הטכניון, מכון טכנולוגי לישראל

מבוא למערכות לומדות

236756

סמסטר אביב 2019

תרגיל מספר: 1

מגישים:

אמיר אביבי	3	7	8	3	8	1	5	0	3	
שם מלא	מספר ת.ז.									

ציון

Exercise 1 – Probability & Python

Question 1:

P(Coin is forged| $\frac{head}{total\ tosses} = \frac{7}{10}$)=

$$\frac{\mathrm{P}(\frac{head}{total\ tosses} = \frac{7}{10} \mid \mathrm{Coin\ is\ forged}) \cdot \mathrm{P(Coin\ is\ forged)}}{P\left(\frac{head}{total\ tosses} = \frac{7}{10}\right)} =$$

$$\frac{P(X\sim Bin(10{,}0.8) = 7) \cdot \frac{1}{1000}}{P\left(\frac{head}{total\ tosses} = \frac{7}{10} | Coin\ is\ forged\right) P(Coin\ is\ forged) + P\left(\frac{head}{total\ tosses} = \frac{7}{10} | Coin\ is\ fair\right) P(coins\ is\ fair)}$$

$$= \frac{\binom{10}{7} \cdot 0.8^7 \cdot 0.2^3 \cdot \frac{1}{1000}}{\binom{10}{7} \cdot 0.8^7 \cdot 0.2^3 \cdot \frac{1}{1000} + \binom{10}{7} \cdot 0.5^7 \cdot 0.5^3 \cdot \frac{999}{1000}} = \frac{0.8^7 \cdot 0.2^3}{\cdot 0.8^7 \cdot 0.2^3 + \cdot 0.5^{10} \cdot 999} = 1.716 * 10^{-3}$$

Question 2:

We are being asked if the probability that the number of boys in a family is bigger than the number of girls.

As we can see the number of girls in a family is distributed as a Geometric random variable meaning:

$The\ \mathrm{number\ of\ girls\ in\ a\ family} \sim \mathrm{Geo}(0.5)$, and the mean behavior of such random variable is $\frac{1}{p}$, which in our case equals 2, meaning that the **common** family will have 1 boy and 1 girl.

For each family we calculate the following:

P(Boys in the family > Girls in the family) = P(The first born baby was a boy) = 0.5 = P(Boys in the family <= Girls in the family)

By using the law of large numbers , and defining the families to be an i.i.d series of random variables, we deduce that the number of males and females in the country converges to its average behavior which means, that in probability, half of the country are **males and half are females, so the answer to the question is: The number of boys and girls is equal.**

Question 3:

Maximum Likelihood Estimation (MLE)

Definition:

$$\widehat{\Theta}_{MLE} \triangleq \underset{\Theta\in\mathcal{R}^p}{\operatorname{argmax}}\, p(D|\theta) = \underset{\Theta\in\mathcal{R}^p}{\operatorname{argmax}}\, logp(D|\theta)$$

Under an i.i.d.(independent, identically distributed) assumption:

$$L(\theta) \triangleq p(D|\theta) = p(x_1, \dots, x_n|\theta) = \prod_{k=1}^{n} p_X(x_k|\theta)$$

$$l(\theta) \triangleq \log\big(L(\theta)\big) = \sum_{k=1}^{n} logp_x(x_k|\theta)$$

Therefore:

$$\widehat{\Theta}_{MLE} = \underset{\Theta\in\mathcal{R}^p}{\operatorname{argmax}}\{l(\theta)\}$$

Binomial distribution:

$$\hat{p}_{MLE} = argmax\,(p(x_1, x_2, \dots, x_n|n, p)) = argmax(\log(p(x_1, x_2, \dots, x_n|n, p))) =$$

$$log\left(\prod_{k=1}^{n}\binom{n}{x_k}\cdot p^{x_k}\cdot(1-p)^{n-x_k}\right) = \sum_{k=1}^{n} log\binom{n}{x_k} + x_k\cdot\log p + (n-x_k)\cdot\log(1-p)$$

Finding the maximum:

$$\frac{\partial l}{\partial p} = \sum_{k=1}^{n}\frac{x_k}{p} - \frac{(n-x_k)}{(1-p)}$$

$$\sum_{k=1}^{n}\frac{x_k}{p} - \frac{(n-x_k)}{(1-p)} = 0 \Rightarrow \sum_{k=1}^{n}\frac{x_k - x_k\cdot p - n\cdot p + x_k\cdot p}{p(1-p)} = \sum_{k=1}^{n}\frac{x_k - n\cdot p}{p(1-p)} = 0 \Rightarrow$$

$$\sum_{k=1}^{n}\frac{x_k}{p(1-p)} - \frac{n^2}{(1-p)} = 0 \Rightarrow$$

$$\boldsymbol{\hat{p}_{MLE}} = \frac{\boldsymbol{\sum_{k=1}^{n} x_k}}{\boldsymbol{n^2}}$$

$$\widehat{\Theta}_{MLE} = \left(\frac{\sum_{k=1}^{n} x_k}{n^2}\right)$$

Normal distribution:

$$\widehat{\Theta}_{MLE} = argmax\ (p(x_1, x_2, \dots, x_n | \mu, \sigma)) = argmax\ (\log p(x_1, x_2, \dots, x_n | \mu, \sigma)) =$$

$$l(\theta) = \log \prod_{k=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{(x_k - \mu)^2}{2\sigma^2}\} = \log(\frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\{-\frac{1}{2\sigma^2} \sum_{k=1}^{n} (x_k - \mu)^2 \})$$

$$= \log \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} - \frac{1}{2\sigma^2} \sum_{k=1}^{n} (x_k - \mu)^2$$

Finding the maximum:

$$\frac{\partial l}{\partial \mu} = -\frac{1}{\sigma^2} \sum_{k=1}^{n} (x_k - \mu) = 0 \Rightarrow \boldsymbol{\hat{\mu}} = \frac{\sum \boldsymbol{x_k}}{\boldsymbol{n}}$$

$$\frac{\partial l}{\partial \sigma^2} = -\frac{n}{2} \cdot \frac{1}{\sigma^2} + \frac{1}{\sigma^4} \sum_{k=1}^{n} (x_k - \mu)^2 = 0 \Rightarrow \boldsymbol{\widehat{\sigma^2}} = \frac{\boldsymbol{1}}{\boldsymbol{n}} \sum_{\boldsymbol{k=1}}^{\boldsymbol{n}} (\boldsymbol{x_k} - \boldsymbol{\mu})^{\boldsymbol{2}}$$

$$\widehat{\Theta}_{MLE} = (\frac{\sum \boldsymbol{x_k}}{\boldsymbol{n}}, \frac{\boldsymbol{1}}{\boldsymbol{n}} \sum_{\boldsymbol{k=1}}^{\boldsymbol{n}} (\boldsymbol{x_k} - \boldsymbol{\mu})^{\boldsymbol{2}})$$

Poisson distribution:

$$l(\theta) = \log \prod_{k=1}^{n} \frac{\lambda^{x_k}}{x_k!} \cdot e^{-\lambda} = \sum_{k=1}^{n} x_k \log \lambda - \log(x_k!) - \lambda$$

Finding the maximum:

$$\frac{\partial l}{\partial \lambda} = \sum_{k=1}^{n} \frac{x_k}{\lambda} - 1 = 0 \Rightarrow \boldsymbol{\hat{\lambda}} = \frac{\sum \boldsymbol{x_k}}{\boldsymbol{n}}$$

$$\widehat{\Theta}_{MLE} = (\frac{\sum \boldsymbol{x_k}}{\boldsymbol{n}})$$

Question 4:

a.

$$f_{x_1}(x_1) = \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \cdot \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right)\right\} dx_2$$

$$let\ z_2 = \frac{x_2-\mu_2}{\sigma_2} \Rightarrow\ dz_2 = \frac{dx_2}{\sigma_2} \Rightarrow dx_2 = dz_2 \cdot \sigma_2$$

$$f_{x_1}(x_1) = \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_1\sqrt{1-\rho^2}} \cdot \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)z_2 + {z_2}^2\right)\right\} dz_2 =$$

$$\frac{1}{2\pi\sigma_1\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} \exp\left\{-\left(\frac{\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2}{2(1-\rho^2)} - \frac{2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)z_2}{2(1-\rho^2)} + \frac{{z_2}^2}{2(1-\rho^2)}\right)\right\} dz_2$$

using the given integral solution we define:

$$\mathbf{a} = \frac{\mathbf{1}}{\mathbf{2(1-\rho^2)}}, \mathbf{b} = -\frac{\boldsymbol{\rho}\left(\frac{\mathbf{x_1-\mu_1}}{\boldsymbol{\sigma_1}}\right)}{\mathbf{(1-\rho^2)}}, \mathbf{c} = \frac{\left(\frac{\mathbf{x_1-\mu_1}}{\boldsymbol{\sigma_1}}\right)^{\mathbf{2}}}{\mathbf{2(1-\rho^2)}}$$

$$f_{x_1}(x_1) = \frac{1}{2\pi\sigma_1\sqrt{1-\rho^2}} \cdot \left(\frac{\pi}{\frac{1}{2(1-\rho^2)}}\right)^{\frac{1}{2}} \exp\left\{\frac{\left(-\frac{\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)}{(1-\rho^2)}\right)^2}{4\left(\frac{1}{2(1-\rho^2)}\right)} - \left(\frac{\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2}{2(1-\rho^2)}\right)\right\} =$$

$$\frac{\sqrt{\pi \cdot 2(1-\rho^2)}}{2\pi\sigma_1\sqrt{1-\rho^2}} \cdot \exp\left\{\frac{\rho^2\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 \cdot 2(1-\rho^2)}{4(1-\rho^2)} - \frac{\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2}{2(1-\rho^2)}\right\} =$$

$$\frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left\{\frac{\rho^2\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - \left(\frac{x_1-\mu_1}{\sigma_1}\right)^2}{2(1-\rho^2)}\right\} = \frac{\mathbf{1}}{\sqrt{\mathbf{2}\boldsymbol{\pi}\boldsymbol{\sigma}_{\mathbf{1}}^{\mathbf{2}}}} \mathbf{exp}\left\{-\frac{\mathbf{1}}{\mathbf{2}}\left(\frac{\boldsymbol{x_1-\mu_1}}{\boldsymbol{\sigma_1}}\right)^{\mathbf{2}}\right\}$$

$$\Rightarrow \boldsymbol{X_1 \sim N(\mu_1, \sigma_1^2)}$$

■

Question 4:

b.

$$f_{X_1|X_2}(x_1|x_2) = \frac{f_{X_1,X_2}(x_1,x_2)}{f_{X_2}(x_2)} =$$

$$f_{X_1|X_2}(x_1|x_2) = \frac{\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \cdot \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right)\right\}}{\frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left\{-\frac{1}{2}\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right\}} =$$

$$\frac{1}{\sqrt{2\pi(\sigma_1^2(1-\rho^2))}} \cdot \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right) + \frac{1}{2}\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right\}$$

$$\frac{1}{\sqrt{2\pi(\sigma_1^2(1-\rho^2))}} \cdot \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2 - (1-\rho^2)\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right)\right\} =$$

$$\frac{1}{\sqrt{2\pi(\sigma_1^2(1-\rho^2))}} \cdot \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \rho^2\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right)\right\} =$$

$$\frac{1}{\sqrt{2\pi(\sigma_1^2(1-\rho^2))}} \cdot \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\left(\frac{x_1-\mu_1}{\sigma_1}\right) - \rho\left(\frac{x_2-\mu_2}{\sigma_2}\right)\right)^2\right\} =$$

$$\frac{1}{\sqrt{2\pi(\sigma_1^2(1-\rho^2))}} \cdot \exp\left\{-\frac{1}{2}\left(\frac{x_1-\left(\mu_1+\frac{\rho\sigma_1(x_2-\mu_2)}{\sigma_2}\right)}{\sigma_1\sqrt{(1-\rho^2)}}\right)^2\right\}$$

$$\Rightarrow X_1|X_2 \sim N\left(\mu_1 + \frac{\rho\sigma_1(x_2-\mu_2)}{\sigma_2}, \sigma_1^2(1-\rho^2)\right)$$

■

Question 5:

The Cauchy–Schwarz inequality states that for all vectors u and v of an inner product space it is true that

$$|\langle \mathbf{u}, \mathbf{v}\rangle|^2 \leq \langle \mathbf{u}, \mathbf{u}\rangle \cdot \langle \mathbf{v}, \mathbf{v}\rangle,$$

where $\langle \cdot, \cdot \rangle$ is the inner product. Examples of inner products include the real and complex dot product, see the examples in inner product. Equivalently, by taking the square root of both sides, and referring to the norms of the vectors, the inequality is written as[2][3]

$$|\langle \mathbf{u}, \mathbf{v}\rangle| \leq \|\mathbf{u}\| \|\mathbf{v}\|.$$

Let's define the mean as the inner product in our inner product space (we can do it based on previous probability courses), we get:

$$< u, v > \triangleq E[u \cdot v]$$

By defining the vectors:

$$u = (x - \mu_x)$$
$$v = (y - \mu_y)$$

And thus

$$< (x - \mu_x), (y - \mu_y) > \stackrel{\text{def}}{=} cov(X, Y)$$

$$< (x - \mu_x), (x - \mu_x) > \stackrel{\text{def}}{=} var(X)$$
$$< (y - \mu_y), (y - \mu_y) > \stackrel{\text{def}}{=} var(Y)$$

We get the following result, based on the inequality:

$$|cov(X, Y)|^2 \leq var(X) \cdot var(Y)$$

*Reminder- $var(\cdot) \geq 0$

$$|cov(X, Y)| \leq \sigma_x \sigma_y$$
$$-\sigma_x \sigma_y \leq cov(X, Y) \leq \sigma_x \sigma_y$$
$$-1 \leq \frac{cov(X, Y)}{\sigma_x \sigma_y} \leq 1$$

■