STAT 581
Homework #5
11/6/2013

Solutions for Hoff Exercises 6.3, 6.6, and 6.8

6.3 Note that the risk advantage of δ_{JS} over $\boldsymbol{X}$ is largest when $\|\boldsymbol{\theta}\|^2$ is small, i.e. $\boldsymbol{\theta} \approx \mathbf{0}$. Suppose instead you thought that $\boldsymbol{\theta} \approx \boldsymbol{\theta}_0$. Derive an alternative estimator that performs well when $\boldsymbol{\theta} \approx \boldsymbol{\theta}_0$, in that it beats δ_{JS} at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ and also dominates $\boldsymbol{X}$ everywhere. Show that your estimator meets these latter two criteria.

Intuitively, a good estimator to try would be a modified James Stein estimator where we shrink towards $\boldsymbol{\theta}_0$ instead of towards $\mathbf{0}$. A good candidate for such an estimator might be

$$\delta_{\boldsymbol{\theta}_0} = \boldsymbol{\theta}_0 + \left(1 - \frac{p-2}{\|\boldsymbol{X} - \boldsymbol{\theta}_0\|^2}\right)(\boldsymbol{X} - \boldsymbol{\theta}_0) = \boldsymbol{X} - \left(\frac{p-2}{\|\boldsymbol{X} - \boldsymbol{\theta}_0\|^2}\right)(\boldsymbol{X} - \boldsymbol{\theta}_0).$$

Letting $g(\boldsymbol{X}) = \left(\frac{p-2}{\|\boldsymbol{X}-\boldsymbol{\theta}_0\|^2}\right)(\boldsymbol{X} - \boldsymbol{\theta}_0)$, we can express our estimator as $\delta_{\boldsymbol{\theta}_0} = \boldsymbol{X} - g(\boldsymbol{X})$. Then to compute the risk, we have

$$\begin{aligned}
R(\boldsymbol{\theta}, \delta_{\boldsymbol{\theta}_0}) &= \mathrm{E}[\|\delta_{\boldsymbol{\theta}_0} - \boldsymbol{\theta}\|^2] \\
&= \mathrm{E}[\|\boldsymbol{X} - g(\boldsymbol{X}) - \boldsymbol{\theta}\|^2] \\
&= \mathrm{E}[\|(\boldsymbol{X} - \boldsymbol{\theta}) - g(\boldsymbol{X})\|^2] \\
&= \mathrm{E}[\|(\boldsymbol{X} - \boldsymbol{\theta})\|^2] + \mathrm{E}[\|g(\boldsymbol{X})\|^2] - 2\mathrm{E}[(\boldsymbol{X} - \boldsymbol{\theta})^T g(\boldsymbol{X})].
\end{aligned}$$

We then compute these three expectations separately. Since $\boldsymbol{X} \sim N_p(\boldsymbol{\theta}, \boldsymbol{I})$, we have $\|\boldsymbol{X} - \boldsymbol{\theta}\|^2 \sim \chi^2_p$, so $\mathrm{E}[\|(\boldsymbol{X} - \boldsymbol{\theta})\|^2] = p$.

For the second term, we have

$$\begin{aligned}
\mathrm{E}\left[\|g(\boldsymbol{X})\|^2\right] &= \mathrm{E}\left[\frac{(p-2)^2\|\boldsymbol{X} - \boldsymbol{\theta}_0\|^2}{\|\boldsymbol{X} - \boldsymbol{\theta}_0\|^4}\right] \\
&= (p-2)^2 \mathrm{E}\left[\frac{1}{\|\boldsymbol{X} - \boldsymbol{\theta}_0\|^2}\right]
\end{aligned}$$

For the third term, applying Stein's identity yields

$$\mathrm{E}[(\boldsymbol{X} - \boldsymbol{\theta})^T g(\boldsymbol{X})] = (p-2)^2 \mathrm{E}\left[\frac{1}{\|\boldsymbol{X} - \boldsymbol{\theta}_0\|^2}\right],$$

so all together we have

$$R(\boldsymbol{\theta}, \delta_{\boldsymbol{\theta}_0}) = p - (p-2)^2 \mathrm{E}\left[\frac{1}{\|\boldsymbol{X} - \boldsymbol{\theta}_0\|^2}\right].$$

Now our two goals are to show i) that this risk is smaller than the risk of $\boldsymbol{X}$ everywhere and ii) that this risk is smaller than the risk of the usual James Stein estimator at $\boldsymbol{\theta}_0$.

For i), we note that the risk of $\boldsymbol{X}$ is p everywhere, so since $(p-2)^2\mathrm{E}\left[\frac{1}{\|\boldsymbol{X}-\boldsymbol{\theta}_0\|^2}\right] > 0$, we have $R(\boldsymbol{\theta}, \delta_{\boldsymbol{\theta}_0}) < R(\boldsymbol{\theta}, \boldsymbol{X})$ everywhere.

For ii), we are comparing the risk functions

$$R(\boldsymbol{\theta}, \delta_{\boldsymbol{\theta}_0}) = p - (p-2)^2\mathrm{E}\left[\frac{1}{\|\boldsymbol{X}-\boldsymbol{\theta}_0\|^2}\right]$$

and

$$R(\boldsymbol{\theta}, \delta_{JS}) = p - (p-2)^2\mathrm{E}\left[\frac{1}{\|\boldsymbol{X}\|^2}\right],$$

so to show that $R(\boldsymbol{\theta}_0, \delta_{\boldsymbol{\theta}_0}) < R(\boldsymbol{\theta}_0, \delta_{JS})$, it is enough to show that

$$\mathrm{E}\left[\frac{1}{\|\boldsymbol{X}-\boldsymbol{\theta}_0\|^2}\right] < \mathrm{E}\left[\frac{1}{\|\boldsymbol{X}\|^2}\right]$$

when $\boldsymbol{X} \sim N_p(\boldsymbol{\theta}_0, \boldsymbol{I})$.

One option is to just appeal to the results of the unassigned exercise 6.7. A more thorough option is to apply Lehman and Romano Lemma 3.4.2. Since the non-central χ^2 distribution has monotone likelihood ratio, for $Z_\theta \sim \chi^2(\theta)$, LR Lemma 3.4.2.i gives $\mathrm{E}[1/Z_\theta^2]$ is non-decreasing in θ. Since $\|\boldsymbol{X}-\boldsymbol{\theta}_0\| \sim \chi^2(0)$ and $\|\boldsymbol{X}\|^2 \sim \chi^2(\|\boldsymbol{\theta}_0\|)$, we have the desired inequality. ■

6.6 For estimating $\boldsymbol{\theta}$ based on $\boldsymbol{X} \sim N_p(\boldsymbol{\theta}, \boldsymbol{I})$, consider the class of adaptive shrinkage estimators of the form $\delta_c(\boldsymbol{x}) = (1 - \frac{c}{\|\boldsymbol{x}\|^2})\boldsymbol{x}$.
6.6.a Under squared error loss, find the simplest expression you can for the risk function of δ_c.

Here we use the same set of tricks as in exercise 6.3. Write $\delta_c = \boldsymbol{X} - \frac{c}{\|\boldsymbol{X}\|^2}\boldsymbol{X} = \boldsymbol{X} - g(\boldsymbol{X})$. Then, as before, we have

$$R(\boldsymbol{\theta}, \delta_c) = \mathrm{E}[\|(\boldsymbol{X}-\boldsymbol{\theta})\|^2] + \mathrm{E}[\|g(\boldsymbol{X})\|^2] - 2\mathrm{E}[(\boldsymbol{X}-\boldsymbol{\theta})^T g(\boldsymbol{X})].$$

Again, the first term in this sum is p. The second term is

$$\begin{aligned}\mathrm{E}\left[\|g(\boldsymbol{X})\|^2\right] &= \mathrm{E}\left[\frac{c^2\|\boldsymbol{X}\|^2}{\|\boldsymbol{X}\|^4}\right] \\ &= c^2\mathrm{E}\left[\frac{1}{\|\boldsymbol{X}\|^2}\right]\end{aligned}$$

and for the third term, applying Stein's identity yields

$$\mathrm{E}[(\boldsymbol{X}-\boldsymbol{\theta})^T g(\boldsymbol{X})] = c(p-2)\mathrm{E}\left[\frac{1}{\|\boldsymbol{X}\|^2}\right],$$

so all together we have

$$R(\boldsymbol{\theta}, \delta_c) = p + (c^2 - 2c(p-2))\mathrm{E}\left[\frac{1}{\|\boldsymbol{X}\|^2}\right].$$

■

6.6.b Now suppose we are in a hierarchical situation where $\boldsymbol{\theta} \sim N_p(\boldsymbol{0}, \tau^2\boldsymbol{I})$. Obtain a closed-form expression for the risk $R(\tau^2, \delta_c)$ of δ_c, where now risk means the loss averaged over $\boldsymbol{X}$ and $\boldsymbol{\theta}$. Find the value $\tilde{c}$ of c that minimizes $R(\tau^2, \delta_c)$.

The only thing we're missing to evaluate this risk is a way to calculate $\mathrm{E}[1/\|\boldsymbol{X}\|^2]$. For this, we can appeal to the known marginal distribution of $\boldsymbol{X}$. We know $\boldsymbol{X} \sim N_p(\boldsymbol{0}, (1+\tau^2)\boldsymbol{I})$. Then, $\|\boldsymbol{X}\|^2$ follows a Gamma$(\frac{p}{2}, \frac{1}{2(1+\tau^2)})$ distribution. So $1/\|\boldsymbol{X}\|^2$ follows an inverse gamma distribution with mean $\frac{1}{(p-2)(1+\tau^2)}$. Plugging this into our expression for the risk, we get

$$R(\tau^2, \delta_c) = p + \frac{c^2 - 2c(p-2)}{(p-2)(1+\tau^2)}.$$

This expression is quadratic in c and minimized at $\tilde{c} = p - 2$, so the optimal estimator of this form is the James-Stein estimator. ■

6.6.c For small, medium, and large values of p, plot $R(\tau^2, \delta_{\tilde{c}})$ as a function of τ^2, along with $R(\tau^2, \frac{\tau^2}{1+\tau^2}\boldsymbol{X})$. Describe and interpret what you see.

Plugging in our expression for $\tilde{c}$ yields $R(\tau^2, \delta_{\tilde{c}}) = p - \frac{p-2}{1+\tau^2}$. From lecture, we have $R(\tau^2, \frac{\tau^2}{1+\tau^2}\boldsymbol{X}) = p\frac{\tau^2}{1+\tau^2}$. [Plots omitted.] For fixed τ^2, the risk of the oracle estimator is always lower than the risk of $\delta_{\tilde{c}}$, but the two risks converge as $p \to \infty$. This makes sense. For large numbers of observations, we can estimate the value of τ^2 with a high degree of precision. ■

6.8 Let $\boldsymbol{X} \sim N_p(\boldsymbol{\theta}, \sigma^2\boldsymbol{I})$ and $S/\sigma^2 \sim \chi^2_n$ be independent. For the case that σ^2 is unknown, consider the class of estimators of the form $\delta_c(\boldsymbol{x}, s) = (1 - \frac{cs}{\|\boldsymbol{x}\|^2})\boldsymbol{x}$.
6.8.a Compute the risk function of δ_c.

We continue to use the same tricks as in the previous problems. Write $\delta_c(\boldsymbol{X}, S) = \boldsymbol{X} - \frac{cS}{\|\boldsymbol{X}\|^2}\boldsymbol{X} = \boldsymbol{X} - g(\boldsymbol{X}, S)$. Then, as before, we have

$$R(\boldsymbol{\theta}, \delta_c) = \mathrm{E}[\|(\boldsymbol{X} - \boldsymbol{\theta})\|^2] + \mathrm{E}[\|g(\boldsymbol{X}, S)\|^2] - 2\mathrm{E}[(\boldsymbol{X} - \boldsymbol{\theta})^T g(\boldsymbol{X}, S)].$$

The first term in this sum is $p\sigma^2$. The second term is

$$\begin{aligned}
\mathrm{E}\left[\|g(\boldsymbol{X}, S)\|^2\right] &= \mathrm{E}\left[\frac{c^2 S^2 \|\boldsymbol{X}\|^2}{\|\boldsymbol{X}\|^4}\right] \\
&= c^2 \mathrm{E}[S^2] \mathrm{E}\left[\frac{1}{\|\boldsymbol{X}\|^2}\right] \\
&= c^2 \sigma^4 (n^2 + 2n) \mathrm{E}\left[\frac{1}{\|\boldsymbol{X}\|^2}\right]
\end{aligned}$$

and for the third term, applying Stein's identity yields

$$\begin{aligned}
\mathrm{E}[(\boldsymbol{X} - \boldsymbol{\theta})^T g(\boldsymbol{X}, S)] &= c(p-2)\sigma^2 \mathrm{E}\left[\frac{S}{\|\boldsymbol{X}\|^2}\right] \\
&= c(p-2) n \sigma^4 \mathrm{E}\left[\|\boldsymbol{X}\|^2\right]
\end{aligned}$$

so all together we have

$$R(\boldsymbol{\theta}, \delta_c) = p\sigma^2 + (c^2\sigma^4(n^2+2n) - 2cn\sigma^4(p-2))\mathrm{E}\left[\frac{1}{\|\boldsymbol{X}\|^2}\right].$$

■

6.8.b Find the value $\tilde{c}$ of c that minimizes the risk function.

Again, we have an expression for risk which is quadratic in c. This expression is minimized at

$$\tilde{c} = -\frac{-2n\sigma^4(p-2)}{2\sigma^4(n^2+2n)} = \frac{p-2}{n+2}.$$

■

6.8.c Compare the risk function of δ_{JS} in the case $\sigma^2 = 1$ is known to the risk of $\delta_{\tilde{c}}$. How much is lost by not knowing σ^2?

The known risk function for the James Stein estimator in the case of $\sigma^2 = 1$ known is given by

$$R(\boldsymbol{\theta}, \delta_{JS}) = p - (p-2)^2 \mathrm{E}\left[\frac{1}{\|\boldsymbol{X}\|^2}\right].$$

Plugging in our value of $\tilde{c}$ and $\sigma^2 = 1$, we get a risk function for our alternative estimator of

$$R(\boldsymbol{\theta}, \delta_{\tilde{c}}) = p - \frac{n}{n+2}(p-2)^2 \mathrm{E}\left[\frac{1}{\|\boldsymbol{X}\|^2}\right].$$

Thus, the difference in risks is given by

$$R(\boldsymbol{\theta}, \delta_{\tilde{c}}) - R(\boldsymbol{\theta}, \delta_{JS}) = \frac{2}{n+2}(p-2)^2 \mathrm{E}\left[\frac{1}{\|\boldsymbol{X}\|^2}\right].$$

For fixed p, as n goes to infinity, this difference in risks goes to zero. ■