# STAT6201: Theoretical Statistics I

Satvik Saha

## Homework 4

1. Let $P$ be supported on $\mathbb{N}$, and let $Q_\lambda$ denote the Poisson($\lambda$) distribution. Further let $p, q$ be the probability mass functions of $P, Q_\lambda$ respectively. Then,

$$
\begin{aligned}
\mathrm{KL}(P, Q_\lambda) &= \sum_{n \in \mathbb{N}} p(n) \log\left(\frac{p(n)}{q_\lambda(n)}\right) \\
&= \sum_{n \in \mathbb{N}} p(n) \log\left(\frac{p(n)\, n!\, e^\lambda}{\lambda^n}\right) \\
&= \sum_{n \in \mathbb{N}} p(n)\, \log(p(n)\, n!) + \lambda \cancelto{1}{\sum_{n \in \mathbb{N}} p(n)} - \log(\lambda) \sum_{n \in \mathbb{N}} n\, p(n) \\
&= C_P + \lambda - \mu \log(\lambda).
\end{aligned}
$$

   It follows that $\mathrm{KL}(P, Q_\lambda)$ is minimized when $\lambda - \mu \log(\lambda)$ is minimized. Differentiating, we see that $\lambda^* = \mu$ is indeed the unique minimizer.

   *Remark:* The map $x \mapsto x - \mu \log x$ diverges to $\infty$ when $x \to 0$ as well as when $x \to \infty$.

2. (a) Let $X_1, \dots, X_n \overset{iid}{\sim} N(0, \theta^{-1})$, so

$$
f(\boldsymbol{x} \mid \theta) = (2\pi)^{-n/2} \theta^{n/2} \exp\left(-\frac{1}{2}\theta \sum_{i=1}^n x_i^2\right)
$$

   Let $\mathscr{G}$ be the family of gamma distributions with densities

$$
\pi(\theta \mid \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha - 1} \exp(-\beta\theta),
$$

   for $\alpha, \beta > 0$. Then, observe that

$$
\pi(\theta \mid \boldsymbol{x}) \propto \theta^{\alpha + n/2 - 1} \exp\left(-\left[\beta + \frac{1}{2}\sum_{i=1}^n x_i^2\right]\theta\right).
$$

   Thus, $\theta \mid \boldsymbol{x} \sim \pi(\theta \mid \alpha + n/2, \beta + \sum_i x_i^2/2) \in \mathscr{G}$. The posterior mean is simply

$$
\mathbb{E}[\theta \mid \boldsymbol{x}] = \frac{\alpha + n/2}{\beta + \sum_i x_i^2/2}.
$$

   (b) Let $X_1, \dots, X_n \overset{iid}{\sim} N(0, \theta^2)$ for $\theta \in (0, \infty)$, so

$$
f(\boldsymbol{x} \mid \theta) = (2\pi\theta^2)^{-n/2} \exp\left(-\frac{1}{2\theta^2}\sum_{i=1}^n x_i^2\right).
$$

Then,

$$\frac{\partial}{\partial\theta}\log f(\boldsymbol{x}\mid\theta) = -\frac{n}{\theta} + \frac{1}{\theta^3}\sum_{i=1}^n x_i^2,$$

$$\frac{\partial^2}{\partial\theta^2}\log f(\boldsymbol{x}\mid\theta) = \frac{n}{\theta^2} - \frac{3}{\theta^4}\sum_{i=1}^n x_i^2,$$

$$-\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2}\log f(\boldsymbol{x}\mid\theta)\right] = -\frac{n}{\theta^2} + \frac{3}{\theta^4}\cdot n\theta^2 = \frac{2n}{\theta^2}.$$

Thus, the Jeffrey's prior is given by $\pi(\theta) \propto \sqrt{I(\theta)} \propto \theta^{-1}$. Note that with this,

$$\pi(\theta\mid\boldsymbol{x}) \propto \theta^{-n-1}\exp\left(-\frac{1}{2}\sum_{i=1}^n x_i^2/\theta^2\right),$$

so

$$\pi(\theta^{-2}\mid\boldsymbol{x}) \propto \theta^3\cdot(\theta^{-2})^{(n+1)/2}\exp\left(-\frac{1}{2}\sum_{i=1}^n x_i^2\theta^{-2}\right),$$

from which $\theta^{-2}\mid\boldsymbol{x} \sim \text{Gamma}(n/2, \sum_i x_i^2/2)$. Thus,

$$\mathbb{E}[\theta\mid\boldsymbol{x}] = \mathbb{E}[(\theta^{-2})^{-1/2}\mid\boldsymbol{x}] = \frac{\Gamma(n/2-1/2)}{\Gamma(n/2)}\cdot\left(\frac{1}{2}\sum_{i=1}^n x_i^2\right)^{1/2}.$$

3. Let $Z \sim N(0,1)$, and let $h \in C^1(\mathbb{R})$ with $\mathbb{E}[|h'(Z)|] < \infty$. Recall the relation $\phi'(z) = -z\,\phi(z)$, where $\phi$ is the density of the standard normal distribution. Then, integration by parts yields

$$\int_{\mathbb{R}} h'(z)\phi(z)\,dz = \cancelto{0}{h(z)\phi(z)\Big|_{-\infty}^{+\infty}} - \int_{\mathbb{R}} h(z)\phi'(z)\,dz = \int_{\mathbb{R}} zh(z)\,\phi(z)\,dz.$$

This is precisely $\mathbb{E}[h'(Z)] = \mathbb{E}[Zh(Z)]$.

*Remark:* It is easily checked that

$$\phi(z) = \frac{1}{\sqrt{2\pi}}e^{-z^2/2}, \qquad \phi'(z) = \frac{1}{\sqrt{2\pi}}e^{-z^2/2}\cdot(-z) = -z\,\phi(z).$$

*Remark:* To verify that the boundary term does indeed vanish, we show that $h(z)\phi(z)\to 0$ as $z\to\infty$; the other case will follow by symmetry. Fix $z_0 > 0$, and note that for $z > z_0$, we have $\phi$ decreasing on $(z_0, z)$, so

$$\begin{aligned}
|h(z)\phi(z)| &\le |h(z_0)|\phi(z) + |h(z)-h(z_0)|\phi(z) \\
&= |h(z_0)|\phi(z) + \left|\int_{z_0}^z h'(t)\phi(z)\,dt\right| \\
&\le |h(z_0)|\phi(z) + \int_{z_0}^z |h'(t)|\phi(t)\,dt \\
&\le |h(z_0)|\phi(z) + \mathbb{E}[|h'(Z)|\,\mathbf{1}_{(z_0,\infty)}(Z)].
\end{aligned}$$

Now, using $\phi(z)\to 0$ as $z\to\infty$,

$$\limsup_{z\to\infty}|h(z)\phi(z)| \le \mathbb{E}[|h'(Z)|\,\mathbf{1}_{(z_0,\infty)}(Z)]$$

for all $z_0 > 0$. Thus, $\mathbb{E}[|h'(Z)|] < \infty$ along with the Dominated Convergence Theorem guarantees that

$$\limsup_{z\to\infty}|h(z)\phi(z)| \le \lim_{z_0\to\infty}\mathbb{E}[|h'(Z)|\,\mathbf{1}_{(z_0,\infty)}(Z)] = 0.$$

4. Let $X \sim N_n(\theta, \sigma^2 \boldsymbol{I}_n)$, and let $h \in C^1(\mathbb{R}^n, \mathbb{R}^n)$ such that $\mathbb{E}[\|\nabla h\|_2] < \infty$.

(a) Write

$$\begin{aligned}\mathbb{E}[\|X - h(X) - \theta\|_2^2] &= \mathbb{E}[\|X - \theta\|_2^2] + \mathbb{E}[\|h(X)\|_2^2] - 2\mathbb{E}[h(X) \cdot (X - \theta)] \\ &= n\sigma^2 + \mathbb{E}[\|h(X)\|_2^2] - 2\mathbb{E}[h(X) \cdot (X - \theta)].\end{aligned}$$

Set $Z = (X - \theta)/\sigma \sim N_n(0, \boldsymbol{I}_n)$, and $g(z) = h(\sigma z + \theta)$, so

$$h(X) \cdot (X - \theta) = \sigma h(\sigma Z + \theta) \cdot Z = \sigma g(Z) \cdot Z.$$

Notate $Z_{-i}$ as the vector $(Z_1, \ldots, Z_n)$ with the $i$-th variable removed. Now,

$$\begin{aligned}\mathbb{E}[h(X) \cdot (X - \theta)] &= \sigma \mathbb{E}[g(Z) \cdot Z] \\ &= \sigma \sum_{i=1}^{n} \mathbb{E}[g_i(Z) Z_i] \\ &= \sigma \sum_{i=1}^{n} \mathbb{E}[\mathbb{E}[g_i(Z) Z_i \mid Z_{-i}]] \\ &= \sigma \sum_{i=1}^{n} \mathbb{E}\left[\frac{\partial g_i(Z)}{\partial z_i}\right] && \text{(Stein's method)} \\ &= \sigma^2 \sum_{i=1}^{n} \mathbb{E}\left[\frac{\partial h_i(X)}{\partial x_i}\right].\end{aligned}$$

Thus, we have shown that

$$\mathbb{E}[\|X - h(X) - \theta\|_2^2] = n\sigma^2 + \mathbb{E}[\|h(X)\|_2^2] - 2\sigma^2 \sum_{i=1}^{n} \mathbb{E}\left[\frac{\partial h_i(X)}{\partial x_i}\right].$$

(b) Let $n > 2$. Setting $\delta_{JS}(X) = X - h(X)$ with

$$h(x) = \frac{(n-2)\sigma^2}{\|x\|_2^2} x,$$

the previous part guarantees that we need only show that for all $\theta \in \mathbb{R}^n$,

$$\mathbb{E}[\|h(X)\|_2^2] - 2\sigma^2 \sum_{i=1}^{n} \mathbb{E}\left[\frac{\partial h_i(X)}{\partial x_i}\right] < 0.$$

Now,

$$\mathbb{E}[\|h(X)\|_2^2] = (n-2)^2 \sigma^4 \mathbb{E}\left[\frac{1}{\|X\|_2^2}\right].$$

Furthermore,

$$\frac{1}{(n-2)\sigma^2} \frac{\partial h_i(x)}{\partial x_i} = \frac{\|x\|_2^2 - x_i \cdot (2x_i)}{\|x\|_2^4} = \frac{1}{\|x\|_2^2} - \frac{2x_i^2}{\|x\|_2^4},$$

so

$$\sum_{i=1}^{n} \frac{\partial h_i(x)}{\partial x_i} = (n-2)\sigma^2 \left[\frac{n}{\|x\|_2^2} - \frac{2\|x\|_2^2}{\|x\|_2^4}\right] = (n-2)^2 \sigma^2 \frac{1}{\|x\|_2^2}.$$

Putting these together, we have

$$\mathbb{E}[\|h(X)\|_2^2] - 2\sigma^2 \sum_{i=1}^n \mathbb{E}\left[\frac{\partial h_i(X)}{\partial x_i}\right] = -(n-2)^2\sigma^4\,\mathbb{E}\left[\frac{1}{\|X\|_2^2}\right] < 0.$$

From this,

$$\mathbb{E}\,\|\delta_{JS}(X) - \theta\|_2^2 = \mathbb{E}\,\|X - h(X) - \theta\|_2^2 < \mathbb{E}\,\|X - \theta\|_2^2$$

for all $\theta \in \mathbb{R}^n$, whence the estimator $X$ for $\theta$ is inadmissible.

*Remark:* It ought to be clear that $h \in C^1$ with $\mathbb{E}\,\|\nabla h(X)\|_2 < \infty$, given that

$$\frac{1}{(n-2)\sigma^2}\frac{\partial h_i(x)}{\partial x_j} = \frac{\delta_{ij}}{\|x\|_2^2} - \frac{2x_i x_j}{\|x\|_2^4},$$

and inverse Gaussians have finite moments.

5. Let $X \mid \theta \sim N_n(\theta, \sigma^2 \boldsymbol{I}_n)$, and $\theta \sim N_n(0, \tau^2 \boldsymbol{I}_n)$ for $\sigma > 0$ known, $\tau > 0$ unknown. Consider the Bayes estimator under squared error loss (in the case $\tau$ is known) for $\theta$,

$$\delta_\pi(X) = \frac{\tau^2}{\tau^2 + \sigma^2}\,X$$

Denote $\alpha = \tau^2/(\tau^2 + \sigma^2)$. We can calculate the posterior

$$\begin{aligned}
\pi(\theta \mid x, \tau^2) &\propto (2\pi\sigma^2)^{n/2}(2\pi\tau^2)^{n/2}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - \theta_i)^2 - \frac{1}{2\tau^2}\sum_{i=1}^n \theta_i^2\right) \\
&\propto \prod_{i=1}^n \exp\left(-\frac{1}{2}\left[\frac{1}{\sigma^2} + \frac{1}{\tau^2}\right]\theta_i^2 + \frac{1}{\sigma^2}x_i\theta_i\right) \\
&\sim N_n(\tau^2 X/(\tau^2 + \sigma^2), \sigma^2\tau^2\boldsymbol{I}_n/(\tau^2 + \sigma^2)) \\
&\sim N_n(\alpha X, \alpha\sigma^2\boldsymbol{I}_n).
\end{aligned}$$

Thus, the marginal

$$\begin{aligned}
f(x \mid \tau^2) &= \frac{f(x \mid \theta)\pi(\theta \mid \tau^2)}{\pi(\theta \mid x, \tau^2)} \\
&= \frac{(\alpha\sigma^2)^{n/2}}{(2\pi\tau^2\sigma^2)^{n/2}}\prod_{i=1}^n \exp\left(-\frac{1}{2\sigma^2}(x_i - \theta_i)^2 - \frac{1}{2\tau^2}\theta_i^2 + \frac{1}{2\alpha\sigma^2}(\theta_i - \alpha x_i)^2\right) \\
&= \frac{1}{(2\pi\tau^2/\alpha)^{n/2}}\prod_{i=1}^n \exp\left(-\frac{(1-\alpha)}{2\sigma^2}x_i^2\right) \\
&\sim N_n(0, (\tau^2 + \sigma^2)\boldsymbol{I}_n).
\end{aligned}$$

(a) Note that $(\tau^2 + \sigma^2)\,\|X\|_2^2 \mid \tau^2 \sim \chi_n^2 \sim \text{Gamma}(n/2, 1/2)$, whence

$$\mathbb{E}\left[(\tau^2 + \sigma^2)\,\|X\|_2^{-2} \mid \tau^2\right] = \frac{1/2}{n/2 - 1} = \frac{1}{n-2}.$$

Thus, $(n-2)\sigma^2\,\|X\|_2^{-2}$ is an unbiased estimator for $\sigma^2/(\tau^2 + \sigma^2) = 1 - \alpha$, so

$$\mathbb{E}[1 - (n-2)\sigma^2\,\|X\|_2^{-2} \mid \tau^2] = \frac{\tau^2}{\tau^2 + \sigma^2}.$$

This gives us an empirical Bayes estimator

$$\delta(X) = \left(1 - \frac{(n-2)\sigma^2}{\|X\|_2^2}\right)X,$$

which is precisely the James-Stein estimator.

(b) Using the standard form of the MLE for Gaussian variances,

$$\hat{\tau}^2 = \arg\max_{\tau^2} f(X \mid \tau^2) = \left( \frac{1}{n} \sum_{i=1}^{n} X_i^2 - \sigma^2 \right)^+ .$$

With this, we have an empirical Bayes estimator

$$\delta(X) = \frac{\hat{\tau}^2}{\hat{\tau}^2 + \sigma^2} X = \frac{\left( \|X\|_2^2 / n - \sigma^2 \right)^+}{\|X\|_2^2 / n} X = \left( 1 - \frac{n\sigma^2}{\|X\|_2^2} \right)^+ X.$$

This resembles the positive-part James-Stein estimator, up to the replacement of $n-2$ with $n$.