Feynman Liang
STAT 210A
Lecture #12
10/04/2016

1 Empirical Bayes / James-Stein Estimator

$\theta_i \sim \mathcal{N}(0, \tau^2)$, $X_i|\theta_i \sim \mathcal{N}(\theta_i, 1)$, $i \leq n$

$$\theta_i^{\text{Bayes}}(X) = \left(1 - \frac{1}{1+\tau^2}\right) X_i \tag{1.1}$$
$$= (1-\zeta)\, X_i \tag{1.2}$$
$$\theta_i^{\text{HBayes}}(X) = (1 - \mathbb{E}[\zeta \mid X])\, X_i \tag{1.3}$$

$$\zeta \sim \text{Unif}(0,1) = \zeta 1_{\zeta \in [0,1]} \tag{1.4}$$
$$X_i \mid \theta^2 \sim \mathcal{N}(0, 1+\tau^2) \tag{1.5}$$
$$X_i \mid \zeta \sim \mathcal{N}(0, \zeta^{-1} I_n) = \left(\frac{\zeta}{2\pi}\right)^{n/2} e^{-\frac{\zeta}{2}\|X\|^2} \tag{1.6}$$

The prior and likelihood for ζ are basically conjugate, so

$$p(\zeta \mid X) \propto \underbrace{\zeta^{\frac{n+2}{2}} e^{-\frac{\zeta}{2}\|X\|^2}}_{\text{Gamma distr}} \quad \underbrace{1_{\zeta \in [0,1]}}_{\text{truncated to } [0,1]} \tag{1.7}$$

$$\sim \text{Gamma}\left(\underbrace{1 + \frac{n}{2}}_{\tilde{k},\text{shape param}}, \underbrace{\frac{1}{\|X\|^2/2}}_{\tilde{\sigma},\text{scale param}} \right) 1_{\zeta < 1} \tag{1.8}$$

If n large

$$\mathbb{E}[\zeta \mid X] \approx \tilde{k}\tilde{\sigma} = \frac{n+2}{\|X\|^2} \approx \frac{1}{\zeta^2 + 1} = \zeta \tag{1.9}$$
$$\text{Var}(\zeta \mid X) \approx \tilde{k}\tilde{\sigma}^2 = \frac{2(n+2)}{\|X\|^4} \to 0 \tag{1.10}$$

1.1 James-Stein Estimator

$$\delta_i^{JS}(X) = \left(1 - \frac{n-2}{\|X\|^2}\right) X_i \tag{1.11}$$

With $n \geq 3$ (to get the James-Stein paradox going).

Proposition 1.1. *If $Y \sim \chi_n^2$, $n \geq 3$, then*

$$\mathbb{E}[\frac{1}{Y}]^n = \frac{1}{n-2} \tag{1.12}$$

Proof.

$$\mathbb{E}\frac{1}{y} = \int \frac{1}{y} \frac{1}{2^{n/2}\Gamma(\frac{n}{2})} y^{n/2-1} e^{-y/2} dy \tag{1.13}$$

+ algebra ☐

$$\mathbb{E}\left[\frac{n-2}{\|X\|_2}\right] = \frac{n-2}{1+\tau^2}\mathbb{E}\left[\frac{1}{\|X\|_2/(1+\tau^2)}\right] = \zeta \tag{1.14}$$

Example 1.2. $X_i \overset{\text{ind}}{\sim} \mathcal{N}(\theta_i, 1)$, $i = 1, \cdots, n$

Many estimators exist:

- $\theta(X) = X$
- MLE
- UMVU
- flat-prior estimator

For all $\theta_i \in \mathbb{R}^n$

$$R(\theta, \delta^{JS}) < R(\theta, X) \tag{1.15}$$

2 Stein's Lemma

Lemma 2.1 (Stein's Lemma (Univariate)). *Suppose $X \sim \mathcal{N}(\theta, \sigma^2)$, $h(x) : \mathbb{R} \to \mathbb{R}$ differentiable, $\mathbb{E}|h'(X)| < \infty$. Then*

$$\underbrace{\mathbb{E}[(X-\theta)h(X)]}_{=Cov(X,h(X))} = \sigma^2 \mathbb{E}[h'(X)] \tag{2.1}$$

Proof. First assume $\theta = 0$, $\sigma^2 = 1$.

$$\mathbb{E}[Xh(X)] = -\int_{-\infty}^{\infty} h(x) \underbrace{(-x\frac{1}{\sqrt{2\pi}}e^{-x^2/2}dx)}_{d\phi(x)} \tag{2.2}$$

$$= \cancel{-h(x)\phi(x)\Big]_{-\infty}^{\infty}}^{\,0} + \int h'(x)\phi(x)dx \tag{2.3}$$

For general θ, σ^2. Write $X = \theta + \sigma Z$ where $Z \sim \mathcal{N}(0,1)$.

$$\mathbb{E}[(X-\theta)h(X)] = \sigma\mathbb{E}[Zh(\theta+\sigma Z)] \tag{2.4}$$
$$= \sigma^2\mathbb{E}[h'(\theta+\sigma Z)] \tag{2.5}$$
$$= \sigma^2\mathbb{E}[h'(X)] \tag{2.6}$$

□

Definition 2.2. $Dh(x) \in \mathbb{R}^{d\times d}$, $[Dh(x)]_{i,j} = \frac{\partial h_i(x)}{\partial x_j}$

Lemma 2.3 (Stein's Lemma (Multivariate))**.** $X \sim \mathcal{N}_d(\theta, \sigma^2 I_d)$, $\theta \in \mathbb{R}^d$, $h : \mathbb{R}^d \to \mathbb{R}^d$.
If $\mathbb{E}\left[\left(\sum^{i,j} Dh(x)_{i,j}^2\right)^{1/2}\right] < \infty$ *(bounded Frobenius norm), then*

$$\mathbb{E}[(X-\theta)^\intercal h(X)] = \sigma^2\mathbb{E}[\operatorname{tr} Dh(X)] \tag{2.7}$$

Proof.

$$\mathbb{E}[(X_i-\theta_i)h_i(X)] = \mathbb{E}[\mathbb{E}[(X_i-\theta_i)h_i(X) \mid X_{\setminus i}]] \tag{2.8}$$
$$= \mathbb{E}[\sigma^2\mathbb{E}[\frac{\partial h_i}{\partial x_i}(X) \mid X_{\setminus i}]] \tag{2.9}$$
$$= \sigma^2\mathbb{E}[Dh(X)_{ii}] \tag{2.10}$$

sum over i. □

3 Stein's Unbiased Risk Estimator (SURE)

Assume $\sigma^2 = 1$.

$$\hat{R} = d + \|h(X)\|^2 - 2\operatorname{tr} Dh(X) \tag{3.1}$$
$$\mathbb{E}_\theta\hat{R} = \text{MSE}(\theta,\delta) \tag{3.2}$$

This is used for tuning hyperparameters of ML methods because it lets us estimate MSE loss

$$MSE(\theta,\delta) = \mathbb{E}_\theta[\|X-\theta-h(X)\|] \tag{3.3}$$
$$= \mathbb{E}_\theta[\|X-\theta\|^2] + \mathbb{E}_\theta[\|h(X)\|^2] - 2\mathbb{E}_\theta[(X-\theta)^\intercal h(X)] \tag{3.4}$$
$$= d + \mathbb{E}_\theta[\|h(X)\|^2] - 2\mathbb{E}_\theta[\operatorname{tr} Dh(X)] \tag{3.5}$$
$$= \mathbb{E}_\theta\hat{R} \tag{3.6}$$

Covariance ($\mathbb{E}_\theta[(X-\theta)^\intercal h(X)] = \text{Cov}(X, h(X))$ is troublesome term; Stein's lemma lets us replace that with something easier.

Example 3.1. $\delta(X) = X$. The natural estimator $h(X) = 0$, $Dh(X) = 0$. $\hat{R} = d$

Example 3.2. $\delta(X) = (1-\zeta)X$. $h(X) = \zeta X = Dh(X) = \zeta I_d$.

$$\begin{aligned}
\hat{R} &= d + \zeta^2 \|X\|^2 - 2\zeta d && (3.7)\\
&= (1-2\zeta)d + \zeta^2\|X\|^2 && (3.8)\\
R(\theta, (1-\zeta)X) &= (1-2\zeta)d + \zeta^2 \mathbb{E}_\theta \sum X_i^2 && (3.9)\\
&= (1-2\zeta)d + \zeta^2(\|\theta\|^2 + d) && (3.10)\\
&= \underbrace{(1-\zeta)^2 d}_{\text{variance}} + \underbrace{\zeta^2\|\theta\|^2}_{\text{bias}^2} && (3.11)
\end{aligned}$$

Example 3.3 (James-Stein estimator)**.**

$$\begin{aligned}
\delta(X) &= \left(1 - \frac{d-2}{\|X\|^2}\right) X && (3.12)\\
h(X) &= \frac{d-2}{\|X\|^2} X && (3.13)\\
\|h(X)\|^2 &= \frac{(d-2)^2}{\|X\|^4}\|X\|^2 = \frac{(d-2)^2}{\|X\|^2} && (3.14)\\
Dh(X)_{ii} &= \frac{\partial h_i(x)}{\partial x_i} = \frac{\partial}{\partial x_i}\frac{(d-2)X_i}{\|X\|^2} && (3.15)\\
&= \frac{\|X\|^2(d-2) - 2(d-2)X_i^2}{\|X\|^4} && (3.16)\\
\implies \operatorname{tr} Dh(X)_{ii} &= (d-2)\left[\frac{d\|X\|^2}{\|X\|^4} - \frac{2\|X\|^2}{\|X\|^4}\right] = \frac{(d-2)^2}{\|X\|^4} && (3.17)
\end{aligned}$$

$$\begin{aligned}
\hat{R}_{JS} &= d + \frac{(d-2)^2}{\|X\|^2} - 2\frac{(d-2)^2}{\|X\|^2} = d - \frac{(d-2)^2}{\|X\|^2} && (3.18)\\
\mathrm{MSE}(\theta, \delta_{JS}) &= d - \mathbb{E}_\theta\left[\frac{(d-2)^2}{\|X\|^2}\right] < d && (3.19)
\end{aligned}$$

$$\implies \delta(X) = X \text{ is inadmissible} \qquad (3.20)$$

4 Stein's Paradox