Homework 4

Frederick J. Boehm

10/15/2020

Question 1

Q1, part i

$$
\begin{aligned}
D(\mathbb{P}||\mathbb{Q}) &= \sum_{x=0}^{\infty}\left(\frac{e^{-\mu_1}\mu_1^x}{x!}\log\left[\frac{e^{-\mu_1}\mu_1^x}{e^{-\mu_2}\mu_2^x}\right]\right)\\
&= \sum_{x=0}^{\infty}\left(\left[\frac{e^{-\mu_1}\mu_1^x}{x!}\right]\left[-\mu_1+\mu_2+x\log\left(\frac{\mu_1}{\mu_2}\right)\right]\right)\\
&= \sum_{x=0}^{\infty}(\mu_2-\mu_1)\frac{e^{-\mu_1}\mu_1^x}{x!}+\sum_{x=1}^{\infty}\frac{e^{-\mu_1}\mu_1^{x-1}}{(x-1)!}\left(\mu_1\log\frac{\mu_1}{\mu_2}\right)\\
&= \mu_2-\mu_1+\mu_1\log\frac{\mu_1}{\mu_2}
\end{aligned}
$$

Q1, part ii

$$
\begin{aligned}
D(\mathbb{P}||\mathbb{Q}) &= \mathbb{E}_1\left(\log\left[\frac{\mu_1 e^{-\mu_1 x}}{\mu_2 e^{-\mu_2 x}}\right]\right)\\
&= \mathbb{E}_1\left(\log\frac{\mu_1}{\mu_2}-\mu_1 x+\mu_2 x\right)\\
&= \log\frac{\mu_1}{\mu_2}-\mu_1^2+\mu_1\mu_2
\end{aligned}
$$

Question 2

We calculate (where we take expectations with respect to the joint distribution):

$$
\begin{aligned}
D(P_{12}||P_1P_2) &= \mathbb{E}\left(-\log(2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}) + \log(2\pi\sigma_1\sigma_2) - \frac{1}{2(1-\rho^2)}\left[\frac{(X_1-\mu_1)^2}{\sigma_1^2} + \frac{(X_2-\mu_2^2)}{\sigma_2^2} - \frac{2\rho(X_1-\mu_1)(X_2-\mu_2)}{\sigma_1\sigma_2}\right] + \frac{(X_1-\mu_1)^2}{\sigma_1^2} + \frac{(X_2-\mu_2)^2}{\sigma_2^2}\right) \\
&= -\frac{1}{2}\log(1-\rho^2) - \frac{1}{2(1-\rho^2)}\left[1+1-\frac{2\rho}{\sigma_1\sigma_2}\rho\sigma_1\sigma_2\right] + \frac{1}{2} + \frac{1}{2} \\
&= -\frac{1}{2}\log(1-\rho^2) - \frac{1}{2(1-\rho^2)}\left[2-2\rho^2\right] + 1 \\
&= -\frac{1}{2}\log(1-\rho^2)
\end{aligned}
$$

Question 3

3a

Let $w_{ij} = P_{12}(X_1 = i, X_2 = j)$. Then $\sum_i w_{ij} = q_j = P_2(X_2 = j)$ and $\sum_j w_{ij} = p_i = P_1(X_1 = i)$.

Then,

$$
\begin{aligned}
I(X_1, X_2) &= \sum_{i,j} w_{ij} \log \frac{w_{ij}}{p_i q_j} \\
&= \sum_{i,j} w_{ij} \log w_{ij} - \sum_{i,j} w_{ij} \log p_i - \sum_{i,j} w_{ij} \log q_j \\
&= \sum_{i,j} w_{ij} \log w_{ij} - \sum_{i} p_i \log p_i - \sum_{j} q_j \log q_j \\
&= -H((X_1, X_2)) + H(X_1) + H(X_2)
\end{aligned}
$$

3b

$$
\begin{aligned}
I(X_1, X_2) &= \sum_{i,j} w_{ij} \log \frac{w_{ij}}{p_i q_j} \\
&= \sum_{i,j} w_{ij} \log \frac{u_{i|j} q_j}{p_i q_j} \\
&= \sum_{i,j} w_{ij} \log \frac{u_{i|j}}{p_i} \\
&= \sum_{i,j} w_{ij} \log u_{i|j} - \sum_{i,j} w_{ij} \log p_i \\
&= \sum_{i,j} w_{ij} \log u_{i|j} - \sum_{i} p_i \log p_i \\
&= H(X) - H(X|Y)
\end{aligned}
$$

Question 4

Q4, Part i

First, define the estimators of the probabilities:

$$
\begin{aligned}
\hat{p}_n^X(x) &= \frac{|\{i : X_i = x\}|}{n} \\
\hat{p}_n^Y(y) &= \frac{|\{i : Y_i = y\}|}{n}
\end{aligned}
$$

$$
\hat{p}_n^{X,Y}(x, y) = \frac{|\{i : X_i = x, Y_i = y\}|}{n}
$$

Then, define:

$$
\hat{I}_n(X, Y) = \hat{H}_n(X) + \hat{H}_n(Y) - \hat{H}_n(X, Y)
$$

$$
\hat{H}(X, Y) = -\sum_{x,y} \hat{p}_n^{X,Y}(x, y) \log \hat{p}_n^{X,Y}(x, y)
$$

Q4, Part ii

We follow the steps in the proof of Theorem 2.15.

$$
\begin{aligned}
&|\hat{I}_n((x_1,y_1),\dots,(x_i,y_i),\dots,(x_n,y_n)) - \hat{I}_n((x_1,y_1),\dots,(x_i',y_i'),\dots,(x_n,y_n))| \le \\
&\le 2\sup|\frac{j+1}{n}\log\frac{j+1}{n} - \frac{j}{n}\log\frac{j}{n} + \frac{k+1}{n}\log\frac{k+1}{n} - \frac{k}{n}\log\frac{k}{n} - \frac{(j+1)(k+1)}{n^2}\log\frac{(j+1)(k+1)}{n^2} + \frac{jk}{n^2}\log\frac{jk}{n^2}| \\
&\le 4\sup|\frac{j+1}{n}\log\frac{j+1}{n} - \frac{j}{n}\log\frac{j}{n}| + 2\sup|-\frac{(j+1)(k+1)}{n^2}\log\frac{(j+1)(k+1)}{n^2} + \frac{jk}{n^2}\log\frac{jk}{n^2}| \\
&\le 4\frac{\log n}{n} + 2\frac{\log n}{n^2}
\end{aligned}
$$

Setting this value equal - call it c - to c_i for all i, we get:

$$\mathbb{P}\left(|\hat{I}_n - I| \ge t\right) \le 2\exp\left(-\frac{2t^2}{nc^2}\right)$$

Question 5

Q5, part i

First, observe that we can Taylor expand $\log f(x;\theta+\epsilon)$.

$$\log f(x;\theta+\epsilon) = \log f(x;\theta) + \epsilon\frac{\partial}{\partial\theta}\log f(x;\theta) + \frac{\epsilon^2}{2}\frac{\partial^2}{\partial\theta^2}\log f(x;\theta) + \delta$$

We take the expectation (with respect to f_θ of the KL divergence:

$$
\begin{aligned}
\int [f(x;\theta)\log f(x;\theta) - f(x;\theta)\log f(x;\theta+\epsilon)]\,dx &\approx \int \epsilon\frac{\partial}{\partial\theta}\log f(x;\theta) + \frac{\epsilon^2}{2}\frac{\partial^2}{\partial\theta^2}\log f(x;\theta) \\
&= \int \frac{\epsilon^2}{2}\frac{\partial^2}{\partial\theta^2}\log f(x;\theta) \\
&= \frac{\epsilon^2}{2}\mathcal{I}(\theta)
\end{aligned}
$$

The second line follows from the first because the mean of the score is equal to zero.

Q5, part ii

This follows from above.

Q5, part iii

Here, we expand both terms and observe that the ϵ^3 terms cancel.

$$\int$$

Question 6

6i

$$
\begin{aligned}
l(\theta) = \log L(\theta) = \log f_\theta(x) &= x \log \theta + (1-x)\log(1-\theta) \\
&\implies \frac{d}{d\theta} l = \frac{x}{\theta} - \frac{1-x}{1-\theta} = \frac{x-\theta}{\theta(1-\theta)} \\
&\implies \mathbb{E}\left(\frac{dl}{d\theta}\right)^2 = \mathbb{E}\left(\frac{(x-\theta)^2}{\theta^2(1-\theta)^2}\right) \\
&= \frac{Var(X)}{\theta^2(1-\theta)^2} \\
&\implies \mathcal{I}(\theta) = \frac{1}{(1-\theta)\theta}
\end{aligned}
$$

6ii

We calculate partial derivatives:

$$
\begin{aligned}
\frac{\partial l}{\partial \mu} &= \frac{X-\mu}{\sigma^2} \\
\frac{\partial l}{\partial \sigma} &= \frac{-1}{\sigma} + \frac{(X-\mu)^2}{\sigma^3} \\
\frac{\partial^2 l}{\partial \mu^2} &= \frac{-1}{\sigma^2} \\
\frac{\partial^2 l}{\partial \mu \partial \sigma} &= \frac{-2(X-\mu)}{\sigma^3} \\
\frac{\partial^2 l}{\partial \sigma^2} &= \frac{1}{\sigma^2} - \frac{3(X-\mu)^2}{\sigma^4}
\end{aligned}
$$

Taking (negative) expectations of the three second partial derivatives, we see that we have:

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

Question 7

7, part i

First, define $Y_i = f(X_i) - \mathbb{E}(f(X_i))$.

Then compute:

$$\begin{aligned}
(n-1)\mathbb{E}V_n^2 &= \sum_{i=1}^{n} \mathbb{E}\left(Y_i - \bar{Y}\right)^2 \\
&= \sum_{i=1}^{n} \mathbb{E}\left(Y_i^2 - 2\bar{Y}Y_i + \bar{Y}^2\right) \\
&= \sum_{i} \mathbb{E}\left(Y_i^2\right) - \frac{2}{n}\sum_{i} \mathbb{E}Y_i^2 + n\mathbb{E}\bar{Y}^2 - 2\sum_{i}\sum_{j\neq i} \mathbb{E}(Y_iY_j) \\
&= (n-2)\mathbb{E}Y_1^2 + \mathbb{E}Y_1^2 \\
&= (n-1)\mathbb{E}(Y_1^2) \\
&= (n-1)Var(f(X))
\end{aligned}$$

For the second conclusion, we want to use the strong law of large numbers. We need to verify that the assumptions of the theorem are met. We note that f is bounded, so the SLLN applies.

7, part ii

We need to show that V_n^2 has bounded differences.

First, observe that it's possible for n values to all have the same value, say, b.

The maximum difference is achieved when this occurs and the replaced value is a.

So, compare the sample variances for the realized values $(a, a, \ldots, a)$ vs. $(b, a, a, \ldots, a)$. The former has sample variance equal to zero, while the latter has sample mean $\bar{y} = \frac{(n-1)a+b}{n}$ and sample variance:

$$
\begin{aligned}
(n-1)s^2 &= \left((\bar{y}-b)^2 + (n-1)(\bar{y}-a)^2\right) \\
&= (n-1)\frac{(b-a)^2}{n^2} + \frac{(n-1)^2}{n^2}(a-b)^2 \\
&= \frac{1}{n^2}(a-b)^2\left(n-1+(n-1)^2\right) = \frac{(a-b)^2 n(n-1)}{n^2} \\
&\implies s^2 = \frac{(a-b)^2}{n}
\end{aligned}
$$

Thus, $c = \frac{(a-b)^2}{n}$.

7, part iii

$$\mathbb{P}(|V_n^2 - \mathbb{E}V_n^2| \geq t) \leq 2\exp\left(-\frac{2nt^2}{(a-b)^4}\right)$$

Question 8

8, part i

$$|G(X) - G(X')| = |\sup_{f\in\mathcal{F}} |\bar{Y} - \mathbb{E}\bar{Y}| - \sup_{g\in\mathcal{F}} |\bar{W} - \mathbb{E}\bar{W}|| \leq |\sup_{f\in\mathcal{F}} |Y - \mathbb{E}\bar{Y}|| \leq 2K$$

8, part ii

We showed in part i the bounded differences hypothesis for Theorem 2.14. Then

$$\mathbb{P}\left(|Y - \mathbb{E}Y| \geq t\right) \leq 2\exp\left(-\frac{2t^2}{\sum_{i=1}^n c_i^2}\right)$$