Some non-Bayesian optimality Thy for SELE of $\gamma(\theta)$
- consider an estimator $\delta(X)$

$$R(\theta,\delta) = E_\theta(\delta(X)-\gamma(\theta))^2$$
$$= MSE_\theta(\delta(X)) = (E_\theta\delta(X)-\gamma(\theta))^2 + Var_\theta\,\delta(X)$$

we'll consider 3 kinds of results

1. Rao-Blackwell (sufficiency + SELE)
2. Lehmann-Scheffe (Best unbiased estimation)
3. Cramér-Rao ("Information Inequalities"/ lower bound on $Var_\theta\,\delta(X)$)

1st Rao-Blackwell

<u>Thm</u> (Rao-Blackwell) For $\delta(X)$ an estimator of $\gamma(\theta)\in\mathbb{R}$ with $E_\theta|\delta(X)|<\infty\ \forall\,\theta$ and $T(X)$ sufficient for θ. Let

$$\delta^*(t) = E\left[\delta(X)\,|\,T(X)=t\right].$$

Then $\delta^*(T(X))$ is an estimator of $\gamma(\theta)$ with

$$MSE_\theta(\delta^*(T(X)) \le MSE_\theta(\delta(X))\quad\forall\,\theta$$

$$\left\{\begin{array}{l}\text{and}\quad MSE_\theta(\delta^*(T(X)) < MSE_\theta(\delta(X))\\ \text{for any } \theta \text{ for which } E_\theta(\delta(X))<\infty \text{ and}\\ \quad P_\theta\left[\delta(X)\ne\delta^*(T(X))\right]>0\end{array}\right\}$$

This theorem says that if the averaging in $E[\delta(X)\,|\,T(X)=t]$ does anything nontrivial we can actually improve on $\delta(X)$! — note BTW that $T(X)=X$ doesn't help - in fact, for maximum averaging/maximal improvement one would want to condition on a <u>minimal</u> sufficient statistic

---

<u>Pf</u> To begin, note that $T(X)$ sufficient implies that

$$E_\theta\left[\delta(X) \mid T(X)=t\right]$$

doesn't depend upon t, so that $\delta^*(T(X))$ really is a statistic/estimator ... the definition makes sense.

Then

$$E_\theta(\delta(X)-\gamma(\theta))^2 = E_\theta E\left[(\delta(X)-\gamma(\theta))^2 \mid T(X)\right]$$

Jensen's inequality since $g(\cdot)=(\cdot-\gamma(\theta))^2$ is convex in $\cdot$

$$\geq E_\theta\left(E[\delta(X)\mid T(X)]-\gamma(\theta)\right)^2$$

$$= E_\theta(\delta^*(T(X))-\gamma(\theta))^2$$

$$= MSE_\theta(\delta^*(T(X))).$$

$\{\ \}$ follows from the strict version of Jensen since $g(\cdot)$ is <u>strictly</u> <u>convex</u> in $\cdot$. ∎

<u>Example</u> $X_1, X_2, \ldots, X_n$ iid Ber(p)

$$T(X)=\sum_{i=1}^n X_i \qquad \delta(X)=X_1 \qquad \left(\text{a stupid estimator of } p\right)$$

Given that $T(X)=t$ the conditional dsn of X is uniform on

$$A\left\{x\in\{0,1\}^n \mid T(x)=t\right\}$$

So given $T(X)=t$, a fraction $\frac{t}{n}$ of the elements of this set have $x_1=1$, i.e.

$$E\left[X_1 \mid T(X)=t\right] = \frac{t}{n}1 + \frac{(n-t)}{n}0 = \frac{t}{n}$$

i.e.

$$\delta^*(t) = \frac{t}{n} \quad \text{and} \therefore \quad \delta^*(T(X)) = \frac{\sum X_i}{n}$$

is at least as good as X_1 for any p. — As long as $p \in (0,1)$ there is positive probability that

$$\frac{\sum X_i}{n} \neq X_1$$

and for such p

$$\underbrace{MSE_p\left(\frac{\sum X_i}{n}\right)}_{\frac{p(1-p)}{n}} < \underbrace{MSE_p(X_1)}_{p(1-p)}$$

This is a pretty encouraging example, but lest we get too enthusiastic, the next example shows that even conditioning on a minimal sufficient statistic doesn't have to produce something sensible

<u>Example</u> $X_1, X_2, \ldots, X_n$ iid $N(\theta, 1)$

Consider SELE of $\gamma(\theta) = E_\theta X_1^2 = \theta^2 + 1$

$\sum X_i$ is minimal sufficient for θ — MOM sets

$$\delta(x) = \frac{1}{n}\sum X_i^2$$

$$= \frac{1}{n}\sum(X_i - \bar{X})^2 + \bar{X}^2$$

independent

So

day 23

$$E\left[\delta(x) \mid \sum X_i\right] = \frac{1}{n} E\left[\sum(X_i-\bar{X})^2 \mid \sum X_i\right] + E\left[\bar{X}^2 \mid \sum X\right]$$

$$\sum (X_i - \bar{X})^2 \sim \chi^2_{n-1} \quad \text{so} \quad E\left[\sum (X_i - \bar{X})^2\right] = n-1$$

and then

$$E\left[\delta(X) \mid \sum X_i\right] = \frac{n-1}{n} + \bar{X}^2$$

$$= 1 + \bar{X}^2 - \frac{1}{n}$$

This is indeed better than $\frac{1}{n}\sum X_i^2$ but is still not sensible, since $\bar{X}^2 - \frac{1}{n}$ can be negative. A better one is

$$\delta^{**}(X) = \begin{cases} 1 + \bar{X}^2 - \frac{1}{n} & \text{if } \bar{X}^2 > \frac{1}{n} \\ 1 & \text{otherwise} \end{cases}$$

Rao Blackwellization can produce "best" estimators in some circumstances — δ best means that for any other δ'

$$MSE_\theta(\delta(X)) \le MSE_\theta(\delta'(X)) \quad \forall \theta$$

(The risk function for δ is always at or below that for $\delta'(X)$)

Unqualified, this is too much to hope for (if I consider all estimators)

<u>Example</u> $X \sim N(\theta, 1)$ $\quad \gamma(\theta) = \theta$

$\delta'(X) = 13$ has 0 risk/MSE if $\theta = 13$

$$MSE_\theta(\delta'(X)) = (13-\theta)^2 \quad \left(\text{squared bias of } \delta'(X)\right)$$

$$\delta''(X) = 17 \text{ has } 0 \text{ risk / MSE if } \theta = 17$$

$$MSE_\theta(\delta''(X)) = (17-\theta)^2$$

etc. – So an overall "best" estimator would need to have $MSE_\theta(\delta(X)) = 0 \;\; \forall \theta$ – It would need to be omniscient/perfect – That's impossible

So instead of "best" overall we have to settle for

"best in some (reasonable) class of estimators"

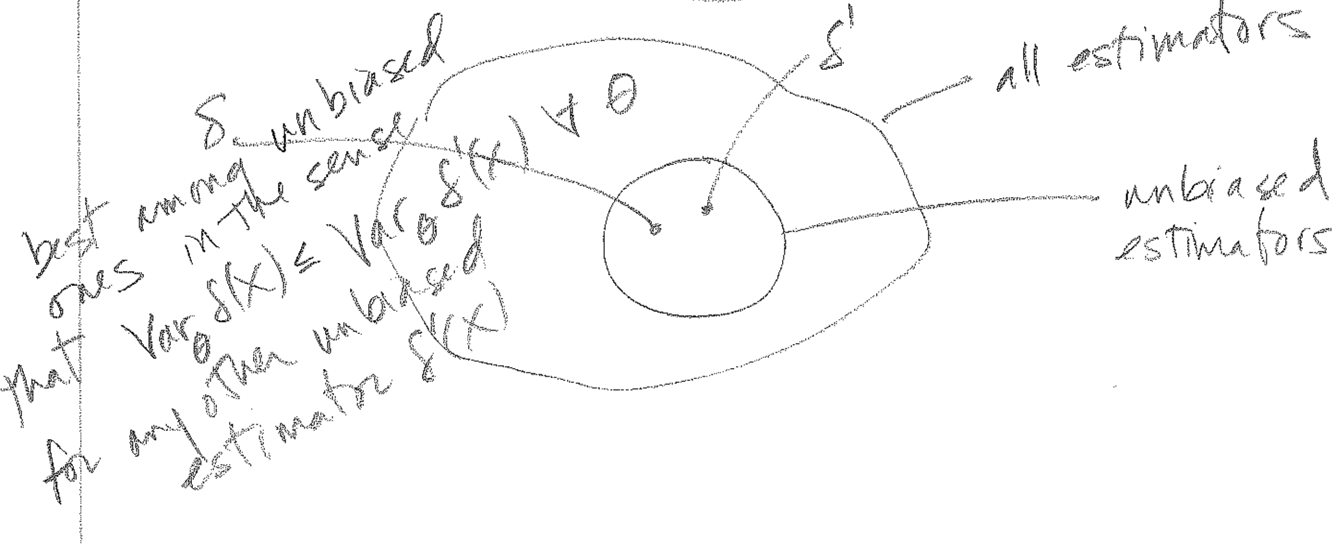
– above δ' and δ'' would be eliminated from consideration as being "unreasonable" – I can, say, restrict attention to <u>unbiased</u> estimators

<u>Def</u> $\delta(X)$ is called an unbiased estimator of $\gamma(\theta)$ provided

$$E_\theta \, \delta(X) = \gamma(\theta) \quad \forall \theta$$

(Note that for unbiased estimators $MSE_\theta(\delta(X)) = Var_\theta \, \delta(X) \quad \forall \theta$)

We can seek "best unbiased" estimators — <u>u</u>niformly (in θ) <u>m</u>inimum <u>v</u>ariance <u>u</u>nbiased



Now consider a "Lehmann-Scheffé" type Thm (not the version usually stated in 543) – The idea here is that Rao-Blackwellization of an unbiased estimator can in the proper context produce a <u>U</u>niformly <u>M</u>inimum <u>V</u>ariance <u>U</u>nbiased <u>E</u>stimator

exp family with T natural sufficient statistic and non-empty interior of parameter space $\Longrightarrow$ "completeness" of sufficient T $\overset{\text{Bahadur}}{\Longrightarrow}$ minimality of T

$\Downarrow$ ← usual "Lehmann-Scheffé"

see handout

Rao-Blackwellization with T gives UMVUE

I'm going to call this a Lehmann-Scheffé result

<u>Theorem</u> In an exponential family with natural sufficient statistic $T(X) = (\sum T_1(X_i), \sum T_2(X_i), \ldots, \sum T_k(X_i))$ (n iid X_i's) suppose $\Xi^* \subset \Xi$ contains an open rectangle. If $\delta(X)$ is unbiased for $\gamma(\theta)$ and

$$\delta^*(t) = E[\delta(X) \mid T(X) = t],$$

then $\delta^*(T(X))$ is a UMVUE of $\gamma(\theta)$

Further, if $\text{Var}_\theta\, \delta^*(T(X)) < \infty \;\; \forall\, \theta$, $\delta^*(T(X))$ is unique.

(If $\delta'(X)$ is another UMVUE of $\gamma(\theta)$

$$P_\theta[\delta^*(T(X)) = \delta'(X)] = 1 \quad \forall\, \theta\;)$$