Online Appendix to Bamboozled by Bonferroni

Conor Mayo-Wilson

In this online appendix, I prove theorem 1 in the body of the article. To ensure this online appendix is self-contained, some definitions appearing in the body of the article are reproduced here.

1 Basic model

Suppose N hypotheses are under investigation, and let $\Theta = \{0,1\}^N$ be the set of all binary strings of length N. A vector $\theta \in \Theta$ specifies which of the N hypotheses are true. For each $k \leq N$, let $H_k = \{\theta \in \Theta : \theta_k = 0\}$ be the set of vectors that say the kth hypothesis is true. For each $k \leq N$, let X_k be a random variable representing an experiment. For each $\theta \in \Theta$, let $\mathbb{P}_\theta(X_1, \ldots, X_N)$ denote the probability measure that specifies the chances of various experimental outcomes.

We assume that for all $\theta \in \Theta$, the N experiments are *mutually independent* with respect to $\mathbb{P}_\theta$. In symbols, let $\vec{X} = \langle X_{i_1}, X_{i_2}, \ldots, X_{i_k} \rangle$ be a random vector, representing some subset of the N experiments. Then:

$$\mathbb{P}_\theta(\vec{X} = \vec{x}) = \prod_{j \leq k} \mathbb{P}_\theta(X_{i_j} = x_{i_j}) \tag{1}$$

for all $\vec{x} = (x_{i_1}, \dots x_{i_k})$. Further, suppose that the truth or falsity of the H_k determines the probabilities of the possible outcomes of the kth experiment; that is, for all $k \leq N$ and all $r \in \{0, 1\}$, there is a probability distribution $\mathbb{P}_{k,r}$ such that $\mathbb{P}_\theta(X_k = x_k) = \mathbb{P}_{k,\theta_k}(X_k = x_k)$. Together with the assumption of mutual independence, this entails that

$$\mathbb{P}_\theta(\vec{X} = \vec{x}) = \prod_{j \leq k} \mathbb{P}_{i_j, \theta_{i_j}}(X_{i_j} = x_{i_j}) \text{ for all } \theta \in \Theta. \tag{2}$$

1.1 Decision adjustment

For each $k \leq n$, let A_k denote a set of *component acts*, and define a *strategy* to be a set S of component acts such that for all k, either $S \cap A_k$ is a singleton or empty. That is, at most, one act can be taken with respect to a hypothesis H_k. A *decision rule* d maps subsets of (values of) the observable variables $X_1, \dots X_N$ to strategies. I require that $d(X_{k_1} = x_{k_1}, \dots, X_{k_m} = x_{k_m})$ contains precisely one element from each of the sets A_{k_m}.

A decision rule d adjusts for multiplicity if there is some x_1 such that

$$d(x_1) \notin d(x_1, \dots x_N) \tag{3}$$

for all values $x_2, \dots x_N$ of $X_2, \dots X_N$.

1.2 Maximin and Baye's rule

Suppose a researcher assigns a utility $u(S, \theta)$ to each strategy S and vector $\theta \in \Theta$ specifying which of the N hypotheses are true. If we fix a vector $\theta \in \Theta$, then the researcher's expected utility (with respect to $\mathbb{P}_\theta$) can be defined straightforwardly, whether she decides to observe one variable or all N variables:[1]

$$\mathbb{E}_\theta^1[d] = \sum_{x_1 \in \mathcal{X}_1} \mathbb{P}_\theta(X_1 = x_1) \cdot u(d(x_1), \theta)$$
$$\mathbb{E}_\theta^N[d] = \sum_{\vec{x} \in \mathcal{X}} \mathbb{P}_\theta(\vec{X} = \vec{x}) \cdot u(d(\vec{x}), \theta).$$

Here, $\mathcal{X}_1$ is the range of X_1, and $\mathcal{X}$ is the range of the random vector $\vec{X} = (X_1, \ldots, X_N)$.

A decision rule d is called *maximin* if $\min_{\theta \in \theta} \mathbb{E}_\theta^j[d] \geq \min_{\theta \in \theta} \mathbb{E}_\theta^j[e]$ for all decision rules e, where $j = 1$ or $j = N$.

Recall that the subjective expected utility of a strategy S with respect to a measure P is given by the following:

$$\mathbb{E}_P[S] := \sum_{\theta \in \Theta} P(\theta) \cdot u(S, \theta). \tag{4}$$

Thus, there is a Bayesian who will adjust for multiplicity if there is a probability measure P, utility function u, and experimental outcomes $\vec{x} = (x_1, \ldots, x_N) \in \mathcal{X}$ such that three conditions hold:

1. $P(\vec{X} = \vec{x}) > 0$;

2. a_1 maximizes $\mathbb{E}_{P(\cdot|X_1 = x_1)}[a]$ over all $a \in A_1$; and

3. $a_1 \notin S$ for some S that maximizes $\mathbb{E}_{P(\cdot|\vec{X} = \vec{x})}[T]$, where T ranges over strategies

[1]For simplicity, I assume all of the sets in this article are finite, including Θ, the ranges of the random variables $X_1, \ldots X_n$, and the range of all decision rules. Under appropriate measure-theoretic assumptions, the sums in the article can be replaced with integrals if one is interested in extending these ideas to continuous spaces.

containing a component act in every A_k.

For simplicity, assume that a decision-maker's utilities are *separable* across component acts in the following sense. Assume that for each hypothesis H_k, there is a "component" utility function $u_k : A_k \times \{0,1\} \to \mathbb{R}$ that specifies the utilities $u(a,0)$ and $u(a,1)$ of taking action $a \in A_k$ when H_k is true and false, respectively. Further, suppose that the utility of a strategy $u(S,\theta)$ in state θ is the sum of the utilities of component acts, that is:

$$u(S,\theta) = \sum_{k \leq N} \sum_{a \in S \cap A_k} u_k(a, \theta_k). \tag{5}$$

2 Theorem and proof

Theorem 1. Suppose utilities are separable in the sense of equation (5). Then there are maximin rules that do not adjust for multiplicity. If in addition, the hypotheses of Θ are mutually independent with respect to P, then one can maximize (subjective) expected utility with respect to P without adjusting. It follows that if the maximin rule is unique, then no decision rule that adjusts is maximin. Similar remarks apply to expected-utility maximization.

Before proving the theorem, we introduce some notation. Given any decision rule d and $k \leq N$, we define a function $d_k : \mathcal{X} \to A_k$ by $d_k(\vec{y}) := A_k \cap d(\vec{y})$. In other words, d_k

picks out the kth component act from each strategy recommended by d.

$$\begin{aligned}
\mathbb{E}_\theta^N[d] &= \sum_{\vec{y}\in\mathcal{X}} \mathbb{P}_\theta(\vec{y}) \cdot u(d(\vec{y}),\theta) \\
&= (\mathbb{P}_\theta(\vec{x}) \cdot u(d(\vec{x}),\theta)) + \left(\sum_{\vec{y}\neq\vec{x}} \mathbb{P}_\theta(\vec{y}) \cdot u(d(\vec{y}),\theta) \right) \\
&= \left(\sum_{1\leq k\leq N} \mathbb{P}_\theta(\vec{x}) \cdot u_k(d_k(\vec{x}),\theta) \right) + \left(\sum_{\vec{y}\neq\vec{x}} \mathbb{P}_\theta(\vec{y}) \cdot u(d(\vec{y}),\theta) \right) \\
&\quad \text{by separability,} \\
&= (\mathbb{P}_\theta(\vec{x}) \cdot u_1(d_1(\vec{x}),\theta)) + \left(\sum_{1<k\leq N} \mathbb{P}_\theta(\vec{x}) \cdot u_k(d_k(\vec{x}),\theta) \right) + \left(\sum_{\vec{y}\neq\vec{x}} \mathbb{P}_\theta(\vec{y}) \cdot u(d(\vec{y}),\theta) \right).
\end{aligned}$$

Call the first, second, and third summands in the previous equation $T_1(\theta,\vec{x},d)$, $T_2(\theta,\vec{x},d)$, and $T_3(\theta,\vec{x},d)$, respectively.

Proof of Theorem 1: The outline of the proof is identical for both maximin and subjective expected-utility (SEU) maximization. We first pick any decision rule d that is maximin (or maximizes SEU). Such a rule exists because we have assumed all the relevant sets to be finite. If d does *not* adjust for multiplicity, we're done. Otherwise, there is some vector $\vec{x} = (x_1, \ldots x_N)$ such that $d(x_1) \not\in d(\vec{x})$. Define a new decision rule—call it e—such that e is like d in all respects except the following. Let $a_1 \in A_1$ be such that $d(x_1) = \{a_1\}$, and let b_1 be the unique element of $A_1 \cap d(\vec{x})$. Define $e(\vec{x}) = (d(\vec{x}) \setminus b_1) \cup \{a_1\}$. And as we said, define $e(\vec{y}) = d(\vec{y})$ for all $\vec{y} \neq \vec{x}$ (regardless of length). We claim that e is also maximin (or maximizes SEU). By repeating this process some finite number of times, we'll obtain a decision rule that is maximin (or maximizes SEU) and that does not adjust for multiplicity.

First, we consider the case in which d is maximin. Because d itself is maximin, to show that e is maximin, it suffices to show the following:

$$\min_{\theta\in\Theta} \mathbb{E}_\theta^1[e] \geq \min_{\theta\in\Theta} \mathbb{E}_\theta^1[d] \text{ and} \tag{6}$$

$$\min_{\theta\in\Theta} \mathbb{E}_\theta^N[e] \geq \min_{\theta\in\Theta} \mathbb{E}_\theta^N[d]. \tag{7}$$

The first equation follows immediately from the definition of e because $e(x) = d(x)$ for all $x \in \mathcal{X}_1$; that is, the values of e and d do not differ on vectors of length 1. So we need to show only that $\min_{\theta\in\Theta} \mathbb{E}_\theta^N[e] \geq \min_{\theta\in\Theta} \mathbb{E}_\theta^N[d]$.

Using the decomposition described previously, we first show that $T_2(\theta, \vec{x}, d) = T_2(\theta, \vec{x}, e)$ and that $T_3(\theta, \vec{x}, d) = T_3(\theta, \vec{x}, e)$ for all θ and $\vec{x}$.

To show $T_2(\theta, \vec{x}, d) = T_2(\theta, \vec{x}, e)$ for all θ, let θ be arbitrary. Notice first that by the definition of e, we know that $d_k(\vec{y}) = e_k(\vec{y})$ for all $k > 1$ and for all $\vec{y}$ (including $\vec{x}$). It follows that for all θ and all $\vec{y}$,

$$\sum_{1<k\leq N} \mathbb{P}_\theta(\vec{y}) \cdot u_k(d_k(\vec{y}), \theta) = \sum_{1<k\leq N} \mathbb{P}_\theta(\vec{y}) \cdot u_k(e_k(\vec{y}), \theta), \tag{8}$$

which is exactly what $T_2(\theta, \vec{x}, d) = T_2(\theta, \vec{x}, e)$ asserts.

To show $T_3(\theta, \vec{x}, d) = T_3(\theta, \vec{x}, e)$, again note that by definition of e, we know that $d_1(\vec{y}) = e_1(\vec{y})$ for all $\vec{y} \neq \vec{x}$. It follows that

$$\mathbb{P}_\theta(\vec{y}) \cdot u(d_1(\vec{y}), \theta) = \mathbb{P}_\theta(\vec{y}) \cdot u(e_1(\vec{y}), \theta) \text{ for all } \theta \text{ and all } \vec{y} \neq \vec{x}. \tag{9}$$

Equations (9) and (8) together entail the following:

$$\sum_{1\leq k\leq n} \mathbb{P}_\theta(\vec{y}) \cdot u_k(d_k(\vec{y}),\theta) = \sum_{1\leq k\leq n} \mathbb{P}_\theta(\vec{y}) \cdot u_k(e_k(\vec{y}),\theta) \text{ for all } \theta \text{ and } \vec{y} \neq \vec{x}. \tag{10}$$

Because u is separable, equation (10) implies that for all $\vec{y} \neq \vec{x}$,

$$\mathbb{P}_\theta(\vec{y}) \cdot u(d(\vec{y}),\theta) = \mathbb{P}_\theta(\vec{y}) \cdot u(e(\vec{y}),\theta) \text{ for all } \theta \text{ and } \vec{y} \neq \vec{x}. \tag{11}$$

And that immediately entails the following:

$$\sum_{\vec{y}\neq\vec{x}} \mathbb{P}_\theta(\vec{y}) \cdot u(d(\vec{y}),\theta) = \sum_{\vec{y}\neq\vec{x}} \mathbb{P}_\theta(\vec{y}) \cdot u(e(\vec{y}),\theta) \text{ for all } \theta \text{ and } \vec{y} \neq \vec{x}. \tag{12}$$

Notice that the previous equation asserts $T_3(\theta,\vec{x},d) = T_3(\theta,\vec{x},e)$, as desired.

So to show that e is maximin, it therefore suffices to show that $\min_{\theta\in\Theta} T_1(\theta,\vec{x},e) \geq \min_{\theta\in\Theta} T_1(\theta,\vec{x},d)$, where we recall the following:

$$T_1(\theta,\vec{x},e) = \mathbb{P}_\theta(\vec{x}) \cdot u_1(e_1(\vec{x}),\theta), \tag{13}$$

and similarly for $T_1(\theta,\vec{x},d)$.

For the sake of contradiction, suppose that

$$\min_{\theta\in\Theta} \mathbb{P}_\theta(\vec{x}) \cdot u_1(e_1(\vec{x}),\theta) < \min_{\theta\in\Theta} \mathbb{P}_\theta(\vec{x}) \cdot u_1(d_1(\vec{x}),\theta). \tag{14}$$

Because the likelihood function factors, by equation (1), it follows that

$$\min_{\theta\in\Theta}\left(\mathbb{P}_\theta(x_1)\cdot\prod_{k\geq 2}\mathbb{P}_\theta(x_k)\right)\cdot u_1(e_1(\vec{x}),\theta) < \min_{\theta\in\Theta}\left(\mathbb{P}_\theta(x_1)\cdot\prod_{k\geq 2}\mathbb{P}_\theta(x_k)\right)\cdot u_1(d_1(\vec{x}),\theta).$$

That inequality cannot be strict unless $\prod_{k\geq 2}\mathbb{P}_\theta(x_k) > 0$ for at least one θ. It follows that

$$\min_{\theta\in\Theta}\mathbb{P}_\theta(x_1)\cdot u_1(e_1(\vec{x}),\theta) < \min_{\theta\in\Theta}\mathbb{P}_\theta(x_1)\cdot u_1(d_1(\vec{x}),\theta)$$

. Recall that $d_1(\vec{x}) = \{b_1\}$, and so the last equation becomes

$$\min_{\theta\in\Theta}\mathbb{P}_\theta(x_1)\cdot u_1(e_1(\vec{x}),\theta) < \min_{\theta\in\Theta}\mathbb{P}_\theta(x_1)\cdot u_1(b_1,\theta)$$

. By separability, the previous equation entails the following:

$$\min_{\theta\in\Theta}\mathbb{P}_\theta(x_1)\cdot u(e(x_1),\theta) < \min_{\theta\in\Theta}\mathbb{P}_\theta(x_1)\cdot u(\{b_1\},\theta)$$

. And because $e(x_1) = d(x_1)$, we obtain the following:

$$\min_{\theta\in\Theta}\mathbb{P}_\theta(x_1)\cdot u(d(x_1),\theta) < \min_{\theta\in\Theta}\mathbb{P}_\theta(x_1)\cdot u(\{b_1\},\theta)$$

. Now if we add $\sum_{y\in\mathcal{X}_1\setminus\{x_1\}}\mathbb{P}_\theta(y)\cdot u(d(y),\theta)$ under the minimum on both sides of the equation, we get

$$\min_{\theta\in\Theta}\left(\sum_{y\in\mathcal{X}_1\setminus\{x_1\}}\mathbb{P}_\theta(y)\cdot u(d(y),\theta)\right)+\mathbb{P}_\theta(x_1)\cdot u(d(x_1),\theta) <$$
$$\min_{\theta\in\Theta}\left(\sum_{y\in\mathcal{X}_1\setminus\{x_1\}}\mathbb{P}_\theta(y)\cdot u(d(y),\theta)\right)+\mathbb{P}_\theta(x_1)\cdot u(\{b_1\},\theta).$$

The left-hand side of that inequality is $\min_{\theta\in\Theta}\mathbb{E}^1_\theta[d]$. And if we let f be the decision rule that is exactly like d except $f(x_1)=\{b_1\}$, then the right-hand side is $\min_{\theta\in\Theta}\mathbb{E}^1_\theta[f]$. So we've shown that

$$\min_{\theta\in\Theta}\mathbb{E}^1_\theta[d] < \min_{\theta\ \in\Theta}\mathbb{E}^1_\theta[f], \tag{15}$$

which contradicts the assumption that d is maximin. That finishes the proof of the claim about maximin.

Next we prove the claim about expected-utility maximization. Suppose that (I) d adjusts for multiplicity and maximizes SEU with respect to the probability measure P, and (II) the hypotheses (i.e., members of Θ) are mutually independent with respect to P. To say that d maximizes SEU with respect to P means that

1. $\mathbb{E}_{P(\cdot|X_1=y)}[d(y)] \geq \mathbb{E}_{P(\cdot|X_1=y)}[a_1]$ for all $a_1\in A_1$ and all $y\in\mathcal{X}_1$, and

2. $\mathbb{E}_{P(\cdot|\vec{X}=\vec{y})}[d(\vec{y})] \geq \mathbb{E}_{P(\cdot|\vec{X}=\vec{y})}[S]$ for all for all strategies $S\subset\bigcup_{k\leq N}A_k$ and all $\vec{y}\in\mathcal{X}$.

As earlier, let $\vec{x}$ be the vector witnessing the fact that d adjusts for multiplicity, and define a decision rule e as in the first half of the proof.

Because $e(y)=d(y)$ for all $y\in\mathcal{X}_1$, it follows immediately that $e(y)$ maximizes SEU with respect to $P(\cdot|X_1=y)$ for all $y\in\mathcal{X}_1$ (because $d(y)$ is a maximizer!).

So it remains to be shown that $e(\vec{y})$ maximizes SEU with respect to $P(\cdot|\vec{X} = \vec{y})$ for all $\vec{y} \in \mathcal{X}$. Because $e(\vec{y}) = d(\vec{y})$ for all $\vec{y} \neq \vec{x}$ and because d is an SEU maximizer, it suffices to show that

$$\mathbb{E}_{P(\cdot|\vec{X}]=\vec{x})}[e(\vec{x})] \geq \mathbb{E}_{P(\cdot|\vec{X}=\vec{x})}[d(\vec{x})].$$

To show this, notice that we can decompose $\mathbb{E}_{P(\cdot|\vec{X}=\vec{x})}[e(\vec{x})]$ as follows:

$$\begin{aligned}
\mathbb{E}_{P(\cdot|\vec{X}=\vec{x})}[e(\vec{x})] &= \sum_{\theta \in \Theta} P(\theta|\vec{X} = \vec{x}) \cdot u(e(\vec{x}), \theta) \\
&= \sum_{\theta \in \Theta} \sum_{k \leq N} P(\theta|\vec{X} = \vec{x}) \cdot u_k(e_k(\vec{x}), \theta_k) \quad \text{ by separability} \\
&= \sum_{\theta \in \Theta} P(\theta|\vec{X} = \vec{x}) \cdot u_1(e_1(\vec{x}), \theta_1) + \sum_{\theta \in \Theta} \sum_{1 < k \leq N} P(\theta|\vec{X} = \vec{x}) \cdot u_k(e_k(\vec{x}), \theta_k).
\end{aligned}$$

Now notice that because $e_k(\vec{y}) = d_k(\vec{y})$ for all $k > 1$, the second summand—that is, the double sum—is equal to the same term in which d_k is substituted for e_k. So it suffices to show that

$$\sum_{\theta \in \Theta} P(\theta|\vec{X} = \vec{x}) \cdot u_1(e_1(\vec{x}), \theta_1) \geq \sum_{\theta \in \Theta} P(\theta|\vec{X} = \vec{x}) \cdot u_1(d_1(\vec{x}), \theta_1). \tag{16}$$

By Bayes's rule and our assumptions about mutual independence of the hypotheses (and of the random variables), we have that for all θ:

$$
\begin{aligned}
P(\theta|\vec{X}=\vec{x}) &= \frac{\mathbb{P}_\theta(\vec{X}=\vec{x})\cdot P(\theta)}{P(\vec{X}=\vec{x})} \\
&= \frac{\prod_{k\leq N}\mathbb{P}_{\theta_k}(X_k=x_k)\cdot P(\theta_k)}{P(\vec{X}=\vec{x})} \\
&= \frac{\prod_{k\leq N}P(X_k=x_k|\theta_k)\cdot P(\theta_k)}{P(\vec{X}=\vec{x})} \\
&= \frac{\prod_{k\leq N}P(\theta_k|X_k=x_k)\cdot P(X_k=x_k)}{P(\vec{X}=\vec{x})} \\
&= \frac{\prod_{k\leq N}P(\theta_k|X_k=x_k)\cdot \prod_{k\leq N}P(X_k=x_k)}{P(\vec{X}=\vec{x})} \\
&= \frac{\prod_{k\leq N}P(X_k=x_k)}{P(\vec{X}=\vec{x})}\cdot\prod_{k\leq N}P(\theta_k|X_k=x_k).
\end{aligned}
$$

It follows that equation (16) holds if and only if:

$$
\sum_{\theta\in\Theta}\prod_{k\leq N}P(\theta_k|X_k=x_k)\cdot u_1(e_1(\vec{x}),\theta_1)\geq\sum_{\theta\in\Theta}\prod_{k\leq N}P(\theta_k|X_k=x_k)\cdot u_1(d_1(\vec{x}),\theta_1). \tag{17}
$$

Recall that $e_1(\vec{x})=d(x_1)$ by construction, and so the last inequality holds if and only if

$$
\sum_{\theta\in\Theta}\prod_{k\leq N}P(\theta_k|X_k=x_k)\cdot u(d(x_1),\theta_1)\geq\sum_{\theta\in\Theta}\prod_{k\leq N}P(\theta_k|X_k=x_k)\cdot u_1(b_1,\theta_1). \tag{18}
$$

Now rewrite the term on the left-hand side of equation (18). To do so, perform the outside sum in two steps, by first summing over values of θ_1 and then by summing over the values of $\theta_2,\ldots,\theta_N$. In other words, observe that we can rewrite the left-hand side of the equation as follows:

$$\sum_{\theta\in\Theta}\prod_{k\leq N} P(\theta_k|X_k=x_k)\cdot u_1(d(x_1),\theta_1)$$

$$=\sum_{\theta_1}\sum_{\theta_2,\ldots\theta_N}\prod_{k\leq N} P(\theta_k|X_k=x_k)\cdot u_1(d(x_1),\theta_1)$$

$$=\sum_{\theta_1}\sum_{\theta_2,\ldots\theta_N}(P(\theta_1|X_1=x_1)\cdot u_1(d(x_1),\theta_1))\cdot\left(\prod_{1<k\leq N} P(\theta_k|X_k=x_k)\right)$$

$$=\sum_{\theta_2,\ldots\theta_N}\sum_{\theta_1}(P(\theta_1|X_1=x_1)\cdot u_1(d(x_1),\theta_1))\cdot\left(\prod_{1<k\leq N} P(\theta_k|X_k=x_k)\right),$$

by reordering the sums,

$$=\sum_{\theta_2,\ldots\theta_N}\left(\prod_{1<k\leq N} P(\theta_k|X_k=x_k)\cdot\left(\sum_{\theta_1} P(\theta_1|X_1=x_1)\cdot u_1(d(x_1),\theta_1)\right)\right)$$

$$=\sum_{\theta_2,\ldots\theta_N}\left(\prod_{1<k\leq N} P(\theta_k|X_k=x_k)\cdot\left(\sum_{\upsilon\in\Theta:\upsilon_1=\theta_1} P(\upsilon|X_1=x_1)\cdot u(d(x_1),\upsilon)\right)\right),$$

as $u(d(x_1),\theta_1)) = u_1(d(x_1),\upsilon)$ if $\upsilon_1=\theta_1$ by separability,

$$=\sum_{\theta_2,\ldots\theta_N}\prod_{1<k\leq N} P(\theta_k|X_k=x_k)\cdot\mathbb{E}_{P(\cdot|X_1=x_1)}[d(x_1)]$$

$$=\mathbb{E}_{P(\cdot|X_1=x_1)}[d(x_1)]\cdot\sum_{\theta_2,\ldots\theta_N}\prod_{1<k\leq N} P(\theta_k|X_k=x_k).$$

□