SUPPLEMENT TO "FROM BLACKWELL DOMINANCE IN LARGE SAMPLES TO RÉNYI DIVERGENCES AND BACK AGAIN"

XIAOSHENG MU
Department of Economics, Princeton University

LUCIANO POMATTO
Division of the Humanities and Social Sciences, Caltech

PHILIPP STRACK
Department of Economics, Yale University

OMER TAMUZ
Division of the Humanities and Social Sciences, Caltech

THIS SUPPLEMENTARY APPENDIX is organized as follows: Section G contains the proof of Proposition 4; Section H presents the proof of Theorem 2; Section I contains examples showing that Theorem 1 does not hold in the absence of genericity assumptions; Section J generalizes our main results to unbounded experiments; in Section K we show that the Rényi order remains a necessary condition for large sample dominance in the case of more than two states; finally, in Section L we show our our main results can be applied to prove a conjecture by Jensen (2019) on the majorization order.

APPENDIX G: PROOF OF PROPOSITION 4

That (i) implies (ii) follows from the fact that Rényi divergences are monotone in the Blackwell order, and additive with respect to independent experiments.

To show (ii) implies (i), we introduce some notation. Given two experiments $P = (\Omega, P_0, P_1)$ and $Q = (\Xi, Q_0, Q_1)$, for each $\alpha \in [0, 1]$ we denote by $\alpha P + (1 - \alpha)Q = (\Psi, M_0, M_1)$ the mixed experiment where the sample space is the disjoint union $\Psi = \Omega \sqcup \Xi$ endowed with the corresponding σ-algebra, and the measures M_0, M_1 satisfy for every measurable $E \subseteq \Psi$

$$M_\theta(E) = \alpha P_\theta(E \cap \Omega) + (1 - \alpha)Q_\theta(E \cap \Xi).$$

Intuitively, the mixed experiment corresponds to a randomized experiment where P is carried out with probability α and Q with probability $1 - \alpha$. The mixture operation and the product operation satisfy $(\alpha P + (1 - \alpha)Q) \otimes R = \alpha(P \otimes R) + (1 - \alpha)(Q \otimes R)$.

Now suppose P dominates Q in the Rényi order, then by Theorem 1, P dominates Q in the large sample order. The next lemma concludes the proof.

LEMMA 4: *Let P, Q be bounded experiments such that P dominates Q in the large sample order. Then there exists a bounded experiment R such that $P \otimes R$ Blackwell dominates $Q \otimes R$.*

Xiaosheng Mu: xmu@princeton.edu
Luciano Pomatto: luciano@caltech.edu
Philipp Strack: philipp.strack@yale.edu
Omer Tamuz: tamuz@caltech.edu

This lemma replicates a more general statement that appears in Duan, Feng, Li, and Ying (2005), Fritz (2017).

PROOF OF LEMMA 4: Assume $P^{\otimes n} \succeq Q^{\otimes n}$. Let

$$R = \frac{1}{n}\big(Q^{\otimes n} + P \otimes Q^{\otimes(n-1)} + P^{\otimes 2} \otimes Q^{\otimes(n-2)} + \cdots + P^{\otimes(n-2)} \otimes Q^{\otimes 2} + P^{\otimes(n-1)} \otimes Q\big).$$

Then

$$
\begin{aligned}
P \otimes R &= P \otimes \frac{1}{n}\big(Q^{\otimes n} + P \otimes Q^{\otimes(n-1)} + \cdots + P^{\otimes(n-2)} \otimes Q^{\otimes 2} + P^{\otimes(n-1)} \otimes Q\big) \\
&= \frac{1}{n}\big(P \otimes Q^{\otimes n} + P^{\otimes 2} \otimes Q^{\otimes(n-1)} + \cdots + P^{\otimes(n-1)} \otimes Q^{\otimes 2} + P^{\otimes n} \otimes Q\big) \\
&\succeq \frac{1}{n}\big(P \otimes Q^{\otimes n} + P^{\otimes 2} \otimes Q^{\otimes(n-1)} + \cdots + P^{\otimes(n-1)} \otimes Q^{\otimes 2} + Q^{\otimes(n+1)}\big) \\
&= Q \otimes \frac{1}{n}\big(Q^{\otimes n} + P \otimes Q^{\otimes(n-1)} + \cdots + P^{\otimes(n-1)} \otimes Q\big) \\
&= Q \otimes R,
\end{aligned}
$$

where the middle step uses the assumption $P^{\otimes n} \succeq Q^{\otimes n}$, so that $P^{\otimes n} \otimes Q \succeq Q^{\otimes(n+1)}$.

Q.E.D.

APPENDIX H: PROOF OF THEOREM 2

Throughout this section, we denote by D an additive divergence that satisfies the data-processing inequality and is finite on bounded experiments.

LEMMA 5: *If a bounded experiment $P = (\Omega, P_0, P_1)$ dominates another bounded experiment $Q = (\Xi, Q_0, Q_1)$ in the Blackwell order, then $D(P_0, P_1) \geq D(Q_0, Q_1)$.*

PROOF: By Blackwell's theorem there exists a measurable function $\sigma\colon \Omega \to \Delta(\Xi)$ such that $Q_\theta(A) = \int \sigma(\omega)(A)\, dP_\theta(\omega)$ for every measurable $A \subseteq \Xi$ and every θ. Let λ be the Lebesgue measure on $[0, 1]$. Since Ω and Ξ are Polish spaces, there exists a measurable function $f\colon \Omega \times [0, 1] \to \Xi$ such that for every $\omega \in \Omega$, $\sigma(\omega) = f(\omega, \cdot)_*(\lambda)$, where $f(\omega, \cdot)_*(\lambda)$ is the push-forward of λ induced by the function $f(\omega, \cdot)$ (see, e.g., Proposition 10.7.6 in Bogachev, 2007). Hence,

$$Q_\theta(A) = \int \lambda\big(\{t \in [0, 1] : f(\omega, t) \in A\}\big)\, dP_\theta(\omega) = f_*(P_\theta \times \lambda)(A),$$

where now $f_*(P_\theta \times \lambda)$ is the push-forward of $P_\theta \times \lambda$ induced by f. Being a divergence, D satisfies $D(\lambda, \lambda) = 0$. Moreover, by additivity, $D(P_0 \times \lambda, P_1 \times \lambda) = D(P_0, P_1)$. The data processing inequality then implies $D(P_0, P_1) = D(P_0 \times \lambda, P_1 \times \lambda) \geq D(Q_0, Q_1)$.   *Q.E.D.*

LEMMA 6: *If the bounded experiments $P = (P_0, P_1)$ and $Q = (Q_0, Q_1)$ satisfy $R_P^\theta(t) \geq R_Q^\theta(t)$ for every $t > 0$ and $\theta \in \{0, 1\}$, then $D(P_0, P_1) \geq D(Q_0, Q_1)$.*

PROOF: Suppose first that the strict inequality $R_P^\theta(t) > R_Q^\theta(t)$ holds for every $t > 0$, including at the limit $t = \infty$ (corresponding to the genericity assumption in the main text). Then, by Theorem 1 there exists n such that $P^{\otimes n}$ dominates $Q^{\otimes n}$ in the Blackwell order. Hence, by applying the previous lemma and by additivity, we obtain

$$nD(P_0, P_1) = D\big(P_0^n, P_1^n\big) \geq D\big(Q_0^n, Q_1^n\big) = nD(Q_0, Q_1).$$

More generally, suppose we only have the weak inequality $R_P^\theta(t) \geq R_Q^\theta(t)$ for $t > 0$. Fix a bounded and nontrivial experiment $S = (S_0, S_1)$. Then, for every $k \in \mathbb{N}$ we have

$$R_{P^{\otimes k} \otimes S}^\theta(t) = kR_P^\theta(t) + R_S^\theta(t) > kR_Q^\theta(t) = R_{Q^{\otimes k}}^\theta(t)$$

for every $t \in (0, \infty]$ and $\theta \in \{0, 1\}$. Given what we just proved, it follows that

$$D\big(P_0^k \times S_0, P_1^k \times S_1\big) \geq D\big(Q_0^k, Q_1^k\big).$$

By additivity, $D(P_0, P_1) + \frac{1}{k}D(S_0, S_1) \geq D(Q_0, Q_1)$. Since this holds for every k and $D(S_0, S_1)$ is finite, the proof is concluded.   Q.E.D.

Let $\overline{\mathbb{R}} = [-\infty, \infty]$ be the extended real line. Given a bounded experiment P we define the function $H_P \colon \overline{\mathbb{R}} \to \mathbb{R}$ as

$$H_P(t) = \begin{cases} R_P^1(t) & \text{if } t \geq 1/2, \\ R_P^0(1-t) & \text{if } t \leq 1/2. \end{cases}$$

Recall that the Rényi divergences of an experiment P satisfy the relation $(1-t)R_P^1(t) = tR_P^0(1-t)$. This implies that the function H_P is well-defined, continuous, and bounded. It is a convenient representation of the Rényi divergences that retains the main properties of the latter, and has the advantage of being strictly positive whenever P is nontrivial. Since $H_P(t)$ is continuous and has a compact domain, it is furthermore bounded away from 0. The functional $P \mapsto H_P$ satisfies two additional properties. An experiment P dominates an experiment Q in the Rényi order if and only if $H_P(t) > H_Q(t)$ for every t. Moreover, the functional is additive: $H_{P \otimes Q}(t) = H_P(t) + H_Q(t)$ for every t.

Thus, to prove Theorem 2 it suffices to show that under the hypotheses of the theorem there exists a finite measure m on $\overline{\mathbb{R}}$ such that for every bounded pair of measures P_0, P_1

$$D(P_0, P_1) = \int_{\overline{\mathbb{R}}} H_P(t)\, dm(t)$$

where P is the experiment (P_0, P_1). The theorem's conclusion (7) follows easily from this by setting $dm_0(t) = -dm(1-t)$ and $dm_1(t) = dm(t)$ for $t \geq \frac{1}{2}$.

Let $C(\overline{\mathbb{R}})$ be the space of continuous functions defined over the compact set $\overline{\mathbb{R}}$. Each function H_P belongs to $C(\overline{\mathbb{R}})$. Consider the set

$$\mathcal{H} = \{H_P : P \text{ is a bounded experiment}\} \subseteq C(\overline{\mathbb{R}}).$$

By Lemma 6, if $H_P = H_Q$ then $D(P_0, P_1) = D(Q_0, Q_1)$. Thus there exists a map $F \colon \mathcal{H} \to \mathbb{R}$ such that $D(P_0, P_1) = F(H_P)$.

By Lemma 6, the functional F is monotone. It is moreover additive: Given two experiments P and Q, the additivity of D and the additivity of $P \mapsto H_P$ imply

$$
\begin{aligned}
F(H_P) + F(H_Q) &= D(P_0, P_1) + D(Q_0, Q_1) \\
&= D(P_0 \times Q_0, P_1 \times Q_1) \\
&= F(H_{P \otimes Q}) \\
&= F(H_P + H_Q).
\end{aligned}
$$

Next, we define $\mathrm{cone}_{\mathbb{Q}}(\mathcal{H}) = \{\sum_{i=1}^{n} \alpha_i H_{P^i} : \alpha_i \in \mathbb{Q}_+, P^i \text{ is a bounded experiment}\}$ to be the rational cone generated by $\mathcal{H}$, where coefficients (α_i) are positive rational numbers. Similarly, define

$$
\mathrm{cone}(\mathcal{H}) = \left\{ \sum_{i=1}^{n} \alpha_i H_{P^i} : \alpha_i \in \mathbb{R}_+, P^i \text{ is a bounded experiment} \right\}
$$

to be the cone generated by $\mathcal{H}$, where coefficients can be all positive numbers. Below we extend the functional F from $\mathcal{H}$ to $\mathrm{cone}_{\mathbb{Q}}(\mathcal{H})$ and then to $\mathrm{cone}(\mathcal{H})$.

Because $P \mapsto H_P$ is additive, $\mathcal{H}$ is itself closed under addition. This implies

$$
\mathrm{cone}_{\mathbb{Q}}(\mathcal{H}) = \bigcup_{n \geq 1} \frac{1}{n} \mathcal{H}.
$$

Define $G \colon \mathrm{cone}_{\mathbb{Q}}(\mathcal{H}) \to \mathbb{R}$ as $G(\frac{1}{n} H_P) = \frac{1}{n} F(H_P)$. The functional G is well-defined: If $\frac{1}{n} H_P = \frac{1}{m} H_Q$, then $H_{P \otimes m} = m H_P = n H_Q = H_{Q \otimes n}$, which implies $m F(H_P) = n F(H_Q)$ by the additivity of F. Similarly, G inherits the monotonicity and additivity of F on the larger domain $\mathrm{cone}_{\mathbb{Q}}(\mathcal{H})$.

We now show G is a Lipschitz functional, where we endow the space $C(\overline{\mathbb{R}})$ with the sup norm. Let S_0 be a nontrivial experiment, so that $H_{S_0}(t)$ is positive and in fact bounded away from 0 for every t. By letting $S = S_0^{\otimes k}$ for large k, we obtain that $H_S(t) > 1$ for every t. Given two functions $f, \hat{f} \in \mathrm{cone}_{\mathbb{Q}}(\mathcal{H})$, we have the pointwise comparison

$$
f(t) \leq \hat{f}(t) + \|f - \hat{f}\| \times H_S(t).
$$

Let $r > \|f - \hat{f}\|$ be a rational number. The additivity and the monotonicity of G imply

$$
G(f) \leq G(\hat{f} + r H_S) = G(\hat{f}) + r G(H_S).
$$

Symmetrically, $G(\hat{f}) \leq G(f + r H_S) = G(f) + r G(H_S)$, so that $|G(f) - G(\hat{f})| \leq r G(H_S)$. By taking the limit $r \to \|f - \hat{f}\|$ we obtain that G is Lipschitz with Lipschitz constant $G(H_S) < \infty$, that is,

$$
\left| G(f) - G(\hat{f}) \right| \leq \|f - \hat{f}\| \cdot G(H_S).
$$

Thus G can be extended to a Lipschitz functional $\overline{G}$ defined on the closure of $\mathrm{cone}_{\mathbb{Q}}(\mathcal{H})$, which contains $\mathrm{cone}(\mathcal{H})$.

We now verify that $\overline{G}$ is still monotone on $\mathrm{cone}(\mathcal{H})$. Let $f \geq \hat{f}$ be two functions in $\mathrm{cone}(\mathcal{H})$, and take any two sequences $\{\frac{1}{p_n} H_{P_n}\}$ and $\{\frac{1}{q_n} H_{Q_n}\}$ in $\mathrm{cone}_{\mathbb{Q}}(\mathcal{H})$ that converge

to f and $\hat{f}$ as $n \to \infty$. For any positive integer m, convergence in the sup-norm implies $\frac{1}{p_n} H_{P_n} \geq f - \frac{1}{2m} H_S$ for all large n, where S is the experiment with $H_S > 1$ everywhere. Similarly, $\frac{1}{q_n} H_{Q_n} \leq \hat{f} + \frac{1}{2m} H_S$. Since $f \geq \hat{f}$, we thus have $\frac{1}{p_n} H_{P_n} \geq \frac{1}{q_n} H_{Q_n} - \frac{1}{m} H_S$ for all large n. By monotonicity and additivity of G, $G(\frac{1}{p_n} H_{P_n}) \geq G(\frac{1}{q_n} H_{Q_n}) - \frac{1}{m} G(H_S)$, which implies $\overline{G}(f) \geq \overline{G}(\hat{f}) - \frac{1}{m} G(H_S)$ by taking $n \to \infty$. As m is arbitrary, we have shown that $\overline{G}$ is monotonic.

We show $\overline{G}$ is additive and satisfies $\overline{G}(af + b\hat{f}) = a\overline{G}(f) + b\overline{G}(\hat{f})$ for any functions $f, \hat{f} \in \mathrm{cone}(\mathcal{H})$ and $a, b \in \mathbb{R}_+$. To show this, first suppose a, b are rational numbers. Consider $\{\frac{1}{p_n} H_{P_n}\} \to f$ and $\{\frac{1}{q_n} H_{Q_n}\} \to \hat{f}$ as above, where f need not be bigger than $\hat{f}$. Then the sequence of functions $\{\frac{a}{p_n} H_{P_n} + \frac{b}{q_n} H_{Q_n}\} \in \mathrm{cone}_{\mathbb{Q}}(\mathcal{H})$ converges to $af + b\hat{f}$. It follows that

$$
\overline{G}(af + b\hat{f}) = \lim_{n \to \infty} G\left(\frac{a}{p_n} H_{P_n} + \frac{b}{q_n} H_{Q_n} \right)
$$

$$
= a \cdot \lim_{n \to \infty} G\left(\frac{1}{p_n} H_{P_n} \right) + b \cdot \lim_{n \to \infty} G\left(\frac{1}{q_n} H_{Q_n} \right) = a \cdot \overline{G}(f) + b \cdot \overline{G}(\hat{f}).
$$

If a, b are real numbers, we can deduce the same result by the Lipschitz property of $\overline{G}$.

Consider next $V = \mathrm{cone}(\mathcal{H}) - \mathrm{cone}(\mathcal{H})$, which is vector subspace of $C(\overline{\mathbb{R}})$. $\overline{G}$ can be further extended to a functional $I : V \to \mathbb{R}$, defined as

$$
I(M_1 - M_2) = \overline{G}(M_1) - \overline{G}(M_2)
$$

for all $M_1, M_2 \in \mathrm{cone}(\mathcal{H})$. The functional I is well-defined and linear because $\overline{G}$ is affine. Moreover, by monotonicity of $\overline{G}$, $I(f) \geq 0$ for any nonnegative function $f \in V$.

The following theorem, a generalization of the Hahn–Banach theorem (see, e.g., Theorem 8.32 in Aliprantis and Border (2006)), shows that I can be further extended to a positive linear functional on the entire space $C(\overline{\mathbb{R}})$:

THEOREM 5—(Kantorovich (1937)): *Let V be a vector subspace of $C(\overline{\mathbb{R}})$ with the property that for every $f \in C(\overline{\mathbb{R}})$ there exists a function $g \in V$ such that $g \geq f$. Then every positive linear functional on V extends to a positive linear functional on $C(\overline{\mathbb{R}})$.*

The "majorization" condition $g \geq f$ is satisfied because every function in $C(\overline{\mathbb{R}})$ is bounded by some n, and V contains the function $n H_S$ which takes values greater than n everywhere.

To summarize, we have obtained a positive linear functional J defined on $C(\overline{\mathbb{R}})$ that extends the original functional $F(H_P) = D(P_0, P_1)$. By the Riesz representation theorem for positive linear functionals over spaces of continuous functions on compact sets, we conclude that $J(f) = \int_{\overline{\mathbb{R}}} f(t) \, dm(t)$ for some finite measure m. Hence $D(P_0, P_1) = F(H_P) = J(H_P)$ is an integral of the Rényi divergences of P, completing the proof of Theorem 2.

APPENDIX I: NECESSITY OF THE GENERICITY ASSUMPTION

Here, we present examples to show that Theorem 1 does not hold without the genericity assumption.

Consider the experiments P and Q described in Example 2 in Section 3.1. Fix $\alpha = \frac{1}{4}$ and $\beta = \frac{1}{16}$, which satisfy (25). Then by Proposition 2, P dominates Q in large samples.

We will perturb these two experiments by adding another signal realization (to each experiment) which strongly indicates the true state is 1. The perturbed conditional probabilities are given below:

$\tilde{P}$:

θ	x_0	x_1	x_2	x_3
0	ε	$\frac{1}{16}$	$\frac{1}{2}$	$\frac{7}{16} - \varepsilon$
1	100ε	$\frac{7}{16}$	$\frac{1}{2}$	$\frac{1}{16} - 100\varepsilon$

$\tilde{Q}$:

θ	y_0	y_1	y_2
0	ε	$\frac{1}{4}$	$\frac{3}{4} - \varepsilon$
1	100ε	$\frac{3}{4}$	$\frac{1}{4} - 100\varepsilon$

If ε is a small positive number, then by continuity $\tilde{P}$ still dominates $\tilde{Q}$ in the Rényi order. Nonetheless, we show below that $\tilde{P}^{\otimes n}$ does not Blackwell dominate $\tilde{Q}^{\otimes n}$ for any n and $\varepsilon > 0$.

To do this, let $\overline{p} := \frac{100^{n-1}}{100^{n-1}+1}$ be a threshold belief. We will show that a decision maker whose indirect utility function is $(p - \overline{p})^+$ strictly prefers $\tilde{Q}^{\otimes n}$ to $\tilde{P}^{\otimes n}$. Indeed, it suffices to focus on posterior beliefs $p > \overline{p}$; that is, the likelihood ratio should exceed 100^{n-1}. Under $\tilde{Q}^{\otimes n}$, this can only happen if every signal realization is y_0, or all but one signal is y_0 and the remaining one is y_1. Thus, in the range $p > \overline{p}$, the posterior belief has the following distribution under $\tilde{Q}^{\otimes n}$:

$$p = \begin{cases} \dfrac{100^n}{100^n + 1} & \text{w.p. } \dfrac{1}{2}(100^n + 1)\varepsilon^n, \\ \dfrac{3 \cdot 100^{n-1}}{3 \cdot 100^{n-1} + 1} & \text{w.p. } \dfrac{n}{8}(3 \cdot 100^{n-1} + 1)\varepsilon^{n-1}. \end{cases}$$

Similarly, under $\tilde{P}^{\otimes n}$ the relevant posterior distribution is

$$p = \begin{cases} \dfrac{100^n}{100^n + 1} & \text{w.p. } \dfrac{1}{2}(100^n + 1)\varepsilon^n, \\ \dfrac{7 \cdot 100^{n-1}}{7 \cdot 100^{n-1} + 1} & \text{w.p. } \dfrac{n}{32}(7 \cdot 100^{n-1} + 1)\varepsilon^{n-1}. \end{cases}$$

Recall that the indirect utility function is $(p - \overline{p})^+$. So $\tilde{Q}^{\otimes n}$ yields higher expected payoff than $\tilde{P}^{\otimes n}$ if and only if

$$\frac{n}{8}(3 \cdot 100^{n-1} + 1)\varepsilon^{n-1} \cdot \left(\frac{3 \cdot 100^{n-1}}{3 \cdot 100^{n-1} + 1} - \overline{p} \right)$$
$$> \frac{n}{32}(7 \cdot 100^{n-1} + 1)\varepsilon^{n-1} \cdot \left(\frac{7 \cdot 100^{n-1}}{7 \cdot 100^{n-1} + 1} - \overline{p} \right).$$

That is,

$$4(3 \cdot 100^{n-1} + 1) \cdot \left(\frac{3 \cdot 100^{n-1}}{3 \cdot 100^{n-1} + 1} - \frac{100^{n-1}}{100^{n-1} + 1} \right)$$
$$> (7 \cdot 100^{n-1} + 1) \cdot \left(\frac{7 \cdot 100^{n-1}}{7 \cdot 100^{n-1} + 1} - \frac{100^{n-1}}{100^{n-1} + 1} \right).$$

The LHS is computed to be $\frac{8 \cdot 100^{n-1}}{100^{n-1}+1}$, while the RHS is $\frac{6 \cdot 100^{n-1}}{100^{n-1}+1}$. Hence the above inequality holds, and it follows that $\tilde{P}^{\otimes n}$ does not Blackwell dominate $\tilde{Q}^{\otimes n}$.

APPENDIX J: GENERALIZATION TO UNBOUNDED EXPERIMENTS

In this section, we present two generalizations of Theorem 1 to experiments that may have unbounded likelihood ratios. Note that the Rényi divergences for an unbounded experiment can still be defined by (3), (4), and (5), so long as we allow these divergences to take the value $+\infty$.

The first result shows that Theorem 1 hold without change so long as the dominated experiment Q is bounded.

THEOREM 6: *For a generic pair of experiments P and Q where Q is bounded, the following are equivalent*:
  (i) *P dominates Q in large samples*.
  (ii) *P dominates Q in the Rényi order*.

To interpret the statement, "generic" means (as in the main text) that $\log \frac{dP_1}{dP_0}$ has different essential maximum and minimum from $\log \frac{dQ_1}{dQ_0}$. In the current setting, P may be unbounded, so that its log-likelihood ratio may have essential maximum $+\infty$ and/or minimum $-\infty$. In those cases, the genericity assumption is automatically satisfied.

We also reiterate that dominance in the Rényi order means the Rényi divergences of P and Q are ranked as $R_P^\theta(t) > R_Q^\theta(t)$ for all $t > 0$ and $\theta \in \{0, 1\}$. Since Q is by assumption bounded, $R_Q^\theta(t)$ is always finite. Thus the requirement in (ii) is that $R_P^\theta(t)$ is either a bigger finite number, or it is $+\infty$.

Our second result in this section deals with pairs of experiments where both P and Q may be unbounded, but they still have finite Rényi divergences. To state the result, we need to generalize the notion of genericity as follows: Say P and Q form a *generic* pair, if for both $\theta = 0$ and $\theta = 1$,

$$\liminf_{t \to \infty} \left| R_P^\theta(t) - R_Q^\theta(t) \right| > 0. \tag{29}$$

Note that when P and Q are bounded, $R_P^\theta(t) \to \max[X^\theta]$ and $R_Q^\theta(t) \to \max[Y^\theta]$ as $t \to \infty$. So in this special case the genericity assumption reduces to the one we introduced in the main text.

The following result shows that under one extra assumption, Theorem 1 once again extends.

THEOREM 7: *Suppose P and Q are a generic pair of* (*possibly unbounded*) *experiments with finite Rényi divergences. Let* (X^θ), (Y^θ) *be the corresponding log-likelihood ratios, and suppose further that their cumulant generating functions satisfy* $\sup_{t \in \mathbb{R}} K_{X^\theta}''(t) < \infty$ *and* $\sup_{t \in \mathbb{R}} K_{Y^\theta}''(t) < \infty$.[1] *Then the following are equivalent*:
  (i) *P dominates Q in large samples*.
  (ii) *P dominates Q in the Rényi order*.

---

[1]Since $K_{X^0}(t) = K_{X^1}(-1-t)$, it suffices to check the assumptions $\sup_{t \in \mathbb{R}} K_{X^\theta}''(t) < \infty$ and $\sup_{t \in \mathbb{R}} K_{Y^\theta}''(t) < \infty$ for one of the two states.

We note that if a random variable X is bounded between $-b$ and b, then its Rényi divergences are finite, and $K_X''(t) \leq b^2$ for every t.[2] Thus Theorem 7 is another strict generalization of Theorem 1 beyond bounded experiments.

More generally, the following is a sufficient condition for Theorem 7 to apply. Roughly speaking, we require the log-likelihood ratios X^θ, Y^θ to have *tails decaying faster than some Gaussian distribution*.

LEMMA 7: *Let X be a random variable whose distribution admits a density $h(x)$ that is positive and twice continuously differentiable. Suppose there exists $\epsilon > 0$ and $M > 0$ such that the following holds*:

$$\frac{\partial^2 \log h(x)}{\partial x^2} \leq -\epsilon \quad \text{for all } |x| > M.$$

Then the cumulant generating function $K_X(t)$ is finite for every t, and $\sup_{t \in \mathbb{R}} K_X''(t) < \infty$.

Note that $\frac{\partial^2 \log h(x)}{\partial x^2} \leq -\epsilon$ implies the standard assumption that the density h is (strictly) log-concave. The requirement that the same ϵ works for all large x makes our assumption stronger, and in particular rules out densities such as $h_1(x) = c_1 \cdot e^{-\lambda_1 |x|}$ or $h_2(x) = c_2 \cdot e^{-\lambda_2 |x|^{1.99}}$.[3] Nonetheless, any Gaussian density h satisfies the assumption regardless of how big the variance is, and so does any other density that decays faster at infinity. Hence Theorem 7 is applicable to a broad class of unbounded experiments.

Below we prove Theorem 6, Theorem 7, and Lemma 7 in turn.

J.1. *Proof of Theorem 6*

That (i) implies (ii) follows from the same argument as in Section 5.1. To prove (ii) implies (i), the idea is to garble P into a bounded experiment $\tilde{P}$ that still has higher Rényi divergences than Q. By Theorem 1, $\tilde{P}^{\otimes n}$ Blackwell dominates $Q^{\otimes n}$ for all large n. But since P Blackwell dominates $\tilde{P}$, $P^{\otimes n}$ also Blackwell dominates $\tilde{P}^{\otimes n}$. Therefore, by transitivity, we would be able to conclude that $P^{\otimes n}$ Blackwell dominates $Q^{\otimes n}$ for all large n.

To construct such a $\tilde{P}$, we first note that by taking $t \to \infty$, $R_P^1(t) > R_Q^1(t)$ implies $\max[X^1] \geq \max[Y^1]$ where X^1 and Y^1 are the log-likelihood ratios. Similarly, $\max[X^0] \geq \max[Y^0]$. By the genericity assumption, both comparisons are in fact strict. We can thus find a pair of positive numbers $b_1 \in (\max[Y^1], \max[X^1])$ and $b_0 \in (\max[Y^0], \max[X^0]) = (-\min[Y^1], -\min[X^1])$. These numbers will be fixed throughout.

Now take any positive number $B \geq \max\{b_1, b_0\}$. We construct a garbling of P, denoted P_B, as follows: All signal realizations under P that induce a log-likelihood ratio $\log \frac{dP_1}{dP_0}$ greater than B (if any) are garbled into a single signal $\bar{s}$, and similarly all realizations with log-likelihood ratio less than $-B$ are garbled into another signal $\underline{s}$. The remaining signal realizations under P (with log-likelihood ratio in $[-B, B]$) are unchanged under P_B. It is easy to see that not only is P_B a garbling of P, but more generally P_B is a garbling of $P_{B'}$

---

[2] The latter follows by showing $K_X''(t)$ to be the variance of some random variable $\hat{X}$ that shares the same support as X. See Proposition 6 and its proof.

[3] It is easy to see that the random variable with density $h_1(x)$ does not have finite Rényi divergences everywhere. It can also be shown that the random variable with density $h_2(x)$ has a cumulant generating function with $K_X''(t) \to \infty$ as $t \to \infty$. Thus, it seems difficult to substantially weaken the condition in Lemma 7 while maintaining the same result.

whenever $B' > B$. Thus, as B increases, the experiment P_B becomes more informative in the Blackwell sense.

Let $R^\theta_{P_B}(t)$ denote the Rényi divergences of P_B. Since the Rényi order extends the Blackwell order, we know that as B increases, $R^\theta_{P_B}(t)$ also increases for each θ and t, with an upper bound of $R^\theta_P(t)$. In fact, we can show that for fixed θ and t,

$$\lim_{B \to \infty} R^\theta_{P_B}(t) = R^\theta_P(t).$$

The proof is technical and deferred to later. Assuming this, we next show that for sufficiently large B, $R^\theta_{P_B}(t) > R^\theta_Q(t)$ holds for *all* $t \geq 1/2$ (thus for all $t > 0$, by (6)). This will prove P_B as the desired garbling $\tilde{P}$ that dominates Q in the Rényi order, which will complete the proof of the theorem.[4]

To this end, fix $\theta = 1$, and define for each B a set

$$T_B = \left\{ t \geq 1/2 : R^1_{P_B}(t) \leq R^1_Q(t) \right\}.$$

By continuity of the Rényi divergences, T_B is a closed set. Moreover, as $t \to \infty$ we have $R^1_{P_B}(t) \to \max[X^1_B]$, where X^1_B is the log-likelihood ratio of state 1 to state 0, distributed under the experiment P_B and true state 1. By the assumption $B \geq b_1$ and the construction of P_B, we have that

$$\mathbb{P}[X^1_B \geq b_1] = \mathbb{P}[X^1 \geq b_1],$$

which is positive because $b_1 < \max[X^1]$. Thus $\max[X^1_B] \geq b_1$. It follows that

$$\lim_{t \to \infty} R^1_{P_B}(t) \geq b_1 > \max[Y^1] = \lim_{t \to \infty} R^1_Q(t).$$

Hence $R^1_{P_B}(t) > R^1_Q(t)$ for all large t and T_B is a bounded set.

We have shown that each T_B is compact set. Note also that because $R^1_{P_B}(t)$ increases in B, the set T_B shrinks as B increases. Therefore, by the finite intersection property, either there exists some t that belongs to every T_B, or T_B is the empty set for all large B. The former is impossible because $R^1_{P_B}(t) \leq R^1_Q(t)$ for all B would imply $R^1_P(t) \leq R^1_Q(t)$ in the limit, contradicting the assumption in (ii).

We thus conclude that T_B must be empty for all large B. In other words, when B is large $R^1_{P_B}(t) > R^1_Q(t)$ holds for all $t \geq \frac{1}{2}$. A symmetric argument shows that $R^0_{P_B}(t) > R^0_Q(t)$ holds for all $t \geq \frac{1}{2}$, completing the proof.

It remains to show $\lim_{B \to \infty} R^\theta_{P_B}(t) = R^\theta_P(t)$. We again fix $\theta = 1$ for easier exposition. Consider the following three cases:

Case 1: $t > 1$.   We recall that $R^1_{P_B}(t) = \frac{1}{t-1} \log \mathbb{E}[e^{(t-1)X^1_B}]$. So we need to show

$$\lim_{B \to \infty} \mathbb{E}[e^{(t-1)X^1_B}] = \mathbb{E}[e^{(t-1)X^1}].$$

Since $R^1_{P_B}(t) \leq R^1_P(t)$ for each B, the LHS above is weakly smaller than the RHS. On the other hand, by construction X^1_B coincides with X^1 conditional on being in the interval

---

$[-B, B]$. As the exponential function is always positive, we have

$$\mathbb{E}\big[e^{(t-1)X_B^1}\big] \geq \mathbb{P}\big[|X_B^1| \leq B\big] \cdot \mathbb{E}\big[e^{(t-1)X_B^1} \mid |X_B^1| \leq B\big]$$
$$= \mathbb{P}\big[|X^1| \leq B\big] \cdot \mathbb{E}\big[e^{(t-1)X^1} \mid |X^1| \leq B\big].$$

Taking the limit as $B \to \infty$, we obtain $\lim_{B \to \infty} \mathbb{E}[e^{(t-1)X_B^1}] \geq \mathbb{E}[e^{(t-1)X^1}]$, which proves they are equal.

Case 2: $t = 1$. Here, we have $R_{P_B}^1(1) = \mathbb{E}[X_B^1]$. So we need to show

$$\lim_{B \to \infty} \mathbb{E}\big[X_B^1\big] = \mathbb{E}\big[X^1\big].$$

Once again we already know the LHS is weakly smaller, so it suffices to show the opposite inequality. By construction, X_B^1 coincides with X^1 on the interval $[-B, B]$. Other than this part, there is probability $\mathbb{P}[X^1 > B]$ that signal $\bar{s}$ occurs under the experiment P_B; when this happens we also have $X_B^1 > B$, which contributes a positive amount to $\mathbb{E}[X_B^1]$.

With remaining probability $\mathbb{P}[X^1 < -B]$, the signal $\underline{s}$ occurs, and the induced log-likelihood ratio X_B^1 is at least $\log \mathbb{P}[X^1 < -B]$ (since this event occurs with probability at most one under state 0). Here the contribution to $\mathbb{E}[X_B^1]$ can be negative, but is no less than $\mathbb{P}[X^1 < -B] \cdot \log \mathbb{P}[X^1 < -B]$.

Summarizing, for each B we have

$$\mathbb{E}\big[X_B^1\big] \geq \mathbb{P}\big[|X^1| \leq B\big] \cdot \mathbb{E}\big[X^1 \mid |X^1| \leq B\big] + \mathbb{P}\big[X^1 < -B\big] \cdot \log \mathbb{P}\big[X^1 < -B\big].$$

Taking the limit as $B \to \infty$, the first summand on the RHS converges to $\mathbb{E}[X^1]$. In addition, the second summand vanishes because $\mathbb{P}[X^1 < -B] \to 0$ and $\lim_{x \to 0} x \log x = 0$. We thus obtain $\lim_{B \to \infty} \mathbb{E}[X_B^1] \geq \mathbb{E}[X^1]$ as desired.

Case 3: $t \in (0, 1)$. In this case, we will again show

$$\lim_{B \to \infty} \mathbb{E}\big[e^{(t-1)X_B^1}\big] = \mathbb{E}\big[e^{(t-1)X^1}\big].$$

Since $R_{P_B}^1(t) \leq R_P^1(t)$, and $R_{P_B}^1(t) = \frac{1}{t-1} \log \mathbb{E}[e^{(t-1)X_B^1}]$, the negative factor $\frac{1}{t-1}$ implies that the LHS above is now weakly *bigger* than the RHS.

To prove it is smaller, we proceed as in Case 2. With probability $\mathbb{P}[X^1 > B]$ the signal $\bar{s}$ occurs, and the induced log-likelihood ratio X_B^1 is *at least* $\log \mathbb{P}[X^1 > B]$. As $t - 1$ is negative here, the contribution of this part to $\mathbb{E}[e^{(t-1)X_B^1}]$ is *at most*

$$\mathbb{P}\big[X^1 > B\big] \cdot \mathbb{E}\big[e^{(t-1)\log \mathbb{P}[X^1 > B]}\big] = \big(\mathbb{P}\big[X^1 > B\big]\big)^t.$$

Similarly, the contribution of the signal $\underline{s}$ is at most $(\mathbb{P}[X^1 < -B])^t$. We thus have

$$\mathbb{E}\big[e^{(t-1)X_B^1}\big] \leq \mathbb{P}\big[|X^1| \leq B\big] \cdot \mathbb{E}\big[e^{(t-1)X^1} \mid |X^1| \leq B\big] + \big(\mathbb{P}\big[X^1 > B\big]\big)^t + \big(\mathbb{P}\big[X^1 < -B\big]\big)^t.$$

As $B \to \infty$, both $(\mathbb{P}[X^1 > B])^t$ and $(\mathbb{P}[X^1 < -B])^t$ vanish since $t > 0$. We therefore conclude $\lim_{B \to \infty} \mathbb{E}[e^{(t-1)X_B^1}] \leq \mathbb{E}[e^{(t-1)X^1}]$, completing the whole proof.

J.2. *Proof of Theorem 7*

We only need to prove (ii) implies (i). Here, we will follow the arguments in Section 5.6 and make necessary modifications. Since Lemma 1 remains valid, it suffices to prove (22), that is,

$$\mathbb{P}\big[X_1^1 + \cdots + X_n^1 \leq na\big] \leq \mathbb{P}\big[Y_1^1 + \cdots + Y_n^1 \leq na\big], \quad \text{for all } a \geq 0.$$

The analysis of the four cases in Section 5.6 relies on Lemma 2 and Proposition 5. We will show later that Lemma 2 continues to hold even if P and Q are unbounded (but have finite Rényi divergences). On the other hand, Proposition 5 cannot hold as stated, but we do have the following modified version where b^2 is replaced by $\sup_{t \in \mathbb{R}} K_X''(t)$:

PROPOSITION 6: *Let X and Y be random variables with finite cumulant generating functions $K_X(t)$ and $K_Y(t)$. Further let $X_1, \ldots, X_n, Y_1, \ldots, Y_n$ be i.i.d. copies of X and Y respectively. Suppose $a \geq \mathbb{E}[Y]$, and $\eta > 0$ satisfies $K_Y^*(a) - \eta > K_X^*(a + \eta)$. Then for all $n \geq 4(1 + \eta)\eta^{-3} \cdot \sup_{t \in \mathbb{R}} K_X''(t)$, it holds that*

$$\mathbb{P}[X_1 + \cdots + X_n > na] \geq \mathbb{P}[Y_1 + \cdots + Y_n > na].$$

Using Lemma 2 and Proposition 6, we can replicate the results in Cases 1, 2, and 4 in Section 5.6. Specifically, let $M = \max\{\sup_{t \in \mathbb{R}} K_{X^1}''(t), \sup_{t \in \mathbb{R}} K_{Y^1}''(t)\}$, then for all $n \geq 4M(1 + \eta)\eta^{-3}$ the inequality $\mathbb{P}[X_1^1 + \cdots + X_n^1 \leq na] \leq \mathbb{P}[Y_1^1 + \cdots + Y_n^1 \leq na]$ holds for values of a outside of the interval $(\mathbb{E}[Y] + \eta, \mathbb{E}[X] - \eta)$ in Case 3.

Turning to $a \in (\mathbb{E}[Y] + \eta, \mathbb{E}[X] - \eta)$, we can still use the Chebyshev inequality to deduce

$$\mathbb{P}\big[X_1^1 + \cdots + X_n^1 \leq na\big] \leq \frac{\text{Var}\big[X^1\big]}{n\eta^2} = \frac{K_{X^1}''(0)}{n\eta^2} \leq \frac{M}{n\eta^2}.$$

Similarly, we also have

$$\mathbb{P}\big[Y_1^1 + \cdots + Y_n^1 \leq na\big] \geq 1 - \frac{\text{Var}\big[Y^1\big]}{n\eta^2} \geq 1 - \frac{M}{n\eta^2}.$$

Thus $\mathbb{P}[X_1^1 + \cdots + X_n^1 \leq na] \leq \mathbb{P}[Y_1^1 + \cdots + Y_n^1 \leq na]$ holds for all $n \geq 2M\eta^{-2}$, and hence for all $n \geq 4M(1 + \eta)\eta^{-3}$. This then implies that $P^{\otimes n}$ Blackwell dominates $Q^{\otimes n}$ for all $n \geq 4M(1 + \eta)\eta^{-3}$.

Below we supply the proofs for Lemma 2 (for unbounded experiments) and Proposition 6.

PROOF OF LEMMA 2 FOR UNBOUNDED EXPERIMENTS: We note that the second part $K_{Y^\theta}^*(a - \eta) < K_{X^\theta}^*(a) - \eta$ continues to hold. This is because, by the same argument as in the case of bounded experiments, $K_{Y^\theta}^*(a) < K_{X^\theta}^*(a)$ holds for all a in the *compact* interval $[0, \mathbb{E}[Y^\theta]]$. Thus by (uniform) continuity, we can "squeeze in" a small positive η without changing the inequality.

The first part of Lemma 2 also holds so long as $\max[Y^\theta]$ is finite, in which case the range of a under consideration is again compact. If instead $\max[Y^\theta] = \infty$, we use a new argument that takes advantage of the genericity assumption. Note that by assumption, $R_P^\theta(t) - R_Q^\theta(t)$ is positive for each θ and t. Given this, the genericity assumption (29)

further implies this difference is bounded away from zero as $t \to \infty$. That is, there exists small $\epsilon > 0$ and large $T > 1$ such that

$$R_P^\theta(t) - R_Q^\theta(t) > \epsilon \quad \text{for all } \theta \in \{0, 1\}, \; t > T.$$

Since $K_X^\theta(t) = t R_P^\theta(t+1)$, we deduce

$$K_X^\theta(t) - K_Y^\theta(t) > \epsilon t > \frac{\epsilon}{2}(t+1) \quad \text{for all } \theta \in \{0, 1\}, \; t > T. \tag{30}$$

We can now prove the first part of Lemma 2. Define $\delta > 0$ by $K'_{X^\theta}(T) = \mathbb{E}[X^\theta] + \delta$. The original proof of Lemma 2 yields that for all sufficiently small $\eta > 0$,

$$K_{Y^\theta}^*(a) - \eta > K_{X^\theta}^*(a + \eta) \quad \text{holds for } \mathbb{E}[X^\theta] - \eta \le a \le \mathbb{E}[X^\theta] + \delta.$$

Note that $\mathbb{E}[X^\theta] + \delta$ is finite, so the range of a considered above is compact, enabling us to use the original argument. We claim that by choosing $\eta < \epsilon/2$, where ϵ is defined earlier, the same inequality holds even if a is bigger than $\mathbb{E}[X^\theta] + \delta$. For this define $\hat{t}$ by $K'_{X^\theta}(\hat{t}) = a + \eta$, then $\hat{t} > T$ by the convexity of K_X. Therefore, by (30),

$$\begin{aligned}
K_{X^\theta}^*(a + \eta) &= \hat{t}(a + \eta) - K_{X^\theta}(\hat{t}) \\
&< \hat{t}(a + \eta) - K_{Y^\theta}(\hat{t}) - \frac{\epsilon}{2}(\hat{t} + 1) \\
&< \hat{t}(a + \eta) - K_{Y^\theta}(\hat{t}) - \eta(\hat{t} + 1) \\
&= \hat{t}a - K_{Y^\theta}(\hat{t}) - \eta \\
&\le K_{Y^\theta}^*(a) - \eta.
\end{aligned}$$

This completes the proof of Lemma 2 for unbounded experiments.   *Q.E.D.*

PROOF OF PROPOSITION 6: Following the original proof of Proposition 5, we just need to show a modified version of Lemma 3 (with $\sup_{t \in \mathbb{R}} K_X''(t)$ replacing b^2):

$$\mathbb{P}[X_1 + \cdots + X_n > na] \ge e^{-n \cdot K_X^*(a + \eta)} \left(1 - \frac{4 \cdot \sup\limits_{t \in \mathbb{R}} K_X''(t)}{n \eta^2} \right).$$

This follows the same proof as in Appendix A, except that in applying the Chebyshev inequality, we now use

$$\mathrm{Var}[\hat{S}_n] = n \, \mathrm{Var}[\hat{X}] = n \cdot K_X''(t) \le n \cdot \sup\limits_{\hat{t} \in \mathbb{R}} K_X''(\hat{t})$$

instead of $\mathrm{Var}[\hat{S}_n] \le nb^2$. The key equality $\mathrm{Var}[\hat{X}] = K_X''(t)$ holds because

$$\mathrm{Var}[\hat{X}] = \mathbb{E}[\hat{X}^2] - \mathbb{E}[\hat{X}]^2 = \frac{\mathbb{E}[X^2 e^{tX}]}{\mathbb{E}[e^{tX}]} - \left(\frac{\mathbb{E}[X e^{tX}]}{\mathbb{E}[e^{tX}]} \right)^2 = K_X''(t).$$

Hence the result.   *Q.E.D.*

J.3. *Proof of Lemma 7*

We first prove K_X is everywhere finite, that is, $\log \mathbb{E}[e^{tX}]$ is finite for every t. Using the density $h(x)$, we can write

$$\mathbb{E}[e^{tX}] = \int_{-\infty}^{\infty} h(x)e^{tx}\,\mathrm{d}x = \int_{-\infty}^{\infty} e^{tx+l(x)}\,\mathrm{d}x,$$

where we define $\ell(x) = \log h(x)$. Since by assumption $\ell''(x) \leq -\epsilon$ for $|x| > M$, it is easy to show $\ell(x) \leq -\frac{\epsilon}{4}x^2$ as $|x| \to \infty$. Hence the above integral is finite.

To prove K_X'' is bounded, we begin with the formula

$$K_X''(t) = \frac{\mathbb{E}[X^2 e^{tX}] \cdot \mathbb{E}[e^{tX}] - \mathbb{E}[X e^{tX}]^2}{\mathbb{E}[e^{tX}]^2}.$$

Let X_1, X_2 be i.i.d. copies of X. Then the denominator above is $\mathbb{E}[e^{tX_1}] \cdot \mathbb{E}[e^{tX_2}] = \mathbb{E}[e^{t(X_1+X_2)}]$. The numerator can be rewritten as

$$\mathbb{E}[X_1^2 e^{tX_1}] \cdot \mathbb{E}[e^{tX_2}] - \mathbb{E}[X_1 e^{tX_1}] \cdot \mathbb{E}[X_2 e^{tX_2}]$$

$$= \mathbb{E}[(X_1^2 - X_1 X_2) \cdot e^{t(X_1+X_2)}]$$

$$= \mathbb{E}\left[\frac{X_1^2 - X_1 X_2 + X_2^2 - X_1 X_2}{2} \cdot e^{t(X_1+X_2)}\right]$$

$$= \mathbb{E}\left[\frac{(X_1 - X_2)^2}{2} \cdot e^{t(X_1+X_2)}\right],$$

where the penultimate step uses the symmetry between X_1 and X_2. Define

$$D(s) = \mathbb{E}[(X_1 - X_2)^2 \mid X_1 + X_2 = s].$$

Then we have shown that

$$K_X''(t) = \frac{\frac{1}{2}\mathbb{E}[D(X_1 + X_2) \cdot e^{t(X_1+X_2)}]}{\mathbb{E}[e^{t(X_1+X_2)}]}.$$

Thus, in order to show K_X'' is bounded, it suffices to show $D(s)$ is bounded as s varies.

Recall that by assumption $\ell''(x) \leq -\epsilon$ for $|x| > M$. We will show (with proof deferred to later) there exists $S > 2M$, such that

$$\ell'(x) - \ell'(s - x) \leq -\frac{\epsilon}{2}(2x - s) \quad \text{for all } s > S,\, x > \frac{s}{2}. \tag{31}$$

Note that (31) in particular implies $\ell'(x) - \ell'(s - x) \leq -1$ for $x > \frac{s}{2} + C$, with $C = \epsilon^{-1}$. Given this, we can show $D(s)$ is bounded.

Without loss consider $s \geq 0$. We use the density $h(x)$ to write

$$D(s) = \frac{\int_{-\infty}^{\infty} h(x)h(s-x)(2x-s)^2\,\mathrm{d}x}{\int_{-\infty}^{\infty} h(x)h(s-x)\,\mathrm{d}x} = \frac{\int_{s/2}^{\infty} h(x)h(s-x)(2x-s)^2\,\mathrm{d}x}{\int_{s/2}^{\infty} h(x)h(s-x)\,\mathrm{d}x}. \tag{32}$$

Since $D(s)$ is continuous, it suffices to prove it is bounded when $s > S$, where S is given earlier. We now break the integral in (32) into two parts, with cutoff $s/2 + 2C$:

$$D(s) = \frac{\int_{s/2}^{s/2+2C} h(x)h(s-x)(2x-s)^2\,dx}{\int_{s/2}^{\infty} h(x)h(s-x)\,dx} + \frac{\int_{s/2+2C}^{\infty} h(x)h(s-x)(2x-s)^2\,dx}{\int_{s/2}^{\infty} h(x)h(s-x)\,dx}.$$

The first term is bounded by $16C^2$, which is the maximum value of $(2x - s)^2$ for $x \in [s/2, s/2 + 2C]$. To bound the second term, we rewrite it as

$$\int_{s/2+2C}^{\infty} \frac{e^{l(x)+l(s-x)}}{\int_{s/2}^{\infty} e^{l(y)+l(s-y)}\,dy} \cdot (2x-s)^2\,dx. \tag{33}$$

As $l'(y) - l'(s - y) \leq -1$ for $y \geq s/2 + C$, we have $l(y) + l(s - y) \geq x - y + l(x) + l(s - x)$ for all $x \geq y \geq s/2 + C$. Thus

$$\int_{s/2}^{\infty} e^{l(y)+l(s-y)}\,dy \geq \int_{s/2+C}^{x} e^{l(y)+l(s-y)}\,dy \geq \int_{s/2+C}^{x} e^{x-y+l(x)+l(s-x)}\,dy = \left(e^{x-s/2-C} - 1\right)e^{l(x)+l(s-x)}.$$

Plugging back into (33), the second term contributing to $D(s)$ is bounded above by

$$\int_{s/2+2C}^{\infty} \frac{1}{e^{x-s/2-C} - 1} \cdot (2x-s)^2\,dx = \int_{C}^{\infty} \frac{1}{e^u - 1} \cdot (2u+2C)^2\,du,$$

where we used change of variable from x to $u = x - s/2 - C$. Since the RHS is a finite constant independent of s, we conclude that $D(s)$ is bounded even as $s \to \infty$.

It remains to prove (31). We write the difference on the LHS as $\int_{s-x}^{x} \ell''(u)\,du$. If $s - x > M$, the result follows from the fact that $\ell''(u) \leq -\epsilon \leq -\frac{\epsilon}{2}$ for every u in the range of integration. Suppose instead that $s - x \leq M$, thus $x \geq s - M$. In this case because $\ell''(u)$ can only be positive for $u \in [-M, M]$, we have

$$\int_{s-x}^{x} \ell''(u)\,du \leq -\epsilon(2x - s - 2M) + \int_{-M}^{M} \left|\ell''(u)\right|\,du$$

$$= -\epsilon(x - s/2) - \epsilon(x - s/2 - 2M) + \int_{-M}^{M} \left|\ell''(u)\right|\,du$$

$$\leq -\epsilon(x - s/2) - \epsilon(s/2 - 3M) + \int_{-M}^{M} \left|\ell''(u)\right|\,du$$

$$\leq -\epsilon(x - s/2).$$

The penultimate inequality uses $x \geq s - M$, whereas the last inequality holds when s is sufficiently large (since $\int_{-M}^{M} |\ell''(u)|\,du$ is finite by the assumption that h is positive and twice continuously differentiable). This completes the proof.

APPENDIX K: NECESSARY CONDITION FOR LARGE SAMPLE DOMINANCE WITH MANY STATES

In this section, we show that the Rényi order can be generalized to more than two states to yield a general necessary condition for large sample dominance. Consider $k + 1$ states $\theta \in \{0, 1, \ldots, k\}$ and two experiments $P = (\Omega, (P_\theta))$, $Q = (\Xi, (Q_\theta))$ revealing information about these states. Conditioning on $\theta = 0$, we consider the moment generating function of the log-likelihood ratio vector $(\frac{dP_0}{dP_1}, \ldots, \frac{dP_0}{dP_k})$, given by

$$M_{X^0}(t) = \int_\Omega e^{\sum_{j=1}^k t_j \log \frac{dP_0(\omega)}{dP_j(\omega)}} \, dP_0(\omega) \tag{34}$$

with $t = (t_1, \ldots, t_k) \in \mathbb{R}^k$. Similarly, define $M_{Y^0}(t)$ for the experiment Q.

By the same argument as in Section 5.1 (see the derivation of (8)), $M_{X^0}(t)$ *would be* the ex ante expected payoff from observing P, in a decision problem with uniform prior and indirect utility function

$$v(p) = (k + 1) p_0^{1+t_1+\cdots+t_k} \cdot p_1^{-t_1} \cdots p_k^{-t_k},$$

where $p = (p_0, p_1, \ldots, p_k)$ represents the belief about the $k + 1$ states. If the function $v(p)$ were convex in p, then it is indeed an indirect utility function. Blackwell dominance of P over Q then requires $M_{X^0}(t) \geq M_{Y^0}(t)$. Since the moment generating function is raised to the nth power when n i.i.d. samples are drawn, we would be able to conclude that $M_{X^0}(t) \geq M_{Y^0}(t)$ also has to hold if P dominates Q in large samples. If instead $v(p)$ were concave, then $-v(p)$ is an indirect utility function, leading to the reverse ranking between the moment generating functions.

We can characterize those parameters $t = (t_1, \ldots, t_k)$ that make the function $v(p)$ globally convex/concave. To make the result easy to state, we make the variables symmetric and consider a function of the form

$$v(p) = (k + 1) p_0^{\alpha_0} \cdot p_1^{\alpha_1} \cdots p_k^{\alpha_k}$$

with $\alpha_0 + \alpha_1 + \cdots + \alpha_k = 1$.

LEMMA 8: *Consider the function $v(p)$ defined above, over the domain $p \in int(\Delta^k)$. Suppose $\alpha_0 + \alpha_1 + \cdots + \alpha_k = 1$ and $\alpha_0 > 0$. Then $v(p)$ is convex in p if and only if $\alpha_1, \ldots, \alpha_k$ are all nonpositive. Conversely, $v(p)$ is concave in p if and only if $\alpha_1, \ldots, \alpha_k$ are nonnegative. Moreover, the convexity/concavity is strict when $\alpha_1, \ldots, \alpha_k$ are strictly negative/positive.*

The proof of this lemma is deferred to the end of the section. Note that unlike the case of two states, there are situations where $v(p)$ is neither convex nor concave.

By rewriting $\alpha_j = -t_j$ for $1 \leq j \leq k$, we obtain the following necessary condition for Blackwell dominance in large samples. Say the experiments P and Q form a generic pair, if for every pair of states $i \neq j$, the maximum and minimum of $\log \frac{d\widetilde{P_i}}{dP_j}$ differ from those of $\log \frac{dQ_i}{dQ_j}$.

PROPOSITION 7: *Suppose P and Q are a generic pair of bounded experiments for $k + 1$ states. If P Blackwell dominates Q in large samples, then the following conditions hold:*[5]

---

[5] We exclude $t = \{\mathbf{0}\}$ from the conditions because $M_X(\mathbf{0}) = M_Y(\mathbf{0}) = 1$ always holds.

(i) *For all $t \in \mathbb{R}_+^k \backslash \{\mathbf{0}\}$, $M_{X^0}(t) > M_{Y^0}(t)$ and symmetrically $M_{X^i}(t) > M_{Y^i}(t)$ if we define the moment generating functions for true state i analogously to* (34);

(ii) *For all $t \in \mathbb{R}_-^k \backslash \{\mathbf{0}\}$ such that $\sum_{j=1}^k t_j > -1$, $M_{X^0}(t) < M_{Y^0}(t)$ and symmetrically $M_{X^i}(t) < M_{Y^i}(t)$ for $1 \le i \le k$;*

(iii) *For every pair of states $i \ne j$, the Kullback–Leibler divergence between P_i and P_j exceeds the divergence between Q_i and Q_j:*

$$\int_\Omega \log \frac{\mathrm{d}P_i(\omega)}{\mathrm{d}P_j(\omega)} \, \mathrm{d}P_i(\omega) > \int_\Xi \log \frac{\mathrm{d}Q_i(\xi)}{\mathrm{d}Q_j(\xi)} \, \mathrm{d}Q_i(\xi).$$

To understand Proposition 7, note from (34) that when t_j are *all* positive, a bigger value of $M_{X^0}(t)$ indicates higher likelihood ratios $\frac{\mathrm{d}P_0}{\mathrm{d}P_j}$ between state 0 and *every* other state j, when state 0 is the true state. It is intuitive that in this case $M_{X^0}(t) > M_{Y^0}(t)$ corresponds to P being (on average) a more informative experiment than Q.[6] This is the content of part (i), which generalizes the comparison of Rényi divergences $R_P^\theta(t) > R_Q^\theta(t)$ in the two state case, for $t > 1$.

Conversely, part (ii) says that when t_j are all negative (subject to the extra condition $\sum_j t_j > -1$), informativeness is captured by the reverse ranking $M_{X^0}(t) < M_{Y^0}(t)$. In this case, the smaller value of $M_{X^0}(t)$ actually indicates higher likelihood ratios $\frac{\mathrm{d}P_0}{\mathrm{d}P_j}$ under true state 0. This part generalizes the comparison $R_P^\theta(t) > R_Q^\theta(t)$ for $t \in (0, 1)$.

Finally, part (iii) directly imposes the Rényi comparison $R_P^\theta(1) > R_Q^\theta(1)$ when it is applied to every pair of states.

We conjecture that the set of necessary conditions identified in Proposition 7 are also sufficient for large sample Blackwell dominance; see Section 6 for discussion of the difficulties.

Below we supply the proof of Lemma 8.

PROOF OF LEMMA 8: The Hessian matrix of $v(\cdot)$ at p is computed as

$$\mathrm{Hess}_v(p) = v(p) \times \begin{pmatrix} \dfrac{\alpha_0(\alpha_0 - 1)}{p_0^2} & \dfrac{\alpha_0 \alpha_1}{p_0 p_1} & \cdots \\ \dfrac{\alpha_0 \alpha_1}{p_0 p_1} & \dfrac{\alpha_1(\alpha_1 - 1)}{p_1^2} & \cdots \\ \cdots & \cdots & \cdots \end{pmatrix}.$$

For any direction $(x_0, x_1, \ldots, x_k)$, the directional second derivative of $v(\cdot)$ at p is thus

$$(x_0, x_1, \ldots) \cdot \begin{pmatrix} \dfrac{\alpha_0(\alpha_0 - 1)}{p_0^2} & \dfrac{\alpha_0 \alpha_1}{p_0 p_1} & \cdots \\ \dfrac{\alpha_0 \alpha_1}{p_0 p_1} & \dfrac{\alpha_1(\alpha_1 - 1)}{p_1^2} & \cdots \\ \cdots & \cdots & \cdots \end{pmatrix} \cdot \begin{pmatrix} x_0 \\ x_1 \\ \cdots \end{pmatrix} = \left(\sum_{i=0}^k \frac{\alpha_i x_i}{p_i} \right)^2 - \sum_{i=0}^k \frac{\alpha_i x_i^2}{p_i^2}, \quad (35)$$

---

[6]To prove the strict inequality $M_{X^0}(t) > M_{Y^0}(t)$, suppose that $t_1, \ldots, t_l$ are positive whereas $t_{l+1}, \ldots, t_k$ are zero, for some $1 \le l \le k$. Let $\tilde{P} = (\Omega, (P_0, \ldots, P_l))$ be the restriction of the experiment P to the first $l+1$ states; similarly, define $\tilde{Q}$. Then $P^{\otimes n} \succeq Q^{\otimes n}$ implies $\tilde{P}^{\otimes n} \succeq \tilde{Q}^{\otimes n}$, which must in fact be a strict comparison by the genericity assumption. Therefore, as the indirect utility function $\tilde{v}(p_0, \ldots, p_l) = (k+1) p_0^{1+t_1+\cdots+t_l} \cdot p_1^{-t_1} \cdots p_k^{-t_l}$ is *strictly* convex on the smaller belief space Δ^l (Lemma 8), the ex ante expected payoff $M_{X^0}(t)$ must be strictly higher than $M_{Y^0}(t)$.

where for simplicity we have ignored the positive factor $v(p)$ as it does not affect the sign.

We first use this to show that if $\alpha_1 > 0$ (or any $\alpha_j > 0$), then the function $v(p)$ is *not* convex for $p \in int(\Delta^k)$. Indeed, consider the direction $(1, -1, 0, 0, \ldots, 0)$, which maintains $p \in int(\Delta^k)$. The directional second derivative can be computed as

$$\frac{\alpha_0(\alpha_0 - 1)}{p_0^2} - \frac{2\alpha_0\alpha_1}{p_0 p_1} + \frac{\alpha_1(\alpha_1 - 1)}{p_1^2}.$$

Suppose $p_0 = \alpha_0 x$, $p_1 = \alpha_1 x$ for some small positive number x, and $p_2, p_3, \ldots$ are arbitrary. Then the above second derivative simplifies to $-\frac{(\alpha_0 + \alpha_1)}{\alpha_0 \alpha_1 x^2} < 0$. Thus $v(p)$ is not convex along this direction.

Suppose instead $\alpha_1, \ldots, \alpha_k \leq 0$, we will show $v(p)$ is convex. For this it suffices to show the RHS of (35) is nonnegative. Indeed, by the Cauchy–Schwarz inequality,

$$\left(\left(\sum_{i=0}^{k} \frac{\alpha_i x_i}{p_i} \right)^2 + \frac{-\alpha_1 x_1^2}{p_1^2} + \cdots + \frac{-\alpha_k x_k^2}{p_k^2} \right) \cdot \left(1 + (-\alpha_1) + \cdots + (-\alpha_k) \right)$$

$$\geq \left(\sum_{i=0}^{k} \frac{\alpha_i x_i}{p_i} + \frac{-\alpha_1 x_1}{p_1} + \cdots + \frac{-\alpha_k x_k}{p_k} \right)^2 = \left(\frac{\alpha_0 x_0}{p_0} \right)^2.$$

Using $\alpha_0 + \alpha_1 + \cdots + \alpha_k = 1$ to simplify, this exactly implies $(\sum_{i=0}^{k} \frac{\alpha_i x_i}{p_i})^2 \geq \sum_{i=0}^{k} \frac{\alpha_i x_i^2}{p_i^2}$ as desired. In fact, $v(p)$ is convex for all $p \gg 0$, including $p \in int(\Delta^k)$.

Moreover, if $\alpha_1, \ldots, \alpha_k$ are *strictly* negative, then the equality condition of the Cauchy–Schwarz inequality above requires $\sum_{i=0}^{k} \frac{\alpha_i x_i}{p_i} = \frac{x_1}{p_1} = \cdots = \frac{x_k}{p_k}$, which in turn implies that $x_0, x_1, \ldots, x_k$ have the same sign (under the assumption $\alpha_0 > 0 > \alpha_1, \ldots, \alpha_k$). Thus, for any direction $(x_0, x_1, \ldots, x_k)$ with $x_0 + x_1 + \cdots + x_k = 0$, the directional second derivative of v is strictly positive. So v is strictly convex for $p \in int(\Delta^k)$.

Next, we will show that if $\alpha_1 < 0$ (or any $\alpha_j < 0$), then the function $v(p)$ is *not* concave for $p \in int(\Delta^k)$. For this, we again consider the second derivative along the direction $(1, -1, 0, 0, \ldots, 0)$, which is $\frac{\alpha_0(\alpha_0-1)}{p_0^2} - \frac{2\alpha_0\alpha_1}{p_0 p_1} + \frac{\alpha_1(\alpha_1-1)}{p_1^2}$. As $\alpha_1 < 0$, we have $\alpha_1(\alpha_1 - 1) > 0$. Thus for p_0 close to 1 and p_1 close to 0, the above second derivative is positive and $v(p)$ is not concave along this direction.

Finally, we show that if $\alpha_1, \ldots, \alpha_k \geq 0$, then the function $v(p)$ is concave. By the Cauchy–Schwarz inequality,

$$\left(\sum_{i=0}^{k} \frac{\alpha_i x_i^2}{p_i^2} \right) \cdot \left(\sum_{i=0}^{k} \alpha_i \right) \geq \left(\sum_{i=0}^{k} \frac{\alpha_i x_i}{p_i} \right)^2.$$

Since $\sum_{i=0}^{k} \alpha_i = 1$, this implies the RHS of (35) is nonpositive. Hence v has nonpositive directional second derivatives and must be globally concave.

Moreover, if $\alpha_1, \ldots, \alpha_k$ are strictly positive, then the equality condition of the Cauchy–Schwarz inequality requires $\frac{x_0}{p_0} = \frac{x_1}{p_1} = \cdots = \frac{x_k}{p_k}$, which in turn requires $x_0, x_1, \ldots, x_k$ to have the same sign. By the same argument as above, we conclude that in this case v is strictly concave for $p \in int(\Delta^k)$. *Q.E.D.*

APPENDIX L: PROOF OF A CONJECTURE REGARDING MAJORIZATION

Jensen (2019) studied the majorization order on finitely supported distributions. Given two such distributions μ and ν, μ is said to *majorize* ν if for every $n \geq 1$ it holds that the sum of the largest n probabilities in μ is greater than or equal to the sum of the n largest probabilities in ν. The Rényi entropy of a distribution μ defined on a finite set S is given by

$$H_\mu(\alpha) = \frac{1}{1-\alpha} \log\left(\sum_{s \in S} \mu(s)^\alpha\right),$$

for $\alpha \in [0, \infty) \setminus \{1\}$. As with our definition of Rényi divergences, this definition is extended to $\alpha = 1$ by continuity to equal the Shannon entropy, and extended to $\alpha = \infty$ to equal $-\log\max_s \mu(s)$. Hence H_μ is defined on $[0, \infty]$.

Note that $H_\mu(0)$ is the size of the support of μ. In his Proposition 3.7, Jensen showed that if $H_\mu(\alpha) < H_\nu(\alpha)$ for all $\alpha \in [0, \infty]$ then the n-fold product $\mu^{\times n}$ majorizes $\nu^{\times n}$.

Commenting on his Proposition 3.7, Jensen wrote "The author cautiously conjectures that … the requirement of a sharp inequality at 0 could be replaced by a similar condition regarding the α-Rényi entropies for negative α."

To understand this statement in terms of the nomenclature and notation of our paper, we identify each distribution μ whose support is a finite set S with the experiment $P^\mu = (S, P_1, P_0)$, where $P_1 = \mu$ and P_0 is the uniform distribution on S. There is a simple connection between the Rényi entropy of μ and the Rényi divergence of P^μ. For $\alpha \geq 0$,

$$H_\mu(\alpha) = \log|S| - R_P^1(\alpha). \tag{36}$$

As Jensen suggests, $H_\mu(\alpha)$ for negative α is also important, as it relates to R_P^0. For $\alpha \leq 0$,

$$H_\mu(\alpha) = \log|S| - \frac{\alpha}{1-\alpha} R_P^0(1-\alpha), \tag{37}$$

which extends to $\alpha = -\infty$ to equal $-\log\min_s \mu(s)$. Moreover, note that

$$H_\mu'(0) = -R_P^0(1) = \log|S| + \frac{1}{|S|} \sum_{s \in S} \log\mu(s). \tag{38}$$

As shown by Torgersen (1985, p. 264), when μ and ν have the same support size, then majorization of ν by μ is equivalent to Blackwell dominance of P^μ over P^ν. Thus Jensen's Proposition 3.7, which assumes that the support sizes are different, has no implications for Blackwell dominance. However, our result on Blackwell dominance does have implications for majorization. In particular, the following proposition follows immediately from the application of Theorem 1 to experiments of the form P^μ.

PROPOSITION 8: *Let μ, ν be finitely supported distributions with the same support size (i.e., $H_\mu(0) = H_\nu(0)$), and such that $H_\mu(\infty) \neq H_\nu(\infty)$ and $H_\mu(-\infty) \neq H_\nu(-\infty)$. Then the following are equivalent*:
  (i) $H_\mu(\alpha) < H_\nu(\alpha)$ for all $\alpha \in (0, \infty]$, $H_\mu(\alpha) > H_\nu(\alpha)$ for all $\alpha \in [-\infty, 0)$ and $H_\mu'(0) < H_\nu'(0)$.[7]

---

[7]This last condition is necessary for majorization, but it was not recognized in the original conjecture of Jensen (2019).

(ii) *There exists an n_0 such that $\mu^{\times n}$ majorizes $\nu^{\times n}$ for every $n \geq n_0$.*

PROOF: For notational ease, let P denote P^μ and Q denote P^ν. The assumption $H_\mu(\alpha) < H_\nu(\alpha)$ for all $\alpha > 0$ is equivalent, via (36), to $R_P^1(t) > R_Q^1(t)$ for all $t > 0$, and to $R_P^0(t) > R_Q^0(t)$ for all $t \in (0, 1)$, using $R_P^0(t) = \frac{t}{1-t} R_P^1(1 - t)$ for $0 < t < 1$.

On the other hand, $H_\mu(\alpha) > H_\nu(\alpha)$ for all $\alpha < 0$ and $H'_\mu(0) < H'_\nu(0)$ is equivalent, via (37) and (38), to $R_P^0(t) > R_Q^0(t)$ for all $t \geq 1$. So (i) is equivalent to P dominating Q in the Rényi order.

Finally, the assumptions that $H_\mu(\infty) \neq H_\nu(\infty)$ and $H_\mu(-\infty) \neq H_\nu(-\infty)$ translate into $\max_s \mu(s) \neq \max_s \nu(s)$ and $\min_s \mu(s) \neq \min_s \nu(s)$, which are in turn equivalent to requiring that P and Q be a generic pair. Therefore, by Theorem 1, (i) is equivalent to $P^{\otimes n}$ Blackwell dominates $Q^{\otimes n}$ for every large n. It follows from Torgersen (1985) that (i) is equivalent to (ii).   *Q.E.D.*

REFERENCES

ALIPRANTIS, C. D., AND K. BORDER (2006): *Infinite Dimensional Analysis*. Springer. [5]

BOGACHEV, V. I. (2007): *Measure Theory*, Vol. 1. Springer Science & Business Media. [2]

DUAN, R., Y. FENG, X. LI, AND M. YING (2005): "Multiple-Copy Entanglement Transformation and Entanglement Catalysis," *Physical Review A*, 71 (4), 042319. [2]

FRITZ, T. (2017): "Resource Convertibility and Ordered Commutative Monoids," *Mathematical Structures in Computer Science*, 27 (6), 850–938. [2]

JENSEN, A. K. (2019): "Asymptotic Majorization of Finite Probability Distributions," *IEEE Transactions on Information Theory*, 65 (12), 8131–8139. [1,18]

KANTOROVICH, L. (1937): "On the Moment Problem for a Finite Interval," *Dokl. Akad. Nauk SSSR*, 14, 531–537. [5]

TORGERSEN, E. (1985): "Majorization and Approximate Majorization for Families of Measures, Applications to Local Comparison of Experiments and the Theory of Majorization of Vectors in r n (Schur Convexity)," in *Linear Statistical Inference*. Springer, 259–310. [18,19]