HW 2: Machine Learning

Juan Vila

Question 1

Part A

For finding the moments through the method of moments we need to solve $E(X_i) = p$, then we know the that $p = \frac{1}{n}\sum_i^n X_i$. Then the expected value is $E(X_i) = \frac{1}{n}\sum_i^n X_i$.

Part B

For get the second moment we have to solve $\mu^2 + \sigma = \frac{1}{n}\sum_i^n X_i^2$ and replace for $mu = \frac{1}{n}\sum_i^n X_i$. Then, $(\frac{1}{n}\sum_i^n X_i)^2 + \sigma = \frac{1}{n}\sum_i^n X_i^2$ Then, if we solveσ we get:

$$(\frac{1}{n}\sum_i^n X_i)^2 + \sigma = \frac{1}{n}\sum_i^n X_i^2$$
$$\sigma = \frac{1}{n}\sum_i^n X_i^2 - (\frac{1}{n}\sum_i^n X_i)^2$$

We know that the mean is p, then:

$$\sigma = \frac{1}{n}\sum_i^n X_i^2 - (p)^2$$

Than as x_i is indicator variable that take values of 1 and zero, will get that $\sum_i^n X_i^2 = \sum_i^n X_i = pn$.

$$\sigma = p - p^2$$
$$\sigma = (1-p)p$$

Part C

For solving the MLE for a Bernoulli distribution, we need to solve:

$$p = \underset{\sigma}{\text{argmin}}\ L(\sigma) = \pi_{i=1}^n f(\sigma/x_i)$$
$$p = \underset{\sigma}{\text{argmin}}\ ln(L(\sigma)) = -\sum_{i=1}^n ln(f(\sigma/x_i))$$

We know that the bernulli distribution is $f(\sigma/x_i) = p^{x_i}(1-p)^{1-x_i}$

$$p = \underset{\sigma}{\text{argmin}}\ ln(L(p)) = -\sum_{i=1}^n ln(p^{x_i}(1-p)^{1-x_i})$$
$$p = \underset{\sigma}{\text{argmin}}\ ln(L(p)) = -\sum_{i=1}^n ln(p)x_i + ln(1-p)(1-x_i))$$

Now we take the derivate agains p and get:

$$\frac{\partial ln(L(\sigma))}{\partial p} = \frac{1}{p}\sum_{i=1}^{n} x_i - \frac{1}{1-p}\sum_{i=1}^{n} 1 - x_i = 0$$
$$\frac{1}{p}\sum_{i=1}^{n} x_i = \frac{1}{1-p}\sum_{i=1}^{n} 1 - x_i$$
$$(1-p)\sum_{i=1}^{n} x_i = p\sum_{i=1}^{n} 1 - x_i$$
$$\sum_{i=1}^{n} x_i - p\sum_{i=1}^{n} x_i = pn - p\sum_{i=1}^{n} x_i$$
$$p = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Part D

For getting the variance we have to solve:

$$Var(p)^{-1} = -E[\frac{\partial^2 ln(L(\sigma))}{\partial p^2}]$$
$$Var(p)^{-1} = -E[\frac{\partial^2 ln(L(\sigma))}{\partial p^2}] = \frac{1}{p^2}\sum_{i=1}^{n} x_i + \frac{1}{(1-p)^2}\sum_{i=1}^{n} 1 - x_i$$
$$Var(p)^{-1} = -E[\frac{\partial^2 ln(L(\sigma))}{\partial p^2}] = \frac{1}{p^2}np + \frac{1}{(1-p)^2}(1-p)n$$

Simplifiying we get:

$$Var(p)^{-1} = -E[\frac{\partial^2 ln(L(\sigma))}{\partial p^2}] = \frac{n}{p(1-p)}$$
$$Var(p) = \frac{p(1-p)}{n}$$

Question 2

Part A

$$p(x|\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}$$
$$p(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)}\lambda^{\alpha-1}e^{-\beta\lambda}$$
$$\lambda = \underset{\lambda}{\operatorname{argmin}} -ln(L(\lambda)) = -\sum_{i=1}^{n} ln(f(\lambda/x_i))$$
$$\lambda = \underset{\lambda}{\operatorname{argmin}} -ln(L(\lambda)) = -\sum_{i=1}^{n} ln(\frac{e^{-\lambda}\lambda^x}{x!})$$
$$ln(L(\lambda)) = -\sum_{i=1}^{n} -\lambda + x_i ln(\lambda) + ln(x_i!)$$

Now we take the derivates:

$$\frac{\partial ln(L(\lambda))}{\partial \lambda} = -\sum_{i=1}^{n} -1 + x_i \frac{1}{\lambda}$$
$$\sum_{i=1}^{n} x_i \frac{1}{\lambda} = n$$
$$\sum_{i=1}^{n} x_i \frac{1}{n} = \lambda$$

Part B

We are going to calculate λ as the average of the values that we get this is 7.2, now we are going to calculate $P(x < 13/\lambda = 7.2) = 1 - P(x = 13/\lambda = 7.2)$

$$P(x = 13/\lambda = 7.2) = \frac{e^{-7.2} 7.2^{13}}{13!}$$
$$P(x = 13/\lambda = 7.2) = 0.01675$$
$$1 - P(x = 13/\lambda = 7.2) = 0.9832$$

Part C

$$\lambda = \underset{\lambda}{\text{argmin}} -ln(L(\lambda)) = -ln(p(\lambda)p(X|\lambda))$$
$$\lambda = \underset{\lambda}{\text{argmin}} -ln(L(\lambda)) = \sum_{i=1}^{n} -ln(p(\lambda)) - ln(p(X|\lambda))$$
$$\lambda = \underset{\lambda}{\text{argmin}} -ln(L(\lambda)) = \sum_{i=1}^{n} -ln(p(\lambda)) - ln(p(X|\lambda))$$
$$p(x|\lambda) = \frac{e^{-\lambda} \lambda^x}{x!}$$
$$-ln(L(\lambda)) = -ln(\frac{\beta^\alpha}{\Gamma(\alpha)}) - ln(\lambda)(\alpha - 1) + \beta\lambda - \sum_{i=1}^{n} -\lambda + x_i ln(\lambda) + ln(x_i!)$$

We take the derivate

$$\frac{\partial ln(L(\lambda))}{\partial \lambda} = -\frac{(\alpha - 1)}{\lambda} + \beta - \sum_{i=1}^{n} -1 + x_i \frac{1}{\lambda} = 0$$
$$\frac{(\alpha - 1 + \sum_{i=1}^{n} x_i)}{\lambda} = \beta + n$$
$$\frac{(\alpha - 1 + \sum_{i=1}^{n} x_i)}{\beta + n} = \hat{\lambda}$$

Then we can see that $\lambda = 5.92$, now we are going to calculate $P(x < 13/\lambda = 5.92) = 1 - P(x = 13/\lambda = 5.92)$

$$P(x = 13/\lambda = 5.92) = \frac{e^{-5.92} 5.92^{13}}{13!}$$
$$P(x = 13/\lambda = 5.92) = 0.004730189$$
$$1 - P(x = 13/\lambda = 5.92) = 0.995269811$$

Question 3

First we have to estimte what is the parameter of the exponential distribution for using the data what we have, for doing this we requiere to estimate the MLE of the distribion

$$\theta = \underset{\theta}{\text{argmin}} -ln(L(\theta)) = -\sum_{i=1}^{n} ln(f(\theta/x_i))$$

$$\theta = \underset{\theta}{\text{argmin}} -ln(L(\theta)) = -\sum_{i=1}^{n} ln(\theta e^{-x_i\theta})$$

$$\theta = \underset{\theta}{\text{argmin}} -ln(L(\theta)) = -\sum_{i=1}^{n} ln(\theta) - x_i\theta$$

$$\frac{\partial ln(L(\theta))}{\partial\theta} = \sum_{i=1}^{n} -\frac{1}{\theta} + x_i = 0$$

$$\frac{n}{\theta} = \sum_{i=1}^{n} x_i$$

$$\frac{n}{\theta} = \sum_{i=1}^{n} x_i$$

$$hat\theta = \frac{n}{\sum_{i=1}^{n} x_i}$$

Now having the estimator for the data what we have, the probability as $1 - e^{-\hat{\theta}x}$, where $\hat{\theta} = \frac{n}{\sum_{i=1}^{n} x_i}$ and x=10.

$$Pr(x < 10) = 1 - e^{-\frac{n}{\sum_{i=1}^{n} x_i}10}$$

Question 4

The Likelihood function in this case is:

$$L(p; x_i) = \prod_{i=1}^{n} \binom{x_i + r - 1}{k} p^r (1-p)^{X_i}$$

$$ln(L(p; x_i)) = \sum_{i=1}^{n} ln(\binom{x_i + r - 1}{k}) + rln(p) + x_i ln(1-p)$$

$$\frac{\partial ln(L(p; x_i)))}{\partial p} = \sum_{i=1}^{n} r\frac{1}{p} - x_i \frac{1}{(1-p)} = 0$$

$$\frac{nr}{p} = \sum_{i=1}^{n} x_i \frac{1}{(1-p)}$$

$$nr - nrp = p\sum_{i=1}^{n} x_i$$

$$p = \frac{nr}{\sum_{i=1}^{n} x_i + nr}$$

$$\hat{p} = \frac{r}{\bar{x} + r}$$

Here x + r is the total of repetitions or n and r the success rate. If we assume that r=10, implies that the total of repetitions if the probability is 1/2 is 20 and x = 10.

Question 5

Section A

$$p(x|\theta) = e^{a(\theta)b(x)+c(x)+d(\theta)}$$
$$ln(p(x|\theta)) = a(\theta)b(x) + c(x) + d(\theta)$$

Section B

$$p(x|\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}$$
$$ln(p(x|\lambda)) = -\lambda + xln(\lambda) - ln(x!)$$
$$a(\theta) = ln(\lambda), b(x) = x, c(x) = -ln(x!), d(\theta) = -\lambda$$

Section C

$$p(x|\theta) = \theta e^{-x\theta}$$
$$ln(p(x|\theta)) = ln(\theta) - x\theta$$
$$a(\theta) = \theta, b(x) = -x, c(x) = 0, d(\theta) = ln(\theta)$$

Question 6

S^2

For getting the expected value of S^2 we could argue using the cochrane theorem that $\frac{(n-1)S^2}{\sigma^2} \backsim \chi^2_{n-1}$ If this is true, we can say that $E[\frac{(n-1)S^2}{\sigma^2}] = n-1$ and $Var[\frac{(n-1)S^2}{\sigma^2}] = 2(n-1)$. Solving this equations we get that $E[S^2] = \sigma^2$ and $Var[S^2] = \frac{2\sigma^4}{n-1}$. This implies that $MSE = \frac{2\sigma^4}{n-1}$, since the estimator is unbiased.

$\hat{\sigma}^2$

For solving this part we can convert $\hat{\sigma^2} = \frac{n-1}{n}S^2$, now if we take the expected value and the variance will get:

$$E[\hat{\sigma^2}] = E[\frac{n-1}{n}S^2] = \frac{n-1}{n}\sigma^2$$

In case of the variance:

$$Var[\hat{\sigma^2}] = Var[\frac{(n-1)}{n}S^2] = \frac{(n-1)^2}{n^2}Var[S^2]$$

But we know the variance of $S^2 = \frac{2\sigma^4}{n-1}$, then :

$$Var[\hat{\sigma^2}] = \cdot\frac{2(n-1)\sigma^4}{n^2}$$

Then the MSE is:

$$MSE = (\frac{n-1}{n}\sigma^2 - \sigma^2)^2 + \frac{2(n-1)\sigma^4}{n^2} = \frac{2n-1}{n^2}\sigma^4$$

Comparation

One way to see which MSE is the greatest is the following, we assume that MSE of S^2 is greater than the σ^2, and compare them trought an inequality, then solve and try to find some inconsistency, that proof the other way around:

$$
\begin{gathered}
\frac{2\sigma^4}{n-1} > \frac{2n-1}{n^2}\sigma^4 \\
\frac{2}{n-1} > \frac{2n-1}{n^2} \\
2n^2 > (2n-1)(n-1) \\
2n^2 > 2n^2 - 3n + 1 \\
n > \frac{1}{3}
\end{gathered}
$$

then this imply for any n>1/3 $MSE_{S^2} > MSE_{\sigma^2}$, since n is a positive integers, implies that the min value is 1, then is true for every possible value relevant for n. This is interesting because we know that σ^2 is a bias

```
In [ ]:
```