STATS 305B: Methods for Applied Statistics I **Winter 2016/17**

Lecture 26: March 13

Lecturer: Jonathan Taylor *Scribes: Kenneth Tay*

26.1 Stein's Unbiased Risk Estimate (SURE)

Say we have data $Y \in \mathbb{R}^n$, and $Y \sim \mathcal{N}(\mu, I_{n\times n})$ (i.e. covariance known). Say we have k estimators $\hat{\theta}_1, \dots, \hat{\theta}_k$ for μ, and we wish to choose the "best" one.

One way to do this is the choose the estimator which minimizes squared error risk, i.e. $R(\mu, \hat{\theta}) = \mathbb{E}_\mu \left[\|\hat{\theta}(Y) - \mu\|^2\right]$.

We can't compute these quantities directly since μ is unknown. Stein's Unbiased Risk Estimate (SURE) gives us a way to estimate $R(\mu, \hat{\theta})$.

Definition 26.1 *We define **SURE** to be*

$$\begin{aligned}
\hat{R}(\mu, \hat{\theta}) &= -n + \|Y - \hat{\theta}(Y)\|^2 + 2\sum_{i=1}^n \frac{\partial \hat{\theta}_i}{\partial Y_i}(Y) \\
&= -n + \|Y - \hat{\theta}(Y)\|^2 + 2\, div\, \hat{\theta}.
\end{aligned}$$

Proposition 26.2 *Under the true distribution, the expected value of SURE is equal to the squared error risk, i.e.*

$$\mathbb{E}_\mu[SURE] = \mathbb{E}_\mu\left[\hat{R}(\mu, \hat{\theta})\right] = R(\mu, \hat{\theta}).$$

The proof uses Stein's Lemma:

Lemma 26.3 *Suppose $Z \sim \mathcal{N}(\mu, 1)$. Then for smooth g (such that $g(z)$ does not grow too quickly as $|z| \to \infty$),*

$$\mathbb{E}_\mu[g(Z)(Z-\mu)] = \mathbb{E}_\mu[g'(Z)].$$

Proof: By integration by parts:

$$\begin{aligned}
\mathbb{E}_\mu[g(Z)(Z-\mu)] &= \int_{-\infty}^{\infty} g(z)(z-\mu)e^{-(z-\mu)^2/2}dz \\
&= \left[-g(z)e^{-(z-\mu)^2/2}\right]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-(z-\mu)^2/2} g'(z)dz \\
&= \mathbb{E}_\mu[g'(Z)].
\end{aligned}$$

■

Proof:[Proof of Proposition 26.2]

Case 1: $Y \in \mathbb{R}$.

Note that

$$\mathbb{E}_\mu[\text{SURE}] = \mathbb{E}_\mu\left[-1 + (Y - \hat{\theta}(Y))^2 + 2\frac{\partial\hat{\theta}}{\partial Y}\right] = -1 + \mathbb{E}_\mu\left[(Y - \hat{\theta}(Y))^2\right] + 2\mathbb{E}_\mu\left[\frac{\partial\hat{\theta}}{\partial Y}\right].$$

Let $g(Y) = \hat{\theta}(Y) - Y$. Then

$$\begin{aligned}
(\hat{\theta}(Y) - \mu)^2 &= (Y - \mu + g(Y))^2 \\
&= (Y - \mu)^2 + 2g(Y)(Y - \mu) + [g(Y)]^2, \\
\mathbb{E}_\mu\left[(\hat{\theta}(Y) - \mu)^2\right] &= 1 + 2\mathbb{E}_\mu[g(Y)(Y - \mu)] + \mathbb{E}_\mu[g(Y)^2] \\
&= 1 + 2\mathbb{E}_\mu[g'(Y)] + \mathbb{E}_\mu\left[(\hat{\theta}(Y) - Y)^2\right] \\
&= 1 + 2\left[-1 + \mathbb{E}_\mu\frac{\partial\hat{\theta}}{\partial Y}\right] + \mathbb{E}_\mu\left[(\hat{\theta}(Y) - Y)^2\right] \\
&= \mathbb{E}_\mu[\text{SURE}].
\end{aligned}$$

Case 2: $Y \in \mathbb{R}^n$.

Let g be as before. Note that

$$\begin{aligned}
\|\hat{\theta}(Y) - \mu\|^2 &= \|Y - \mu + g(Y)\|^2 \\
&= \|Y - \mu\|^2 + 2g(Y)^T(Y - \mu) + \|g(Y)\|^2, \\
\mathbb{E}_\mu\left[\|\hat{\theta}(Y) - \mu\|^2\right] &= \mathbb{E}\|Y - \mu\|^2 + 2\sum_{i=1}^n \mathbb{E}_\mu[g_i(Y)(Y - \mu)] + \mathbb{E}_\mu\|g(Y)\|^2 \\
&= n + 2\sum_{i=1}^n \mathbb{E}_\mu[g_i(Y)(Y - \mu)] + \mathbb{E}_\mu\left[\|\hat{\theta}(Y) - Y\|^2\right],
\end{aligned}$$

since $\|Y - \mu\|^2 \sim \chi^2_n$. For each i,

$$\begin{aligned}
\mathbb{E}_\mu[g_i(Y)(Y - \mu)] &= \mathbb{E}_\mu\left[\mathbb{E}_\mu\left[g_i(Y)(Y - \mu) \mid Y_{-i}\right]\right] \\
&= \mathbb{E}_\mu\left[\frac{\partial g_i(Y)}{\partial Y_i}\right] \\
&= \mathbb{E}_\mu\left[-1 + \frac{\partial\hat{\theta}_i(Y)}{\partial Y_i}\right].
\end{aligned}$$

Hence,

$$\begin{aligned}
\mathbb{E}_\mu\left[\|\hat{\theta}(Y) - \mu\|^2\right] &= n + 2\sum_{i=1}^n \mathbb{E}_\mu\left[-1 + \frac{\partial\hat{\theta}_i(Y)}{\partial Y_i}\right] + \mathbb{E}_\mu\left[\|\hat{\theta}(Y) - Y\|^2\right] \\
&= -n + 2\mathbb{E}_\mu[\text{div }\hat{\theta}] + \mathbb{E}_\mu\left[\|\hat{\theta}(Y) - Y\|^2\right] \\
&= \mathbb{E}_\mu[SURE].
\end{aligned}$$

∎

Definition 26.4 *In this context, we define the degrees of freedom of $\hat{\theta}$ to be*

$$df(\hat{\theta}) := div\ \hat{\theta}.$$