Stats 300A HW 2 Solutions

Stephen Bates

October 15, 2018

Problem 1: Tweedie's formula

Part a

$$\begin{aligned} p(x) &= \int \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-1}{2\sigma^2}(x-\theta)^2} Q(d\theta) \\ \implies \frac{d}{dx}\log p(x) &= \frac{\frac{1}{\sigma^2}\int \frac{1}{\sqrt{2\pi\sigma^2}}(\theta - x) e^{\frac{-1}{2\sigma^2}(x-\theta)^2} Q(d\theta)}{\int \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-1}{2\sigma^2}(x-\theta)^2} Q(d\theta)} \\ &= \frac{1}{\sigma^2}(\hat{\theta}_B - x) \end{aligned}$$

Part b

In the general case we have:

$$\begin{aligned} p(x) &= \int \frac{1}{Z(\theta)} e^{\langle \theta, x\rangle} h(x) Q(d\theta) \\ \implies \frac{\partial}{\partial x_i}\log p(x) &= \frac{1}{\int \frac{1}{Z(\theta)} e^{\langle \theta, x\rangle} h(x) Q(d\theta)} \left[\int \frac{1}{Z(\theta)} \theta_i e^{\langle \theta, x\rangle} h(x) Q(d\theta) + \int \frac{1}{Z(\theta)} e^{\langle \theta, x\rangle} (\frac{\partial}{\partial x_i} h)(x) Q(d\theta)\right] \\ &= \hat{\theta}_{B,i} + \frac{\int \frac{1}{Z(\theta)} e^{\langle \theta, x\rangle} (\frac{\partial}{\partial x_i} h)(x) Q(d\theta)}{\int \frac{1}{Z(\theta)} e^{\langle \theta, x\rangle} h(x) Q(d\theta)} \\ &= \hat{\theta}_{B,i} + \frac{\partial}{\partial x_i}\log(h(x)) \end{aligned}$$

Note in the Gaussian case $h(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-1}{2\sigma^2}x^2}$ and the natural parameter is $\theta = \frac{\mu}{\sigma^2}$, so this agrees with the explicit calculation in part a.

Problem 2: Estimating a single bit

Part a

The posterior distribution of θ given the data is

$$\begin{aligned} p(\theta = 1 | X = x) &= \frac{qP(X = x|\theta = 1)}{qP(X = x|\theta = 1) + \bar{q}P(X = x|\theta = 0)} \\ &= \frac{qp_1(x)}{qp_1(x) + \bar{q}p_2(x)}. \end{aligned}$$

The the Bayes optimal estimator minimizes

$$\mathbb{E}[L(\hat{\theta}(x),\theta)|X=x] = c_2\frac{qp_1(x)}{qp_1(x)+\bar{q}p_0(x)}P(\hat{\theta}(x)=0|x) + c_1\frac{\bar{q}p_0(x)}{qp_1(x)+\bar{q}p_0(x)}P(\hat{\theta}(x)=1|x)$$

for each x.

The Bayes estimator is then

$$\hat{\theta}_B(x) = \begin{cases} 0 \text{ when } c_0 q p_1(x) \le c_1 \bar{q} p_0(x) \\ 1 \text{ otherwise .} \end{cases}$$

Part b

Let $R_0 = \{x \in \mathcal{X} : \hat{\theta}_B(x) = 0\}$ and $R_1 = \{x \in \mathcal{X} : \hat{\theta}_B(x) = 1\}$.

$$\begin{aligned} \mathbb{E}[L(\hat{\theta},\theta)] &= c_1 P(\theta = 0, x \in R_1) + c_2 P(\theta = 1, x \in R_0) \\ &= c_1\bar{q}\int_{R_1} p_0(x) + c_2 q \int_{R_0} p_1(x) \end{aligned}$$

Part c

Notice that when $c_1 = c_2$ and $q = \bar{q}$ we have that $R_0 = \{x \in \mathcal{X} : p_0(x) \ge p_1(x)\}$ and $R_1 = R_0^{\complement}$.

$$\begin{aligned} 1 - \|P_0 - P_1\|_{tv} &= 1 - \frac{1}{2}\int_{\mathcal{X}} |p_0(x) - p_1(x)| dx \\ &= 1 - \frac{1}{2}\int_{R_0} p_0(x) - p_1(x) - \frac{1}{2}\int_{R_1} p_1(x) - p_0(x) \\ &= \frac{1}{2}\int_{R_1\cup R_0} p_0(x) + \frac{1}{2}\int_{R_1\cup R_0} p_1(x) - \frac{1}{2}\int_{R_0} p_0(x) - p_1(x) - \frac{1}{2}\int_{R_1} p_1(x) - p_0(x) \\ &= \int_{R_1} p_0(x) + \int_{R_0} p_1(x) \end{aligned}$$

The result follows after diving this expression by 2.

Part d

Consider $\mathcal{X} = \{\pm 1\}$. Let $P_0(X=1) = 0$ and $P_1(X_1 = .85)$. With $q = .5$ we see that the Bayes estimator is $\hat{\theta}_B(x) = x$ and the risk is $\mathbb{E}\mathrm{I}_{\hat{\theta}\neq\theta} = .075$ With $q = .9$, we see that the Bayes estimator is $\hat{\theta}_B(x) = 1$ and the Bayes risk is .1.

Problem 3: Stochastic Block Model

Part a

Note that the problem is symmetric in the group labelings, so we want a lost function that does not change when changing the ground truth θ to $-\theta$. Simply counting the number of correctly labeled data points would not suffice. The proposed loss function equivalent to counting the number of correctly labeled points after accounting for this ambiguity.

Part b

Let $\hat{\theta}_0(x)$ be a Bayes optimal estimator, i.e.

$$R_B(\hat{\theta}_0(G), Q) = \inf_A R_B(A, Q) = R_B(Q).$$

For a permutation π, denote the inverse permutaiton by $\bar{\pi}$. We can construct a symmetric version of this estimator by taking a uniformly random permutation and then applying this estimator to the permuted data: $\hat{\theta}_1(G) = \hat{\theta}_0(G^\pi)^{\bar{\pi}}$ where π is a permutation chosen uniformly at random. We will now show that $R_B(\hat{\theta}_1(G), Q) = R_B(\hat{\theta}_0(G), Q)$, which implies that $\hat{\theta}_1$ is a Bayes estimator.

Fix any permutation π and consider the risk of the estimator $\hat{\theta}_\pi(G) = \hat{\theta}_0(G^\pi)^{\bar{\pi}}$:

$$\begin{aligned}
R_B(\hat{\theta}_\pi, Q) &= \int_\theta \int_G L(\hat{\theta}_0(G^\pi)^{\bar{\pi}}, \theta) dP_\theta DQ \\
&= \int_\theta \int_G L(\hat{\theta}_0(G^\pi), \theta^\pi) dP_\theta DQ \text{ using the symmetry of } L \\
&= \int_\theta \int_G L(\hat{\theta}_0(G), \theta) dP_\theta DQ \text{ since } (G^\pi, \theta^\pi) \overset{d}{=} (G, \theta) \\
&= R_B(\hat{\theta}_0, Q) \\
&= R_B(Q).
\end{aligned}$$

Since $\hat{\theta}_1$ is a mixture of the estimators $\hat{\theta}_\pi$, this implies that $R_B(\hat{\theta}_1, Q) = R_B(Q)$ and hence $\hat{\theta}_1$ is a Bayes estimator.

To show the desired symmetry, let π be a fixed permutation and let α be a uniformly random permutation. Then

$$\begin{aligned}
(\theta^\pi, \hat{\theta}_1(G^\pi)) &= (\theta^\pi, \hat{\theta}_0(G^{\pi\alpha})^{\bar{\alpha}}) \\
&= (\theta^\pi, \hat{\theta}_0(G^{\alpha'})^{\bar{\alpha'}\pi}) \\
&\overset{d}{=} (\theta, \hat{\theta}_1(G)).
\end{aligned}$$

Part c

Let $\hat{\theta}$ be the symmetric Bayes estimator from the previous part.

$$\begin{aligned}
1 - \frac{1}{n^2}\mathbb{E}\langle\hat{\theta}, \theta\rangle^2 &= 1 - \frac{1}{n^2}\sum_{i \neq j} \mathbb{E}[\hat{\theta}_i\hat{\theta}_j\theta_i\theta_j] - 1/n \\
&= (1 - \frac{1}{n}) - \frac{n(n-1)}{n^2}\mathbb{E}\sum_{i \neq j}\mathbb{E}[\hat{\theta}_1\hat{\theta}_2\theta_1\theta_2] \quad \text{by the symmetry we proved in (b)} \\
&= (1 - \frac{1}{n})(1 - \mathbb{E}\sum_{i \neq j}\mathbb{E}[\hat{\theta}_1\hat{\theta}_2\theta_1\theta_2])
\end{aligned}$$

Part d

We continue on from the previous part:

$$\begin{aligned}
(1 - \frac{1}{n})(1 - \mathbb{E}\sum_{i \neq j}\mathbb{E}[\hat{\theta}_1\hat{\theta}_2\theta_1\theta_2]) &\geq (1 - \frac{1}{n}) \inf_{\hat{\theta}_1(G)\hat{\theta}_2(G)} (1 - \mathbb{E}\sum_{i \neq j}\mathbb{E}[\hat{\theta}_1(G)\hat{\theta}_2(G)\theta_1\theta_2]) \text{ (since the lhs is a special case of the rhs)} \\
&\geq (1 - \frac{1}{n}) \inf_{\hat{\theta}_1(G, \theta_{\sim 1})\hat{\theta}_2(G, \theta_{\sim 1})} (1 - \mathbb{E}\sum_{i \neq j}\mathbb{E}[\hat{\theta}_1(G, \theta_{\sim 1})\hat{\theta}_2(G, \theta_{\sim 1})\theta_1\theta_2]) \text{ (inf over larger set)}
\end{aligned}$$

Lastly, notice that $\hat{\theta}_1(G, \theta_{\sim 1})\hat{\theta}_2(G, \theta_{\sim 1})\theta_2$ can be combined into a single function of G and $\theta_{\sim 1}$, which we can write as $\hat{\theta}_1(G; \theta_{\sim 1})$.

Part e

$$\begin{aligned} P(G, \theta_{\sim 1})_{\theta_1} &= P(G|\theta_{\sim 1}\theta_1)P(\theta_{\sim 1}|\theta_1) \\ &\propto P(G|\theta_{\sim 1}\theta_1) \\ &\propto \prod_{(i,j)\in E} (a/n)^{I_{\theta_1=\theta_j}} (b/n)^{I_{\theta_1\neq\theta_j}} \prod_{(1,j)\notin E} (1-a/n)^{I_{\theta_1=\theta_j}} (1-b/n)^{I_{\theta_1\neq\theta_j}} \\ &\propto (a/n)^{X_+ I_{\theta_1=1} + X_- I_{\theta_1=-1}} (b/n)^{X_+ I_{\theta_1=-1} + X_- I_{\theta_1=1}} \\ &\quad (1-a/n)^{(N_+-X_+)I_{\theta_1=1} + (N_- - X_-)I_{\theta_1=-1}} (1-b/n)^{(N_+-X_+)I_{\theta_1=-1} + (N_- - X_-)I_{\theta_1=1}} \end{aligned}$$

The likelihood is only a function of (N_+, N_-, X_+, X_-) as well as the parameter θ_1, so by the Fisher-Neyman criterion, these are the sufficient statistics.

Part f

We have now reduced it to a one-bit estimation problem. We can use problem 2c to get an exact expression for the Bayes risk, so it suffices to find an expression for the TV distance $\|P_{\theta=1}(Z_+, Z_-) - P_{\theta=-1}(Z_+, Z_-)\|_{TV}$

$$\begin{aligned} \|P_{\theta=1}(Z_+, Z_-) - P_{\theta=-1}(Z_+, Z_-)\|_{TV} &= \frac{1}{2}\int |P_1(Z_+, Z_-) - P_0(Z_+, Z_-)| \\ &= \frac{1}{2}\sum_{x=0}^{\infty}\sum_{y=0}^{\infty} \frac{e^{-a/2-b/2}}{x!y!2^{x+y}} |a^x b^y - a^y b^x| \end{aligned}$$

This expression is exact, but challenging to work with, so we will reduce to a simpler problem. Consider estimators that are a function only of Z_+. Call the Bayes risk in this new problem $R_B^{restricted}(a, b)$. Since this is a restriction of the problem above, we have that $R_B^{restricted}(a, b) \geq R_B^s(a, b)$, so it suffices to show that $R_B^{restricted}(a, b) \to 0$.

From problem 2c, we have that $\|P_0 - P_1\|_{TV} = 1/2 \int |p_1(x) - p_0(x)|$ and using since $\int p_1(x) - p_0(x) = 0$ we have that $\|P_0 - P_1\|_{TV} = \int_{R_1} p_1(x) - p_0(x) = P_1(R_1) - P_0(R_1)$. Since $R_1 = \{x : p_1(x) > p_0(x)\}$, we have that for any A, $P_1(R_1) - P_0(R_1) \geq P_1(A) - P_0(A)$, so we arrive at a well-known property of total variation distance:

$$\|P_0 - P_1\|_{TV} \geq P_1(A) - P_0(A).$$

We will consider the set $A = [\frac{a+b}{4}, \infty)$. Notice that $P_{\theta=1}(Z_+ \in A) \geq 1 - 8a/(a-b)^2$ by Chebyshev's inequality. We similarly have $P_{\theta=-1}(Z_+ \in A) \leq 8b/(a-b)^2$. Together with the above, this implies that if $a - b \to \infty$ such that $b/(a-b), a/(a-b)$ are bounded, then $\|P_{\theta=1}(Z_+) - P_{\theta=-1}(Z_+)\|_{TV} \to 1$ and using 2c this implies that $R_B^{restricted}(a, b) \to 0$, as desired.