STAT 581
Homework #1
10/7/2013

Solutions for Hoff Exercises 3.1, 3.2, 4.2, and 4.4

3.1 Suppose we want to estimate a parameter θ under a strictly convex loss function. Let X be the data and let $T = T(X)$ be a sufficient statistic that is not a 1-1 function of X. Show that any estimator $\hat{\theta}(X)$ that is a function of X and not T is inadmissible. In your construction, where did you use the fact that T is sufficient, and not just any function of X?

Let $\hat{\theta}$ be an estimator that is a function of X and not T. Consider the alternative estimator $\delta = \mathrm{E}_{X|T}[\hat{\theta}(X)|T]$. By sufficiency of T, the distribution of $X|T$ does not depend on θ, so δ is a function of T but not of θ. Then

$$\begin{aligned} R(\theta, \hat{\theta}) &= \mathrm{E}_X[L(\theta, \hat{\theta}(X))] \\ &= \mathrm{E}_T[\mathrm{E}_{X|T}[L(\theta, \hat{\theta}(X))|T]] \\ &> \mathrm{E}_T[L(\theta, \mathrm{E}_{X|T}[\hat{\theta}(X)|T])] \qquad \text{by Jensen's Inequality} \\ &= \mathrm{E}_T[L(\theta, \delta(T(X)))] \\ &= \mathrm{E}_X[L(\theta, \delta(T(X)))] \\ &= R(\theta, \delta). \end{aligned}$$

Thus, since δ dominates $\hat{\theta}$, we have that $\hat{\theta}$ is inadmissible. ■

3.2 Based on $\bar{X} \sim N(\theta, 1/n)$, suppose you need to decide among three actions: 1) stating nothing, 2) stating $\theta < 0$, or 3) stating $\theta > 0$. Refer to these decisions numerically as $d = 0, d = -1$, and $d = 1$, respectively, and let the loss be $L(\theta, d) = 1 - d \cdot \mathrm{sign}(\theta)$.
3.2.a Consider a decision rule of the form $\delta(\bar{x}) = \mathrm{sign}(\bar{x}) \cdot 1(|\bar{x}| > c)$. Compute the risk as a function of θ and n, and plot the risk function for several values of n.

For convenience, note that $\bar{X} \sim N(\theta, 1/n) \Rightarrow \sqrt{n}(\bar{X} - \theta) \sim N(0, 1)$. Now,

$$\begin{aligned}
R(\theta, \delta) &= \mathrm{E}[L(\theta, \delta)] \\
&= \mathrm{E}[1 - \delta \operatorname{sign}(\theta)] \\
&= \mathrm{E}[I(\delta = 1)(1 - \operatorname{sign}(\theta)) + I(\delta = 0) \cdot 1 + I(\delta = -1)(1 + \operatorname{sign}(\theta))] \\
&= \mathrm{E}[I(\bar{X} > c)(1 - \operatorname{sign}(\theta)) + I(-c \leq \bar{X} \leq c) \cdot 1 + I(\bar{X} < -c)(1 + \operatorname{sign}(\theta))] \\
&= \Pr(\bar{X} > c)(1 - \operatorname{sign}(\theta)) + \Pr(-c \leq \bar{X} \leq c) \cdot 1 + \Pr(\bar{X} < c)(1 + \operatorname{sign}(\theta)) \\
&= \Pr(\sqrt{n}(\bar{X} - \theta) > \sqrt{n}(c - \theta))(1 - \operatorname{sign}(\theta)) \\
&\quad + \Pr(\sqrt{n}(-c - \theta) \leq \sqrt{n}(\bar{X} - \theta) \leq \sqrt{n}(c - \theta)) \cdot 1 \\
&\quad + \Pr(\sqrt{n}(\bar{X} - \theta) < \sqrt{n}(-c - \theta))(1 + \operatorname{sign}(\theta)) \\
&= (1 - \Phi(\sqrt{n}(c - \theta)))(1 - \operatorname{sign}(\theta)) \\
&\quad + \Phi(\sqrt{n}(c - \theta)) - \Phi(\sqrt{n}(-c - \theta)) \\
&\quad + \Phi(\sqrt{n}(-c - \theta))(1 + \operatorname{sign}(\theta)),
\end{aligned}$$

where Φ is the standard normal CDF.

Plots of this function are below. ∎

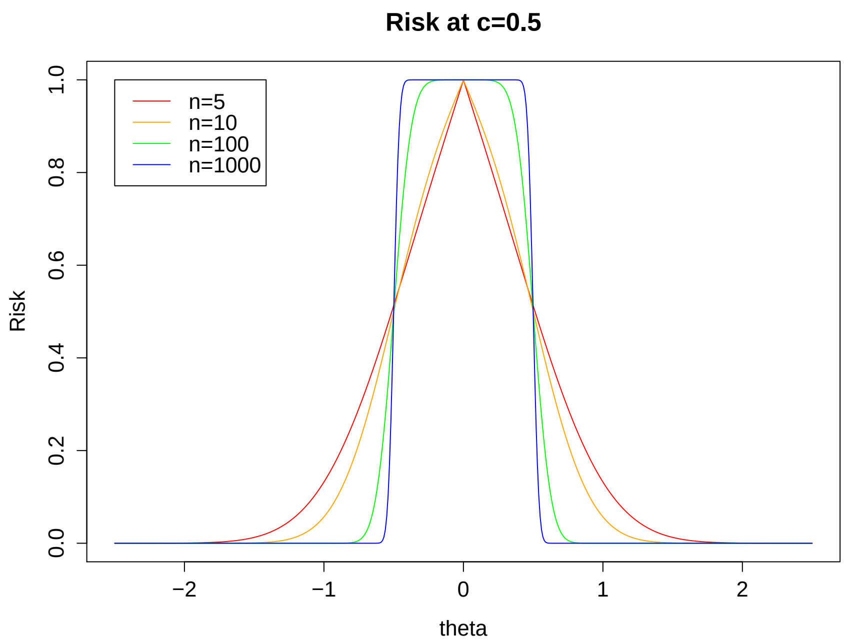

Figure 1: Plots of the risk function for $c = 0.5$ and $n \in \{5, 10, 100, 1000\}$.

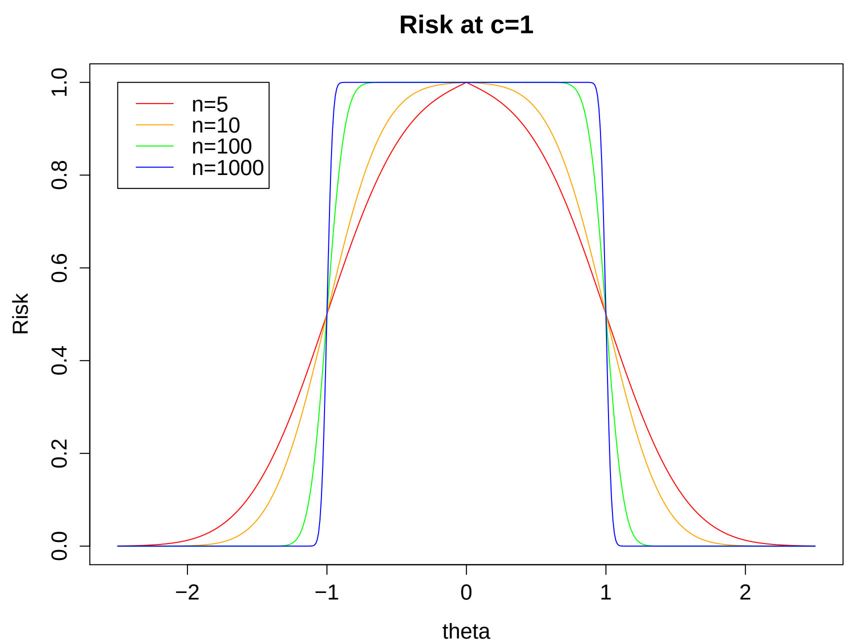

Figure 2: Plots of the risk function for $c = 1$ and $n \in \{5, 10, 100, 1000\}$.

3.2.b Consider a decision rule based on the z-test: If the test of $\theta = 0$ is rejected at level $\alpha = 0.05$, then take d to be the sign of $\bar{X}$. If the test doesn't reject, then take $d = 0$. Compute the risk function of this procedure as a function of n for several values of n, compare to the risk function in (a) and comment.

For each fixed value of n, this decision rule just reduces to a decision rule based on a cutoff value of $c = 1.96/\sqrt{n}$. We know the risk function for such a decision rule from part a, so our new risk function is just

$$\begin{aligned} R(\theta, \delta) = &(1 - \Phi(\sqrt{n}(c/\sqrt{n} - \theta)))(1 - \text{sign}(\theta)) \\ &+ \Phi(\sqrt{n}(c/\sqrt{n} - \theta)) - \Phi(\sqrt{n}(-c/\sqrt{n} - \theta)) \\ &+ \Phi(\sqrt{n}(-c/\sqrt{n} - \theta))(1 + \text{sign}(\theta)). \end{aligned}$$

We can plot this to compare against the risk function from part (a):

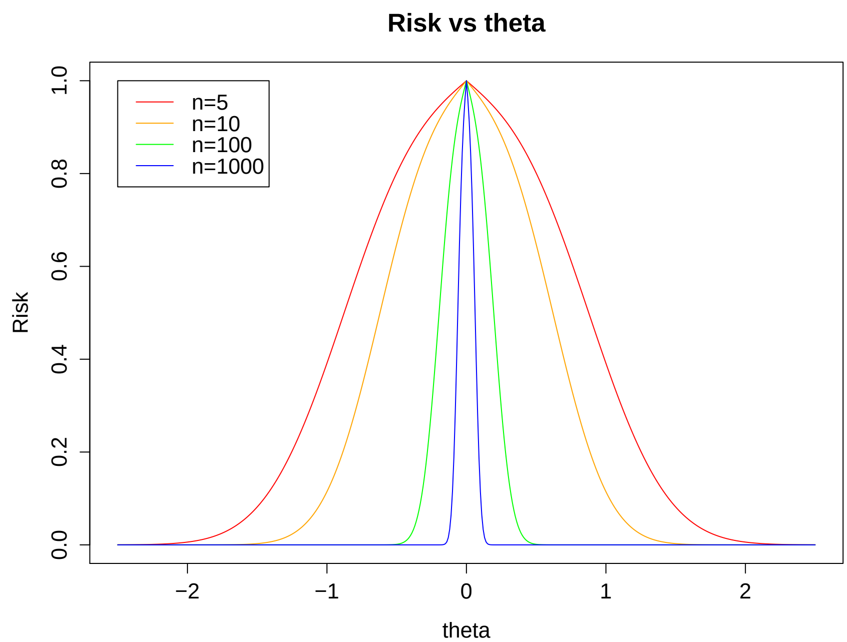

Figure 3: Plots of the risk function for the decision rule based on the z-test with $n \in \{5, 10, 100, 1000\}$.

For this decision rule, increasing n yields a decision rule with uniformly lower risk (except at $\theta = 0$). This is different than the decision rule from part (a) in which we saw that increasing n was detrimental in the range of θ values given by $(-c, c)$.

■

4.2 Suppose we want to estimate θ under a strictly convex loss function. Let X be the data and let T_1, T_2 be any two sufficient statistics such that $T_2 = g(T_1)$ for some known function g. Let $\mathcal{C}_1, \mathcal{C}_2$ be the classes of estimators that are functions of T_1 and T_2, respectively.

4.2.a Show that both $\mathcal{C}_1$ and $\mathcal{C}_2$ are complete classes.

Let δ be an arbitrary estimator in $\mathcal{C}_1^c$. To show that $\mathcal{C}_1$ is complete, it is enough to show that δ is inadmissible. Since $\delta \notin \mathcal{C}_1$, δ is not a function of T_1. Thus, by the sufficiency of T_1 and Exercise 3.1, δ is inadmissible, as desired. An analogous argument shows that $\mathcal{C}_2$ is complete. ■

4.2.b Show that $\mathcal{C}_2 \subset \mathcal{C}_1$.

We need only show that $\delta \in \mathcal{C}_2 \Rightarrow \delta \in \mathcal{C}_1$. Take an arbitrary $\delta \in \mathcal{C}_2$. Our goal is to show that δ is a function of T_1.

By definition, since δ is in $\mathcal{C}_2$, there is an h such that $\delta(X) = h(T_2(X)) = h(g(T_1(X)))$. Thus, letting $f \equiv h \circ g$, we have $\delta = h(T_1(X))$, so $\delta \in \mathcal{C}_1$, as desired. ■

4.2.c Show that if $\delta \in \mathcal{C}_1 \setminus \mathcal{C}_2$, then δ is inadmissible.

We already know that both $\mathcal{C}_1$ and $\mathcal{C}_2$ are complete classes. If we have $\delta \in \mathcal{C}_1 \setminus \mathcal{C}_2$, then we must have $\delta \notin \mathcal{C}_2$, and since δ is not in the complete class $\mathcal{C}_2$, it must be inadmissible. ■

4.2.d Based on the result, what sort of sufficient statistic should be used to construct an estimator?

Given two statistics on which to base complete classes, we'd prefer to look at the one which leads to the smaller complete class. That is, we'd like to look at a sufficient statistic which is a function of other sufficient statistics, if possible. By definition, a *minimal* sufficient statistic is a function of all other sufficient statistics, so we should construct our estimator based on a minimal sufficient statistic whenever possible. ■

4.4 Let $X_j \sim N(\theta_j, 1)$, $j = 1, 2$, and let $L((\theta_1, \theta_2), d) = (\theta_1 - d)^2$. Show that $\delta((X_1, X_2)) = \text{sign}(X_2)$ is an admissible procedure, and explain this counterintuitive result.

Assume δ' is another estimator which is at least as good as δ. If we can show that $\delta' = \delta$ a.e., then it must be the case that δ is not dominatable and therefore δ is admissible.

First, consider the risk of δ.

$$\begin{aligned} R((\theta_1, \theta_2), \delta) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (\theta_1 - \text{sign}(x_2))^2 p(x_1|\theta_1) p(x_2|\theta_2)\, dx_1\, dx_2 \\ &= \int_{-\infty}^{\infty} (\theta_1 - \text{sign}(x_2))^2 p(x_2|\theta_2)\, dx_2, \end{aligned}$$

where $p(x|\mu)$ is the density function of the $N(\mu, 1)$ distribution.

Similarly, for δ', we have the risk function

$$\begin{aligned}R((\theta_1, \theta_2), \delta') &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (\theta_1 - \delta'(x_1, x_2))^2 p(x_1|\theta_1) p(x_2|\theta_2)\, dx_1\, dx_2 \\ &= \int_{-\infty}^{\infty} h(x_2|\theta_1) p(x_2|\theta_2)\, dx_2,\end{aligned}$$

where $h(x_2|\theta_1)$ is defined to be $\int_{-\infty}^{\infty} (\theta_1 - \delta'(x_1, x_2))^2 p(x_1|\theta_1)\, dx_1$.

Now, since δ' is at least as good as δ, we have $R((\theta_1, \theta_2), \delta') \leq R((\theta_1, \theta_2), \delta)$ for all values of θ_1 and θ_2. Equivalently,

$$\frac{R((\theta_1, \theta_2), \delta')}{R((\theta_1, \theta_2), \delta)} \leq 1.$$

Consider what happens to the risk ratio at $\theta_1 = 1$ as we let θ_2 grow without bound. At $\theta_1 = 1$, we have

$$\begin{aligned}\frac{R((\theta_1, \theta_2), \delta')}{R((\theta_1, \theta_2), \delta)} &= \frac{\int_{-\infty}^{\infty} h(x_2|1) p(x_2|\theta_2)\, dx_2}{\int_{-\infty}^{\infty} (1 - \text{sign}(x_2))^2 p(x_2|\theta_2)\, dx_2} \\ &= \frac{\int_{-\infty}^{\infty} h(x_2|1) p(x_2|\theta_2)\, dx_2}{\int_{-\infty}^{0} 4 \cdot p(x_2|\theta_2)\, dx_2} \\ &\geq \frac{\int_{0}^{\infty} h(x_2|1) p(x_2|\theta_2)\, dx_2}{\int_{-\infty}^{0} 4 \cdot p(x_2|\theta_2)\, dx_2}.\end{aligned}$$

By an exponential family property, this ratio grows without bound as θ_2 goes to infinity unless $h(x_2|1) = 0$ almost everywhere on $x_2 > 0$. If the ratio grows without bound, we have a contradiction with the inequality

$$\frac{R((\theta_1, \theta_2), \delta')}{R((\theta_1, \theta_2), \delta)} \leq 1,$$

so it must be the case that $h(x_2|1) = 0$ almost everywhere on $x_2 > 0$. By the definition of h, this requires $(1 - \delta'(x_1, x_2))^2 p(x_1|\theta_1)$ to be 1 a.e. on $x_2 > 0$, or equivalently, $\delta' = 1$ a.e. on $x_2 > 0$.

A similar argument considering $\theta_1 = -1$ and letting θ_2 go to $-\infty$ shows that $\delta' = -1$ a.e. on $x_2 < 0$. Together, this yields $\delta' = \delta$ a.e., which shows that δ is admissible.

Although δ is an absurd estimator for most parts of the (θ_1, θ_2) parameter space, it performs exceptionally well when $\theta_1 = 1$ and $\theta_2 \gg 0$ and also when $\theta_1 = -1$ and $\theta_2 \ll 0$—well enough that the only estimator which has any chance of competing is forced to be 1 when $x_2 > 0$ and -1 when $x_2 < 0$. ■