How are Rao-Blackwell estimators "better"?

In #7(a) on the fourth problem set we see how a horrible flaw in an estimator is remedied by the Rao-Blackwell process. That is one respect in which that particular Rao-Blackwell estimator is better than the flawed estimator in that problem.

Sometimes the goodness of an estimator δ of an unobservable quantity θ is measured by the smallness of its mean squared error $\mathbf{E}((\delta - \theta)^2)$. Another respect in which Rao-Blackwell estimators are better than the estimators upon which they improve, is that they typically have smaller mean squared errors, and the never have bigger ones (the M.S.E. in some cases remains the same rather than getting smaller, however).

To the question

"Which is bigger: $\mathbf{E}(Y^2)$ or $(\mathbf{E}(Y))^2$?"

recall that the answer follows from the observation that

$$\mathbf{E}(Y^2) - (\mathbf{E}(Y))^2 = \mathrm{var}(Y)$$

and $\mathrm{var}(Y)$ necessarily is ≥ 0.

Armed with this observation, let us examine the mean squared error of a Rao-Blackwell estimator:

The "crude" estimator is $\delta(X)$.

The Rao-Blackwell estimator is $\delta_0(X) = \mathbf{E}(\delta(X) \mid T(X))$

The Rao-Blackwell estimator's mean squared error is

$$\begin{aligned}
\mathbf{E}((\delta_0(X) - \theta)^2) &= \mathbf{E}((\mathbf{E}(\delta(X) \mid T(X)) - \theta)^2) \\
&= \mathbf{E}((\mathbf{E}(\delta(X) - \theta \mid T(X)))^2) \text{ (since } \theta \text{ is constant)} \\
&\leq \mathbf{E}(\mathbf{E}((\delta(X) - \theta)^2 \mid T(X))) \text{ (since } (\mathbf{E}(Y))^2 \leq \mathbf{E}(Y^2)) \\
&= \mathbf{E}((\delta(X) - \theta)^2) \text{ (since } \mathbf{E}(\mathbf{E}(U \mid V)) = \mathbf{E}(U)) \\
&= \text{the "crude" estimators mean squared error.}
\end{aligned}$$

Summary:

The R-B estimator's M.S.E. $\leq$ the "crude" estimator's M.S.E.

This bottom-line summary is the Rao-Blackwell Theorem.

When is a Rao-Blackwell estimator the "best" estimator? The answer to that involves the concept of "completeness."

Completeness

Suppose $X_1, \ldots, X_n \sim$ i. i. d. $N(\mu, 1^2)$. Let $\overline{X}_n = (X_1 + \cdots + X_n)/n$. Observe that

- $X_1 - \overline{X}_n$ depends only on the "data" $X_1, \ldots, X_n$ and not on the unobservable μ, i.e., $X_1 - \overline{X}_n$ is a statistic.
- $\mathbf{E}(X_1 - \overline{X}_n) = 0$ **regardless of the value of** μ. Changing the value of the unobservable μ does not change the fact that the expectation of this statistic is zero.

In other words $X_1 - \overline{X}_n$ is an **"unbiased estimator of zero."**

Suppose $W_1, \ldots, W_9 \sim$ i. i. d. Uniform$(\theta, \theta{+}1)$. Let $D = \max\{W_1, \ldots, W_9\} - \min\{W_1, \ldots, W_9\}$. It can be shown that $\mathbf{E}(D) = 0.8$ **regardless of the value of** θ. And the value of D depends only on the data and not on the unobservable θ, so D is a statistic. This statistic is not an unbiased estimator of zero, but $g(D) = D - 0.8$ is an unbiased estimator of zero. And $h(W_1) = \sin(2\pi W_1)$ is also an unbiased estimator of zero.

A locution about to be defined allows us to encapsulate the information above in these simple statements:

- $X_1 - \overline{X}_n$ is not a complete statistic.
- D is not a complete statistic.
- W_1 is not a complete statistic.

DEFINITION: A statistic U is complete iff there is no function g such that $g(U)$ is an unbiased estimator of zero (except, of course, the indentically zero function $g \equiv 0$).

LEHMANN-SCHEFFE THEOREM: If $T(X)$ is a **complete** sufficient statistic for θ and $\delta(X)$ is an unbiased estimator of θ then $\delta_0(X) = \mathbf{E}(\delta(X) \mid T(X))$ is an unbiased estimator of θ that has a smaller mean squared error than any other unbiased estimator of θ.

PROOF: Suppose $\gamma(X)$ is some other unbiased estimator of θ. Let $\gamma_0(X) = \mathbf{E}(\gamma(X) \mid T(X))$. Then $\gamma_0(X) - \delta_0(X)$ is an unbiased estimator of zero. Because of completeness, we must therefore have $\gamma_0 = \delta_0$, i.e., they are both the same estimator. In other words, the Rao-Blackwell process, which by the Rao-Blackwell theorem improves any unbiased estimator, will always yield the same result no matter which estimator we start with.

EXAMPLE: Suppose $X_1, \ldots, X_n \sim$ i. i. d. $N(\mu, 1^2)$. Then there are many functions of $X_1, \ldots, X_n$ that are unbiased estimators of μ (e.g., ...). But $\overline{X}_n$ is a complete, sufficient, and unbiased. Therefore $\overline{X}_n$ is the best unbiased estimator of μ.

Q: How do we know $\overline{X}_n$ is complete?

A: Think about two-sided Laplace transforms.......