Online Appendices

C Additional Analysis of Sufficient Operational Statistics

Additional Result About Sufficiency. There is an additional way to identify sufficient operational statistics based on ancillary and complete statistics. However, this result is restrictive and mostly covered by Theorem 1. Specifically, given a random sample $\boldsymbol{Y}$ from a distribution parameterized by θ, (1) a statistic S is said to be ancillary if its distribution does not depend on θ; (2) a statistic T is said to be complete if for every measurable function $g(\cdot)$, $\mathbb{E}_\theta(g(T)) = 0$ for all θ implies $P_\theta(g(T) = 0) = 1$ for all θ.

As shown in Lemma 1, for a random sample $\boldsymbol{Y}$ parameterized by an unknown scale parameter θ, a condition leading to the sufficiency of $t(\cdot)$ is that $t(\boldsymbol{Y})$ is independent of $\boldsymbol{Y}/t(\boldsymbol{Y})$. Clearly, the statistics $\boldsymbol{Y}/t(\boldsymbol{Y})$ are ancillary given the scale parameter. Thus, by Basu's Theorem, if $t(\boldsymbol{Y})$ is a complete sufficient statistic, then $t(\boldsymbol{Y})$ is independent of $\boldsymbol{Y}/t(\boldsymbol{Y})$ and hence is a sufficient operational statistic.

While this result appears promising, its applicability is in fact limited. The limitation is due to the completeness with a scale parameterization (note that the factorization rule in Theorem 1 itself does not require a scale parameterization although the decision problem assumes it). First, identifying a complete statistic by using its definition is difficult, except in some special cases. Even in those special cases, the parameterization is often not a scale parameterization. In statistical theory, the most commonly used result regarding completeness is with respect to the exponential family: If each observation Y_i in the sample $\boldsymbol{Y}$ has a density function of the form $f_{Y_i}(y_i) = h(y_i)c(\theta)e^{w(\theta)t(y_i)}$, then $\sum_{i=1}^n t(Y_i)$ is a complete statistic (adapted from Theorem 6.2.25 in Casella and Berger 2002 to fit our context). However, the scale parameterization restricts the applicability of this result: given the scale parameterization, Y_i has a density of the form $f_{Y_i}(y_i) = (1/\theta)f_Z(y_i/\theta)$ for some base distribution $f_Z(\cdot)$. Further, if a function of the form $u(x/\theta)$ can be multiplicatively decomposed as $u(x/\theta) = v(x)w(\theta)$, then this relationship is equivalent to $u(x/\theta)u(1) = u(x)u(1/\theta)$. Functions satisfying this relationship are *completely multiplicative functions*, which primarily concern power functions, i.e., $u(x) = x^a, a > 0$ (completely multiplicative functions also include some special functions like Liouville function, Dirichlet characters, Jacobi symbol and Legendre symbol, but these are not commonly used in practice). Because power functions are homogeneous, such cases are already covered by Theorem 1.

In summary, although this characterization of sufficient operational statistics based on ancillary and complete statistics may cover some additional special cases, most of the commonly used cases have already been covered by Theorem 1. Theorem 1 is much more general than this result. We thus do not focus on this case. $\square$

Proof of Auction Example. From Theorem 1 and Equation (3), $r^*(\hat{\boldsymbol{x}}) = a^* t(\hat{\boldsymbol{x}})$, where a^* is the maximum solution to the problem

$$\mathbb{E}\left[Z_{[2]}\mathbb{I}(Z_{[2]} > at(\hat{\boldsymbol{Z}})) + at(\hat{\boldsymbol{Z}})\mathbb{I}(Z_{[2]} \le at(\hat{\boldsymbol{Z}}) \le Z_{[1]})\right].$$

Its first-order condition can be shown to be:

$$\mathbb{E}_{t(\hat{\boldsymbol{Z}})}\left\{t(\hat{\boldsymbol{Z}})m\left[1 - F_Z(at(\hat{\boldsymbol{Z}})) - at(\hat{\boldsymbol{Z}})f_Z(at(\hat{\boldsymbol{Z}}))\right]F_Z(at(\hat{\boldsymbol{Z}}))^{m-1}\right\} = 0. \tag{5}$$

This condition applies to any value distribution. We next apply this condition to the uniform distribution case, where the base value Z follows a standard uniform distribution, $f_Z(z) = 1$ and $F_Z(z) = z$ for $z \in [0, 1]$. When $a \le 1$, this condition becomes:

$$\begin{aligned}
\mathbb{E}_{t(\hat{\boldsymbol{Z}})}\left\{t(\hat{\boldsymbol{Z}})\left[1 - at(\hat{\boldsymbol{Z}}) - at(\hat{\boldsymbol{Z}})\right](at(\hat{\boldsymbol{Z}}))^{m-1}\right\} &= 0\\
\mathbb{E}_{t(\hat{\boldsymbol{Z}})}\left[t(\hat{\boldsymbol{Z}})^m\right] - 2a\mathbb{E}_{t(\hat{\boldsymbol{Z}})}\left[t(\hat{\boldsymbol{Z}})^{m+1}\right] &= 0\\
\int_0^1 y^m ny^{n-1}dy - 2a\int_0^1 y^{m+1}ny^{n-1}dy &= 0.
\end{aligned}$$

Here, the last step follows from $t(\hat{\boldsymbol{Z}}) = \max(\hat{\boldsymbol{Z}})$, which has a density of ny^{n-1}. We solve this equation and obtain

$$a^* = \frac{m+n+1}{2(m+n)}. \tag{6}$$

When $a > 1$, in the condition (5), the expression inside $\mathbb{E}_{t(\hat{\boldsymbol{Z}})}\{\}$ is nonzero only when $0 < at(\hat{\boldsymbol{Z}}) < 1$. By assumption, $at(\hat{\boldsymbol{Z}}) > 0$. That $at(\hat{\boldsymbol{Z}}) < 1$ means $t(\hat{\boldsymbol{Z}}) < 1/a$. Thus, (5) becomes

$$\begin{aligned}
\mathbb{E}_{t(\hat{\boldsymbol{Z}})}\left\{t(\hat{\boldsymbol{Z}})\left[1 - at(\hat{\boldsymbol{Z}}) - at(\hat{\boldsymbol{Z}})\right](at(\hat{\boldsymbol{Z}}))^{m-1}\mathbb{I}(t(\hat{\boldsymbol{Z}}) < 1/a)\right\} &= 0\\
\mathbb{E}_{t(\hat{\boldsymbol{Z}})}\left[t(\hat{\boldsymbol{Z}})^m\mathbb{I}(t(\hat{\boldsymbol{Z}}) < 1/a)\right] - 2a\mathbb{E}_{t(\hat{\boldsymbol{Z}})}\left[t(\hat{\boldsymbol{Z}})^{m+1}\mathbb{I}(t(\hat{\boldsymbol{Z}}) < 1/a)\right] &= 0\\
\int_0^{1/a} y^m ny^{n-1}dy - 2a\int_0^{1/a} y^{m+1}ny^{n-1}dy &= 0.
\end{aligned}$$

It is easy to verify that this equation only holds when $a = \infty$, which cannot be the optimal solution. Thus, the solution is given by (6). $\square$

D Additional Details About Mean-Bounded Examples

The examples in Section 3 are standard results, with some technical details provided here. First, consider the sample mean $t(\hat{\boldsymbol{Z}})$, where the sample $\hat{\boldsymbol{Z}}$ is drawn from a sub-Gaussian distribution. By Hoeffding's or Khintchine's inequality, $\mathbb{P}\left(|t(\hat{\boldsymbol{Z}})-1| \geq a\right) \leq 2\exp(-cna^2)$ for all $a \geq 0$ and some $c \geq 0$; here, we have used the fact that centering does not harm the sub-Gaussian property. This means

$$\mathbb{E}\left(|t(\hat{\boldsymbol{Z}})-1|\right) \leq \frac{c}{\sqrt{n}}, \text{ for some } c \geq 0.$$

In addition, if $\hat{\boldsymbol{Z}}$ comes from a sub-exponential distribution, then by Bernstein's inequality, $\mathbb{P}\left(|t(\hat{\boldsymbol{Z}})-1| \geq a\right) \leq 2\exp\left(-c_1 n \min\{ac_2, a^2c_2^2\}\right)$ for all $a \geq 0$ and some $c_1, c_2 \geq 0$. Thus,

$$\mathbb{E}\left(|t(\hat{\boldsymbol{Z}})-1|\right) \leq \frac{c}{n} + \frac{d}{\sqrt{n}}, \text{ for some } c, d \geq 0.$$

Finally, consider the sample maximum Y, where the sample is drawn from a bounded distribution F_Z with a support of $[0,1]$ and without a probability mass at 1. Here, without loss of generality, we have normalized the support to $[0,1]$. Thus, $F_Y(y) = F_Z(y)^n$ on $[0,1]$. For any $\epsilon \in [0,1]$, $\mathbb{P}(1-Y \geq \epsilon) = F_Z(1-\epsilon)^n$ and for $\epsilon > 1$, $\mathbb{P}(1-Y \geq \epsilon) = 0$. Therefore, $\mathbb{E}|Y-\mathbb{E}(Y)| \leq [1-\mathbb{E}(Y)] + \mathbb{E}(1-Y) = 2\mathbb{E}(1-Y) = 2\int_0^1 F_Z(1-\epsilon)^n d\epsilon$. The right-hand-side clearly decreases in n as $F_Z < 1$ almost surely, and converges to 0 by dominated convergence theorem.

E Location-Scale Parametric Setting

We next consider a location-scale parametric setting, where the random variables in the decision-making problem $\boldsymbol{X} = \tau\boldsymbol{e} + \theta\boldsymbol{Z}$ and the sample $\hat{\boldsymbol{X}} = \tau\boldsymbol{e} + \theta\hat{\boldsymbol{Z}}$, where $\tau \in \mathbb{R}$ and $\theta \in \mathbb{R}^+$ are unknown, $\boldsymbol{Z} = (Z_1, ..., Z_m)$ and $\hat{\boldsymbol{Z}} = (\hat{Z}_1, ..., \hat{Z}_n)$ have known distributions, and, with slight abuse of notation, $\boldsymbol{e}$ is a "dimension filler," a vector of ones to match the dimension in an expression. We focus on decision rules and payoff functions with the following properties.

Definition 2 *A function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be shift-invariant degree-one homogeneous (SH1) if $f(\alpha\boldsymbol{e} + \beta\boldsymbol{x}) = \alpha + \beta f(\boldsymbol{x}) + \Delta$ for $\alpha \in \mathbb{R}$, $\beta \in \mathbb{R}^+$, $\boldsymbol{x} \in \mathbb{R}^n$, and some $\Delta \in \mathbb{R}$ that does not contain $\boldsymbol{x}$. The function is said to be strictly shift-invariant degree-one homogeneous (SSH1) if $\Delta = 0$.*

For example, a major model with an SH1 payoff function is the newsvendor model, and the $\mathscr{H}_1^e$ decision rule class in Chu et al. (2008) contains SSH1 functions. To further our analysis, we again need to identify sufficient operational statistics. Since there are two unknown parameters τ and θ, we generally need two statistics, $t_\tau(\cdot)$ and $t_\theta(\cdot)$, and represent the decision as a function of the two statistics. As discussed in the main text, such a representation only preserves sufficiency under certain conditions. The basic idea to derive the conditions is to first normalize the sample by subtracting a location statistic of the sample. The resultant sample then has only a scale structure and is solvable by Theorem 2. Formally, suppose $t_\tau(\cdot)$ is an SSH1 function, and let $\hat{\boldsymbol{Y}} = \hat{\boldsymbol{X}} - t_\tau(\hat{\boldsymbol{X}})\boldsymbol{e}$ be the normalized sample and $\hat{\boldsymbol{B}} = \hat{\boldsymbol{Z}} - t_\tau(\hat{\boldsymbol{Z}})\boldsymbol{e}$ be the normalized base sample. Then,

$$\hat{\boldsymbol{Y}} = \hat{\boldsymbol{X}} - t_\tau(\hat{\boldsymbol{X}})\boldsymbol{e} = \tau\boldsymbol{e} + \theta\hat{\boldsymbol{Z}} - t_\tau(\tau\boldsymbol{e} + \theta\hat{\boldsymbol{Z}})\boldsymbol{e} = \tau\boldsymbol{e} + \theta\hat{\boldsymbol{Z}} - \left[\tau + \theta t_\tau(\hat{\boldsymbol{Z}})\right]\boldsymbol{e} = \theta\left[\hat{\boldsymbol{Z}} - t_\tau(\hat{\boldsymbol{Z}})\boldsymbol{e}\right] = \theta\hat{\boldsymbol{B}},$$

which has a scale structure. We then apply Theorem 2 to $\hat{\boldsymbol{Y}}$ and $\hat{\boldsymbol{B}}$ to obtain the solution.

We say a function is of an exponential form if $f(x) = c\prod_i d_i^{a_i x + b_i}$ for some a_i, b_i, c, d_i (e.g., ce^{-ax} and $p^{a_1 x}(1-p)^{a_2 x + b}$). Then

Proposition 1 *If the sample is drawn from a density/mass function of an exponential form, then any SSH1 statistic is a sufficient operational statistic for* τ*.*

Proposition 2 *In a problem with a location-scale parametric setting, if the payoff function is SH1 and the decision rule is SSH1, then* $r^*(\hat{\boldsymbol{x}}) = a^* t_\theta(\hat{\boldsymbol{x}} - t_\tau(\hat{\boldsymbol{x}})\boldsymbol{e}) + t_\tau(\hat{\boldsymbol{x}})$*, where*

$$a^* = \arg\max_a \mathbb{E}\left[\phi\left(at(\hat{\boldsymbol{B}}), \boldsymbol{B}\right)\right].$$

In the first proposition, it is easy to find an SSH1 function of data (e.g., the sample mean). Any of such functions can be used as a sufficient operational statistic for τ. The density/mass function needs to have an exponential form, which covers some distributions, but is more restrictive than the factorization rule in Theorem 1.

Proofs. The following result is necessary: For a random sample $\hat{\boldsymbol{Z}}$ drawn from a density/mass function of an exponential form, if $t(\cdot)$ is an SSH1 statistic, then $t(\hat{\boldsymbol{Z}})$ is independent of $\hat{\boldsymbol{Z}} - t(\hat{\boldsymbol{Z}})\boldsymbol{e}$.

The proof is similar to that of Lemma 1. Consider the transformation from $\hat{\boldsymbol{Z}}$ to $\boldsymbol{Y} = [\hat{Z}_1 - t(\hat{\boldsymbol{Z}}), ..., \hat{Z}_{n-1} - t(\hat{\boldsymbol{Z}}), t(\hat{\boldsymbol{Z}})]$. Clearly, $\hat{z}_i = y_i + y_n$ for $i = 1, ..., n-1$, and

$$\begin{aligned}\hat{z}_n &= \arg_{\hat{z}}[t(\hat{z}_1, ..., \hat{z}_{n-1}, \hat{z}) = y_n] = \arg_{\hat{z}}[t(\hat{z}_1 - y_n, ..., \hat{z}_{n-1} - y_n, \hat{z} - y_n) = 0] \\ &= \arg_w[t(y_1, ..., y_{n-1}, w) = 0] + y_n = s(y_1, ..., y_{n-1}) + y_n,\end{aligned}$$

where $s(\cdot)$ represents some function of $y_1, ..., y_{n-1}$, and the second equality holds since $t(\cdot)$ is SSH1.

Thus, the Jacobian determinant

$$\det(J) = \begin{vmatrix} 1 & 0 & \cdots & 1 \\ 0 & 1 & \cdots & 1 \\ \vdots & \vdots & \vdots & \vdots \\ s_1(y_1, ..., y_{n-1}) & s_2(y_1, ..., y_{n-1}) & \cdots & 1 \end{vmatrix},$$

where $s_i(\cdot)$ is the derivative of $s(\cdot)$ with respect to its ith argument. By the Leibniz formula and following the steps in the proof of Lemma 1, we have

$$\det(J) = 1 - \sum_{i=1}^{n-1} s_i(y_1, ..., y_{n-1}).$$

Thus, the joint pdf of $\boldsymbol{Y}$ is:

$$\begin{aligned} f_{\boldsymbol{Y}}(\boldsymbol{y}) =& f_{\hat{\boldsymbol{Z}}}(y_1 + y_n, ..., y_{n-1} + y_n, s(y_1, ..., y_{n-1}) + y_n)|\det(J)| \\ =& h(y_1, ..., y_{n-1}, s(y_1, ..., y_{n-1}))g(y_n)|\det(J)|, \end{aligned}$$

where second equality holds by the exponential form of the density function. $f_{\boldsymbol{Y}}(\boldsymbol{y})$ is then multiplicatively separable in $(y_1, ..., y_{n-1})$ and y_n. A same argument as that in the proof of Lemma 1 concludes that $\hat{\boldsymbol{Z}} - t(\hat{\boldsymbol{Z}})\boldsymbol{e}$ is independent of $t(\hat{\boldsymbol{Z}})$.

Now we prove the propositions. As defined, $\hat{\boldsymbol{Y}} = \hat{\boldsymbol{X}} - t_\tau(\hat{\boldsymbol{X}})\boldsymbol{e}$ and $\hat{\boldsymbol{B}} = \hat{\boldsymbol{Z}} - t_\tau(\hat{\boldsymbol{Z}})\boldsymbol{e}$ are the normalized sample and base sample, and $\boldsymbol{Y} = \boldsymbol{X} - t_\tau(\hat{\boldsymbol{X}})\boldsymbol{e}$ and $\boldsymbol{B} = \boldsymbol{Z} - t_\tau(\hat{\boldsymbol{Z}})\boldsymbol{e}$ are the normalized random variables and base random variables. By the aforementioned result, $\boldsymbol{Y}$ and $\hat{\boldsymbol{Y}}$ are independent. In addition, the transformed sample and random variables have a scale structure. Then,

$$\begin{aligned} \max_{r(\cdot)} \mathbb{E}\left[\phi\left(r(\hat{\boldsymbol{X}}), \boldsymbol{X}\right)\right] =& \max_{r(\cdot)} \mathbb{E}\left[\phi\left(r\left(\hat{\boldsymbol{X}} - t_\tau(\hat{\boldsymbol{X}})\boldsymbol{e}\right), \boldsymbol{X} - t_\tau(\hat{\boldsymbol{X}})\boldsymbol{e}\right) + t_\tau(\hat{\boldsymbol{X}}) + \Delta\right] \\ =& \max_{r(\cdot)} \mathbb{E}\left[\phi\left(r\left(\hat{\boldsymbol{Y}}\right), \boldsymbol{Y}\right)\right] + \mathbb{E}\left[t_\tau(\hat{\boldsymbol{X}}) + \Delta\right], \end{aligned}$$

where the first equality holds because $r(\cdot)$ is SSH1 and $\phi(\cdot)$ is SH1. By Theorem 2, the solution of the transformed problem is: $a^* t_\theta(\hat{\boldsymbol{y}})$, where $a^* = \arg\max_a \mathbb{E}\left[\phi\left(at(\hat{\boldsymbol{B}}), \boldsymbol{B}\right)\right]$. Since by the normalization $r^*(\hat{\boldsymbol{x}}) = r^*(\hat{\boldsymbol{y}}) + t_\tau(\hat{\boldsymbol{x}})$, the propositions hold.

References

Casella, G., R. L. Berger. 2002. *Statistical Inference*. 2nd ed. Duxbury, Pacific Grove, CA.

Chu, L. Y., J. G. Shanthikumar, Z.-J. M. Shen. 2008. Solving operational statistics via a Bayesian analysis. *Operations research letters* **36**(1) 110–116.