Classical and Quantum Statistical Inference (and a bit more)

M. Skotiniotis*

1 Classical Hypothesis Testing

Exercise 1

For the coin example given in the lectures compute the following;

(i) The likelihood for each hypothesis given that the first coin toss comes out heads

(ii) Determine the average probability of success

(iii) Determine the Stein error rates

Exercise 2

The p-value is defined as the probability of observing an extreme value of some statistic given that the null hypothesis holds true. P-values play a vital role in scientific discoveries, and you may know them better as the "five sigma rule": if $p \leq 0.05$ then the null hypothesis is rejected. The mathematical definition of a p-value is as follows. Let t be a realization. Assuming that the null hypothesis is true the *right* and *left* p-values are

$$\begin{aligned} p_R &:= \Pr(X \geq t|H_0) \\ p_L &:= \Pr(X \leq t|H_0)\,, \end{aligned} \tag{1}$$

whereas the *symmetric* p-value is defined as

$$p_S := 2\min\{p_L,\, p_R\}. \tag{2}$$

Consider the following hypothesis testing scenario. A coin is tossed and we are to decide whether the coin is fair or not.

(i) The coin is thrown 100 times, 60 of which turn out to be heads. Determine the symmetric p-value and whether the null hypothesis is true.

(ii) The 101st coin toss comes up heads. What is the p-value now? Do you accept or reject the null hypothesis

(iii) The 102nd coin toss comes up tails. What is the symmetric p-value now? Do you accept or reject the null hypothesis.

*Email: mskotiniotis@onsager.ugr.es

Exercise 3

Let $X \in \mathbb{R}$ and consider N independent and identically distributed samplings from X. The *critical region* of size N, C_N, consists of all realizations $\boldsymbol{x} \in X^N$ that lead us to reject the null hypothesis, i.e.,

$$C_N := \{\boldsymbol{x} \in \mathbb{R} \,|\, f(\boldsymbol{x}) = H_1\}. \tag{3}$$

(i) Assume that under the null hypothesis X is normally distributed with mean μ_0 and variance σ^2, whereas under the alternative hypothesis X is normally distributed with mean μ_1 and variance σ^2 with $\mu_1 > \mu_0$. Consider the *likelihood ratio* statistic,

$$t = \frac{p(\boldsymbol{x}|H_1)}{p(\boldsymbol{x}|H_0)}\,. \tag{4}$$

Determine the critical region defined by $t \geq 0.05$.

(ii) Repeat the above calculation assuming that under the null hypothesis X follows a Poisson distribution with mean μ_0, whereas under the alternative hypothesis X follows a Poisson distribution with mean μ_1 (again $\mu_1 > \mu_0$).

2 Classical Parameter Estimation

Exercise 4

(i) Consider the exponential distribution $p(x\,|\,\lambda) = \lambda e^{-\lambda x}$. Suppose we take a sample of size n. Determine the maximum likelihood estimate for λ. Check whether it is unbiased.

(ii) Let X be a normally distributed random variable. Find the maximum likelihood estimates for the mean, μ, and variance σ^2. Check whether these are unbiased estimates.

Exercise 5

(i) Compute the Fisher information of the Bernoulli distribution with parameter p

(ii) Compute the Fisher information matrix for the exponential distribution $p(x\,|\,\lambda) = \lambda e^{-\lambda x}$.

(iii) Compute the Fisher Information matrix of the normal distribution.

3 Quantum Parameter Estimation

Exercise 6

Consider the set of qubit pure states that lie in the equator of the Bloch sphere, i.e., states with Bloch vector $\mathbf{v}^T = (v_x, v_y, 0)$.

(i) Write the components of the vector in terms of the azimutal angle ϕ.

(ii) Show that the operator L_ϕ can be written in the form

$$L_\phi = a\mathbb{1} + \mathbf{b}\cdot\boldsymbol{\sigma} \tag{5}$$

where a is a scalar, and $\mathbf{b}$ is a 3-D vector.

(iii) Show that the $\mathbf{b} = \frac{\mathrm{d}\mathbf{v}}{\mathrm{d}\phi} - a\mathbf{v}$.

(iv) Compute the Quantum Fisher Information, $\mathcal{F}_\phi$.

(v) Show that measuring the orthogonal bases $\{|+\rangle, |-\rangle\}$, the Fisher Information of the probability distribution of the outcomes is equal to the Quantum Fisher Information.

Solutions to the Exercises

Solution 1

(i) Recall that the probability distributions for each hypothesis were

$$p(\text{heads}|\text{H}_0) = 0.53, \quad p(\text{tails}|\text{H}_0) = 0.47$$
$$p(\text{heads}|\text{H}_1) = 0.59, \quad p(\text{tails}|\text{H}_1) = 0.41,$$

and that each hypothesis occurs with prior probability $\eta_0 = 0.65, \eta_1 = 0.35$. The likelihood function $\ell(x|\text{H}_k)$ is thus

$$\ell(0|\text{H}_0) = 0.53, \quad \ell(0|\text{H}_1) = 0.59.$$

(ii) The average probability of success is given by

$$\begin{aligned} P_S &= \frac{1}{2}\left(1 + \frac{1}{2}\|p_0\,\eta_0 - p_1\eta_1\|\right) \\ &= \frac{1}{2}\left(1 + \frac{1}{2}(|0.53 * 0.65 - 0.59 * 0.65| + |0.47 * 0.35 - 0.41 * 0.35|)\right) \\ &= 0.515 \end{aligned}$$

(iii) The Stein error rate is given by the relative entropy between p_0 and p_1 specifically

$$\begin{aligned} D(p_0\|p_1) &= 0.53 \log_2 \frac{0.53}{0.59} + 0.47 \log_2 \frac{0.47}{0.41} \\ &= 0.0106 \end{aligned}$$

Solution 2

(i) We compute

$$p_R = \frac{1}{2^{100}} \sum_{n=60}^{100} \binom{100}{x} = 0.02844 = \frac{1}{2^{100}} \sum_{n=0}^{40} \binom{100}{x} = p_L$$

Hence the symmetric p value is

$$p_S = 2 * 0.02844 = 0.0568$$

and we accept the fair coin hypothesis.

(ii) We now obtain the following

$$p_R = \frac{1}{2^{101}} \sum_{n=61}^{101} \binom{101}{x} = 0.023022 = \frac{1}{2^{101}} \sum_{n=0}^{40} \binom{101}{x} = p_L$$

and the p value now is $p_S = 0.0460$. So a single coin toss later we are now lead to the rejection of the hypothesis.

(iii) We now compute

$$p_R = \frac{1}{2^{102}} \sum_{n=61}^{102} \binom{102}{x} = 0.0297 = \frac{1}{2^{102}} \sum_{n=0}^{41} \binom{102}{x} = p_L$$

and the corresponding p value now reads $p = 0.594$. After yet another coin toss we are now lead to accepting the fair coin hypothesis.

This is an example of what is known as p-hacking, and why you should be very skeptical when people use p-values in statistics.

Solution 3

(i) As we are dealing with normally distributed, i.i.d. random variables the statistics t is explicitly given by

$$t = \frac{e^{-\sum_{m=1}^{n} \frac{(x_m-\mu_1)^2}{2\sigma^2}}}{e^{-\sum_{m=1}^{n} \frac{(x_m-\mu_0)^2}{2\sigma^2}}} \geq 0.05\,.$$

Taking the natural logarithm on both sides gives

$$\ln t = -\sum_{m=1}^{n} \frac{(x_m - \mu_1)^2}{2\sigma^2} + \sum_{m=1}^{n} \frac{(x_m - \mu_0)^2}{2\sigma^2} \geq \ln 0.05\,.$$

Expanding and the squares and re-arranging the above reduces to

$$\ln t = \sum_{m=1}^{n} \geq \frac{2\sigma^2 \, \ln 0.0.5 - \frac{n}{2}(\mu_0^2 - \mu_1^2)}{\mu_1 - \mu_0}$$

Dividing by the number of samples n and recalling that $\frac{1}{n}\sum_{m=1}^{n} x_m = \mathbb{E}[\mathbf{x}]$ we finally arrive at

$$\frac{1}{n} \ln t = \mathbb{E}[\mathbf{x}] \geq \frac{2\sigma^2 \, \ln 0.0.5 - \frac{n}{2}(\mu_0^2 - \mu_1^2)}{n(\mu_1 - \mu_0)}\,. \tag{6}$$

Hence our critical region consists of all n-dimensional vectors $\mathbf{x}$ whose average satisfies the inequality in Eq. (6).

(ii) Recalling that the Poisson distribution is given by

$$P(x, \mu) = \frac{\mu^x e^{-\mu}}{x!}$$

a similar computation to the one of **(i)** yields

$$\frac{1}{n} \ln t = \mathbb{E}[\mathbf{x}] \geq \frac{\ln 0.05 + (\mu_1 - \mu_0)}{n(\ln \mu_1 - \ln \mu_0)}\,. \tag{7}$$

Hence our critical region consists of all n-dimensional vectors $\mathbf{x}$ whose mean value satisfies the inequality in Eq. (7).

Solution 4

(i) The likelihood, is

$$\ell\left(\lambda \,|\, x_1, \ldots, x_n\right) = \prod_{i=1}^{n} \left(\lambda e^{-\lambda x_i}\right) = \lambda^n \exp\left(-\lambda \sum_{i=1}^{n} x_i\right) = \lambda^n \exp(-n\lambda\bar{x}) \,.$$

From which one gets the following expression for the log-likelihood

$$\ln \ell = n \ln \lambda - n\lambda\bar{x} \,,$$

and so

$$\frac{\mathrm{d} \ln \ell}{\mathrm{d}\lambda} = \frac{n}{\lambda} - n\bar{x} \,.$$

It follows that $\ln \ell$ (and hence ℓ) has a unique maximum at $\hat{\lambda} = 1/\bar{x}$ and this is therefore the maximum likelihood estimator of λ.

(ii) Computing $\mathbb{E}\left[\hat{\lambda}_{\mathrm{MLE}}\right]$ one obtains

$$\mathbb{E}\left[\hat{\lambda}_{\mathrm{MLE}}\right] = \int_{\mathbb{R}_+^n} \frac{n}{\sum_{i=1}^n x_i} \lambda^n e^{-\lambda \sum_{i=1}^n x_i} \mathrm{d}^n \mathbf{x}. \tag{8}$$

Inserting the identity

$$\int_0^\infty \delta\left(\textstyle\sum_{i=1}^n x_i - w\right) \mathrm{d}w = 1, \tag{9}$$

where $\delta(x)$ is the Dirac delta distribution, into Eq. (8) we have

$$\mathbb{E}\left[\hat{\lambda}_{\mathrm{MLE}}\right] = n\lambda^n \int_0^\infty \frac{e^{-\lambda w}}{w} \mathrm{d}w \int_{\mathbb{R}_+^n} \delta\left(\textstyle\sum_{i=1}^n x_i - w\right) \mathrm{d}^n \mathbf{x}.$$

Now rescale x_i as $x_i = w y_i$, so that $\mathrm{d}^n x = w^n \mathrm{d}^n y$, and

$$\begin{aligned}
\mathbb{E}\left[\hat{\lambda}_{\mathrm{MLE}}\right] &= n\lambda^n \int_0^\infty w^{n-1} e^{-\lambda w} \mathrm{d}w \int_{\mathbb{R}_+^n} \delta\left[w\left(\textstyle\sum_{i=1}^n y_i - 1\right)\right] \mathrm{d}^n \mathbf{y} \\
&= n\lambda^n \int_0^\infty w^{n-2} e^{-\lambda w} \mathrm{d}w \int_{\mathbb{R}_+^n} \delta\left(\textstyle\sum_{i=1}^n y_i - 1\right) \mathrm{d}^n \mathbf{y} \\
&= n(n-2)!\lambda \,\mathrm{vol}(\Delta^n),
\end{aligned}$$

where $\mathrm{vol}(\Delta^n)$ is the volume of the simplex $\Delta^n = \{(y_1, \ldots, y_n) \mid \sum_{k=1}^n y_k = 1\}$. Using the identity in Eq. (9) and the fact that the exponential distribution is a *bona fide* distribution we obtain

$$\begin{aligned}
1 &= \int_{\mathbb{R}_+^n} \lambda^n e^{-\lambda \sum_{i=1}^n x_i} \mathrm{d}^n \mathbf{x} \\
&= \lambda^n \int_0^\infty w^{n-1} e^{-\lambda w} \mathrm{d}w \int_{\mathbb{R}_+^n} \delta\left(\textstyle\sum_{i=1}^n y_i - 1\right) \mathrm{d}^n \mathbf{y} \\
&= (n-1)! \,\mathrm{vol}(\Delta^n).
\end{aligned}$$

Hence $\mathrm{vol}(\Delta^n) = 1/(n-1)!$ and

$$\mathbb{E}\left[\hat{\lambda}_{\mathrm{MLE}}\right] = \frac{n(n-2)!}{(n-1)!}\lambda = \frac{n}{n-1}\lambda.$$

It follows that $\hat{\lambda}_{\mathrm{MLE}}$ is *not* unbiased for any finite sample but the estimator is *asymptotically unbiased.*

Solution 5

(i) For the Bernouli distribution of a single parameter the Fisher Information is

$$F(p) = \frac{1^2}{p} + \frac{(-1)^2}{1-p} = \frac{1}{p(1-p)} \,.$$

Notice that the Fisher information is inversely proportional to the variance of the Bernouli distribution.

(ii) For the exponential distribution of a single parameter, the Fisher Information reads

$$F[\lambda] = \frac{1}{\lambda}\int_0^\infty e^{-\lambda x}(1-\lambda x)^2\,\mathrm{d}x = \frac{1}{\lambda^2}\,.$$

(iii) The normal distribution has two variables, μ and σ^2. We thus need to build the Fisher information matrix whose elements are

$$F_{ij} = \int_{-\infty}^{\infty} p(x|\mu,\sigma^2)\left(\frac{\partial p(x|\mu,\sigma^2}{\partial \mu}\right)\left(\frac{\partial p(x|\mu,\sigma^2}{\partial \sigma^2}\right)\mathrm{d}x\,.$$

Computing the matrix elements results in

$$F(\mu,\sigma^2) = \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{pmatrix}$$

Solution 6

(i) Using polar coordinates any vector can be written as $\mathbf{r} = (\sin\theta\cos\phi,\, \sin\theta\sin\phi,\, \cos\theta)$ where $\theta \in (0,\pi)$, and $\phi \in (0, 2\pi)$. Given that $\mathbf{v} = (v_x,\, v_y,\, 0)^T$, it follows that $\theta = \frac{\pi}{2}$ and hence

$$\begin{aligned} v_x &= \cos\phi \\ v_y &= \sin\phi \end{aligned}$$

(ii) Since L_ϕ is a linear operator acting on $\mathcal{H}_2$ it can be expanded in terms of the operator basis $\{\sigma_i\}_{i=0}^3$ as

$$L_\phi = \sum_i \mathrm{Tr}(\sigma_i\, L_\phi)\,\sigma_i$$

Using the fact that $\sigma_0 = \mathbb{1}$ and defining $a = \mathrm{Tr}L_\phi$ and $\mathbf{b} = (\mathrm{Tr}(\sigma_1 L_\phi),\, \mathrm{Tr}(\sigma_2 L_\phi),\, \mathrm{Tr}(\sigma_3 L_\phi))$ gives the final result.

(iii) For the definition of the SLD we have

$$\frac{\mathrm{d}\rho}{\mathrm{d}\phi} = \frac{1}{2}\left(L_\phi\rho + \rho L_\phi\right)\,. \tag{10}$$

Using the Bloch representation of ρ and the solution of **(i)** it follows that

$$\begin{aligned} \frac{\mathrm{d}\rho}{\mathrm{d}\phi} &= \frac{\mathrm{d}}{\mathrm{d}\phi}\left(\frac{\mathbb{1} + \mathbf{v}\cdot\boldsymbol{\sigma}}{2}\right) \\ &= \frac{1}{2}\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}\phi}\cdot\boldsymbol{\sigma}\,, \end{aligned} \tag{11}$$

with

$$\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}\phi} = (-\sin\phi,\, \cos\phi, 0)\,.$$

Therefore Eq. (10) reads

$$\begin{aligned}
2\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}\phi}\cdot\boldsymbol{\sigma} &= (a\mathbb{1} + \mathbf{b}\cdot\boldsymbol{\sigma})(\mathbb{1} + \mathbf{v}\cdot\boldsymbol{\sigma}) + (\mathbb{1} + \mathbf{v}\cdot\boldsymbol{\sigma})(a\mathbb{1} + \mathbf{b}\cdot\boldsymbol{\sigma}) \\
&= 2a\mathbb{1} + 2(\mathbf{b} + a\mathbf{v})\cdot\boldsymbol{\sigma} + \sum_{ij}(b_i v_j + b_j v_i)\sigma_i\sigma_j \\
&= 2(a + \mathbf{b}\cdot\mathbf{v})\mathbb{1} + 2(\mathbf{b} + a\mathbf{v})\cdot\boldsymbol{\sigma}\,,
\end{aligned}$$

where in the last line we have used the identity of Eq. (**??**). It follows that $a + \mathbf{b}\cdot\mathbf{v} = 0$ and $\mathbf{b} + a\mathbf{v} = \frac{\mathrm{d}\mathbf{v}}{\mathrm{d}\phi}$. Since $\mathbf{v}\cdot\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}\phi} = 0$, the second condition implies the first one and $\mathbf{b} = \frac{\mathrm{d}\mathbf{v}}{\mathrm{d}\phi} - a\mathbf{v}$ for any a. Hence, *the SLD is not uniquely defined*:

$$L_\phi = a\mathbb{1} + (\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}\phi} - a\mathbf{v})\cdot\boldsymbol{\sigma} \quad \text{for any } a \in \mathbb{R}\,.$$

(iv) The Quantum Fisher Information is given by

$$\begin{aligned}
\mathcal{F}_\phi &= \mathrm{Tr}(L_\phi^2\rho) \\
&= \mathrm{Tr}\left((a\mathbb{1} + (\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}\phi} - a\mathbf{v})\cdot\boldsymbol{\sigma})^2(\frac{\mathbb{1} + \mathbf{v}\cdot\boldsymbol{\sigma}}{2})\right) \\
&= a^2 + \frac{\mathrm{d}\mathbf{v}}{\mathrm{d}\phi}\cdot\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}\phi} + a^2\mathbf{v}\cdot\mathbf{v} + 2a(\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}\phi} - a\mathbf{v})\cdot\mathbf{v} \\
&= \frac{\mathrm{d}\mathbf{v}}{\mathrm{d}\phi}\cdot\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}\phi} = 1
\end{aligned}$$

(v) The probability distribution we obtain if we perform the a measurement in the $|\pm\rangle$ basis is given by

$$p(\pm|\phi) = \begin{cases} \cos^2\frac{\phi}{2} & \text{for } + \\ \sin^2\frac{\phi}{2} & \text{for } - \end{cases} \tag{12}$$

Computing the Fisher information for this probability distribution gives

$$\begin{aligned}
F(p(\pm|\phi) &= \frac{\left(\frac{\mathrm{d}}{\mathrm{d}\phi}p(+|\phi)\right)^2}{p(+|\phi)} + \frac{\left(\frac{\mathrm{d}}{\mathrm{d}\phi}p(-|\phi)\right)^2}{p(-|\phi)} \\
&= \sin^2\frac{\phi}{2} + \cos^2\frac{\phi}{2} = 1
\end{aligned}$$

The same as the quantum Fisher information. Hence this measurement is optimal.