Ellie Makin (erm67)

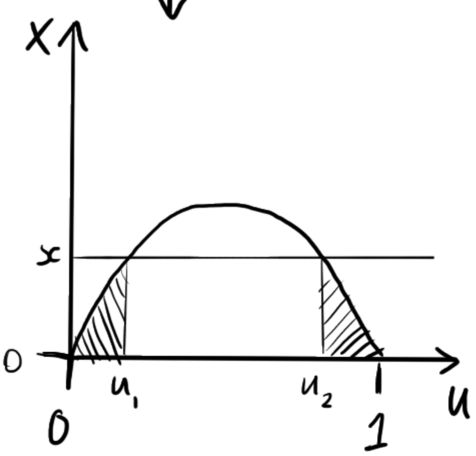
Data Science supervision 1

Example sheet 1

1.

1. $U \sim [0,1)$, $X = U(1-U)$

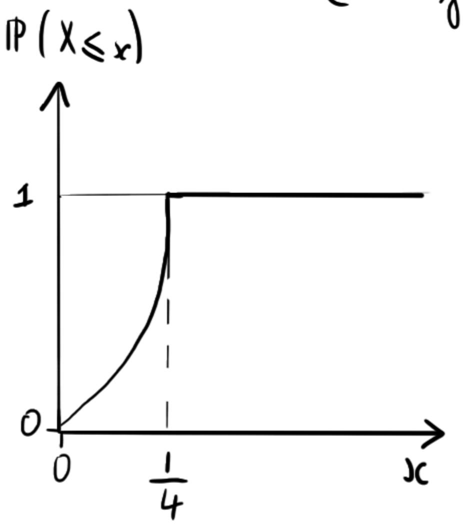
$\mathbb{P}(X \leq x) = \mathbb{P}(U(1-U) \leq x)$



$u - u^2 = x \rightarrow u^2 - u + x = 0$

$u = \dfrac{1 \pm \sqrt{1-4x}}{2}$

$$\therefore \mathbb{P}(X \leq x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 - \sqrt{1-4x} & \text{for } 0 \leq x \leq \frac{1}{4} \\ 1 & \text{for } \frac{1}{4} < x \end{cases}$$



$$\frac{d}{dx}\left(1 - (1-4x)^{\frac{1}{2}}\right)$$

$$= 2(1-4x)^{-\frac{1}{2}}$$

$$\text{pdf}(x) = \frac{\partial}{\partial x}\mathbb{P}(X \leq x) = \begin{cases} 0 & \text{for } x < 0 \\ \dfrac{2}{\sqrt{1-4x}} & \text{for } 0 \leq x \leq \frac{1}{4} \\ 0 & \text{for } \frac{1}{4} < x \end{cases}$$

2. $$\Pr(x_i;\lambda) = \frac{\lambda^{x} e^{-\lambda}}{x_i!}$$

$$\log \Pr(x_i;\lambda) = x_i \ln(\lambda) - \lambda - \ln(x_i!)$$

$$\therefore \log \Pr(x_1, x_2 \dots x_n; \lambda) = \sum_{i=1}^{n} \log \Pr(x_i;\lambda)$$

$$= \ln(\lambda) \sum_{i=1}^{n} x_i - n\lambda - \sum_{i=1}^{n} \ln(x_i!)$$

$$\frac{\partial}{\partial \lambda} \log \Pr(x_1, x_2 \dots x_n; \lambda) = \frac{1}{\lambda} \sum_{i=1}^{n} x_i - n$$

$$\frac{1}{\hat{\lambda}} \sum_{i=1}^{n} x_i - n = 0 \Rightarrow \hat{\lambda} = n^{-1} \sum_{i=1}^{n} x_i \quad \square$$

3.

```
import scipy.optimize
import numpy as np

x = [3,2,8,1,5,0,8]
n = len(x)

def logPr(t):
    l = np.exp(t)   # use exponential transform so that l > 0
    return np.log(l)*sum(x) - n*l - sum(map(
        lambda x_i: np.log(np.math.factorial(x_i)),
        x
    ))


[tMax] = scipy.optimize.fmin(lambda t: -logPr(t), 1)

lMax = np.exp(tMax)     # undo the transform
print(lMax)
```

4. $$\Pr(x;\theta) = \begin{cases} 0 & \text{for } x<0 \\ \frac{1}{\theta} & \text{for } 0 \leq x \leq \theta \\ 0 & \text{for } x>\theta \end{cases}$$

$$\log \Pr(x;\theta) = \begin{cases} -\ln\theta & 0 \leq x \leq \theta \\ -\infty & \text{otherwise} \end{cases}$$

$$\log \Pr(x_1,\dots,x_n;\theta) = \sum_{i=1}^{n} \log \Pr(x_i;\theta)$$

$$= \begin{cases} -n\ln\theta & \max_i x_i \leq \theta \\ -\infty & \text{otherwise} \end{cases}$$

maximize $-n\ln\theta \Rightarrow$ minimize $\ln\theta \Rightarrow$ minimize θ

but $\theta \geq \max_i x_i$

$\therefore \hat{\theta} = \max_i x_i$ $\square$

5. $Pr(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$ Assuming independence.

$$\log Pr(x;\mu,\sigma) = -\ln(\sigma\sqrt{2\pi}) - \frac{(x-\mu)^2}{2\sigma^2}$$

$$\log P_r(x_1,\dots,x_n;\mu,\sigma) = -m\ln(\sigma\sqrt{2\pi}) - \frac{1}{2\sigma^2}\sum_{i=1}^{m}(x_i-\mu)^2$$

$$\log P_r(x_1,\dots,x_n,y_1\dots,y_n;\mu,\sigma,\delta) =$$

$$-(m+n)\ln(\sigma\sqrt{2\pi}) - \frac{1}{2\sigma^2}\left(\sum_{i=1}^{m}(x_i-\mu)^2 + \sum_{i=1}^{n}(y_i-\mu-\delta)^2\right)$$

$$\frac{\partial}{\partial\mu} = \frac{1}{\sigma^2}\left(\sum_{i=1}^{m}(x_i-\mu) + \sum_{i=1}^{n}(y_i-\mu-\delta)\right)$$

$$\frac{\partial}{\partial\sigma} = \frac{-(m+n)}{\sigma} + \frac{1}{\sigma^3}\left(\sum_{i=1}^{m}(x_i-\mu)^2 + \sum_{i=1}^{n}(y_i-\mu-\delta)^2\right)$$

$$\frac{\partial}{\partial\delta} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i-\mu-\delta)$$

$$\frac{\partial}{\partial\mu}, \frac{\partial}{\partial\sigma}, \frac{\partial}{\partial\delta} = 0$$

$$\therefore \frac{1}{\hat\sigma^2}\sum_{i=1}^{m}(x_i-\hat\mu) = 0 \rightarrow -m\hat\mu + \sum_{i=1}^{m}x_i = 0 \rightarrow \boxed{\hat\mu = m^{-1}\sum_{i=1}^{m}x_i}$$

$$\therefore -n\hat\mu - n\hat\delta + \sum_{i=1}^{n}y_i = 0 \rightarrow \boxed{\hat\delta = n^{-1}\sum_{i=1}^{n}y_i - m^{-1}\sum_{i=1}^{m}x_i}$$

$$\therefore \hat\sigma^2 = \sqrt{\left(\frac{1}{m+n}\left(\sum_{i=1}^{m}(x_i-\hat\mu)^2 + \sum_{i=1}^{n}(y_i-\hat\mu-\hat\delta)^2\right)\right)}$$

6. $\Pr(y; x, \lambda) = \dfrac{(\lambda x)^y e^{-\lambda x}}{y!}$

$$\log \Pr(y; x, \lambda) = y \ln(\lambda x) - \lambda x - \ln(y!)$$

$$\log \Pr(y_1, \dots, y_n; x_1, \dots, x_n, \lambda) = \sum_{i=1}^{n} \log \Pr(y_i; x_i, \lambda)$$

$$= \sum_{i=1}^{n} y_i \ln(\lambda x_i) - \lambda \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \ln(y_i!)$$

$$\frac{\partial}{\partial \lambda} = \frac{1}{\lambda} \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} x_i \quad , \frac{\partial}{\partial \lambda} = 0$$

$$\therefore \hat{\lambda} = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i}$$

7.

We can use the following model formula to make sure that the two lines meet at the inflection point:

$$f(x) = \begin{cases} m_1(x - x_{inf}) + c & x \leq x_{inf} \\ m_2(x - x_{inf}) + c & x_{inf} < x \end{cases}$$

where m_1 and m_2 are the gradients of the two lines, (x_{inf}, c) is the position of the inflection point. Assuming we are given the points as a labelled dataset $(x_1, y_1), \dots, (x_n, y_n)$, we can adapt this to account for a random component:

$$Y_i \sim \text{Normal}(f(x_i), \sigma^2)$$

This has likelihood

$$\Pr(y_i; x_i, x_{inf}, m_1, m_2, c, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(y_i - f(x_i))^2 / 2\sigma^2}$$

which can then be $\log$'d and optimised numerically over parameters m_1, m_2, c, and σ in order to fit the dataset, probably including an exponential transform over σ to keep it positive.