165} = 0.909$$

accuracy = 91%.

Recall or True positive Rate or Senstivity

$$= \frac{TP}{\text{actual positive}} = \frac{d}{c+d} = \frac{100}{5+100} = 0.9523$$

Precision or predicted positive value = $\frac{TP}{\text{Predicted Positive}}$

$$= \frac{d}{b+d} = \frac{100}{10+100} = 0.91$$

False alarm or False positive rate = $\frac{FP}{\text{Actual negative}}$

$$= \frac{b}{a+b} = \frac{10}{50+10} = 0.1666 = \underline{\underline{0.17}}$$

Que 7 → K-Means Algorithm ~~d~~ advantages :

① If variables are huge, then ~~K-mean~~ K-Means most of the times computationally faster than hierarchical clustering if we keep K smalls.

② K-Means produce tighter clusters than hierarchical clustering, especially if the clusters are globular.

K-Means Algorithm disadvantages :-

① Difficult to predict K-Value
② With globular cluster, it didn't work well.
③ Different initial partitions can result in ~~diff~~ different final clusters
④ It does not work very well with clusters (in the original data) of different ~~si~~ size and different density.

K-Means Algorithm ~~is~~ is a clustering algorithm.

classmate

Date

Page

Que 8 $\longrightarrow$ Naive Bayes Classification probability,

$$P(\text{yes} \mid \text{today or Feature}) = \frac{P(\text{Feature} \mid \text{yes}) \times P(\text{yes})}{P(\text{Feature} \mid \text{Yes}) \times P(\text{yes}) + P(\text{Feature} \mid \text{No}) \times P(\text{No})} \quad -(1)$$

Feature values for the given data point are:-

Outlook = Sunny, temperature = Hot

Humidity = Normal, Windy = False

Now since all the features are independent, we can use Multiplication theorem of probability to get

$P(\text{Feature} \mid \text{Yes})$ and $P(\text{Feature} \mid \text{No})$

i.e; $P(\text{feature} \mid \text{Yes}) = P(\text{Sunny} \mid \text{yes}) \times P(\text{Hot} \mid \text{yes}) \times P(\text{Normal} \mid \text{yes}) \times P(\text{False} \mid \text{yes})$

$$= \frac{3}{9} \times \frac{2}{9} \times \frac{6}{9} \times \frac{6}{9} \quad - (2)$$

in the same way

$$P(\text{Feature} \mid \text{No}) = \frac{2}{5} \times \frac{2}{5} \times \frac{1}{5} \times \frac{2}{5} \quad - (3)$$

Now, $P(\text{yes}) = \frac{9}{14}$ $\quad - (4)$

$P(\text{No}) = \frac{5}{14}$ $\quad - (5)$

Putting all of these values in eq ①, we get.

$$P(\text{yes} \mid \text{Feature}) = \frac{\left(\frac{3}{9} \times \frac{2}{9} \times \frac{6}{9} \times \frac{6}{9}\right) \times \frac{9}{14}}{\left(\frac{3}{9} \times \frac{2}{9} \times \frac{6}{9} \times \frac{6}{9}\right) \times \frac{9}{14} + \left(\frac{2}{5} \times \frac{2}{5} \times \frac{1}{5} \times \frac{2}{5}\right) \times \frac{5}{14}}$$

$$= \frac{\frac{1944}{91854}}{\frac{1944}{91854} + \frac{40}{8750}} = 0.8223$$

$$P(\text{yes} \mid \text{Feature}) = \underline{\underline{0.82}}$$