From Blackwell dominance in large samples to Rényi Divergences and back again

Xiaosheng Mu, Luciano Pomatto, Philipp Strack, Omer Tamuz

Princeton University
Caltech University
Yale University
Caltech University

May 6, 2022

Introduction

- Blackwell order provides a partial order for comparing experiments in terms of their informativeness.
- However, it is demanding for two experiments to be ranked in the Blackwell order.
- In many applications, an experiment does not consist of a single observation but of multiple i.i.d. samples.
- We study a weakening of the Blackwell order that is appropriate for comparing experiments in terms of their large sample properties.

Model: Statistical Experiments

- Consider binary state of world: $\theta \in \{0,1\}$.
- A Blackwell-Le Cam experiment

$$P = (\Omega, P_0, P_1)$$

  consists of a sample space Ω and a pair of Borel probability measures (P_0, P_1) defined over Ω, where $P_\theta(A)$ is the probability of observing $A \subseteq \Omega$ in state $\theta \in \{0,1\}$.
- Given two experiments $P = (\Omega, P_0, P_1)$ and $Q = (\Xi, Q_0, Q_1)$, we can form the product experiment $P \otimes Q$ given by

$$P \otimes Q = (\Omega \times \Xi, P_0 \times Q_0, P_1 \times Q_1)\,,$$

  where $P_\theta \times Q_\theta$, given $\theta \in \{0,1\}$, denotes the product of the two measures.
- We denote by

$$P^{\otimes n} = P \otimes \cdots \otimes P$$

  the n-fold product experiment where n independent observations are generated according to the experiment P.

Model: Statistical Experiments

- Consider a Bayesian decision maker whose prior belief assigns probability $\frac{1}{2}$ to the state being 1.
- To each experiment $P = (\Omega, P_0, P_1)$, we associate a Borel probability measure π over $[0, 1]$ that represents the distribution over posterior beliefs induced by the experiment.
- Formally, let $p(\omega)$ be the posterior belief that the state is 1 given the realization $\omega \in \Omega$:
$$p(\omega) = \frac{\mathrm{d}P_1(\omega)}{\mathrm{d}P_1(\omega) + \mathrm{d}P_0(\omega)}.$$
- Furthermore, define for every Borel set $B \subseteq [0, 1]$
$$\pi_\theta(B) = P_\theta(\{\omega : p(\omega) \in B\})$$
as the probability that the posterior belief will belong to B, given state θ.
- We then define $\pi = (\pi_0 + \pi_1)/2$ as the unconditional measure over posterior beliefs.

The Blackwell Order

- Consider two experiments P and Q and their induced distribution over posterior beliefs denoted by π and τ, respectively.
- The experiment P Blackwell dominates Q, denoted $P \succeq Q$, if

$$\int_0^1 v(p)\mathrm{d}\pi(p) \geq \int_0^1 v(p)\mathrm{d}\tau(p)$$

  for every convex function $v : (0,1) \to \mathbb{R}$.
- Equivalently, $P \succeq Q$ if π is a mean-preserving spread of τ. We write $P \succ Q$ if $P \succeq Q$ and $Q \not\succeq P$. So, $P \succ Q$ if and only if (1) holds with a strict inequality whenever v is strictly convex, that is, π is a *mean-preserving spread* of τ and $\pi \neq \tau$.

The Blackwell Order

Definition (1. Large Sample Order)

An experiment P dominates an experiment Q in large samples if there exists an $n_0 \in \mathbb{N}$ such that

$$P^{\otimes n} \succeq Q^{\otimes n} \quad \text{for every } n \geq n_0.$$

- As shown by Blackwell (1951, Theorem 12), dominance of P over Q implies dominance of $P^{\otimes n}$ over $Q^{\otimes n}$, for every n. So dominance in large samples is an extension of the Blackwell order.
- This extension is strict, as shown by examples in Torgersen (1970) and Azrieli (2014).

Rényi Divergence and the Rényi Order

- Given two probability measures μ, ν on a measurable space Ω and a parameter $t > 0$, the Rényi t-divergence is given by

$$R_t(\mu \| \nu) = \frac{1}{t-1} \log \int_\Omega \left(\frac{\mathrm{d}\mu}{\mathrm{d}\nu}(\omega) \right)^{t-1} \mathrm{d}\mu(\omega)$$

  when $t \neq 1$, and, ensuring continuity,

$$R_1(\mu \| \nu) = \int_\Omega \log \left(\frac{\mathrm{d}\mu}{\mathrm{d}\nu}(\omega) \right) \mathrm{d}\mu(\omega).$$

- Equivalently, $R_1(\mu \| \nu)$ is the Kullback-Leibler divergence between the measures μ and ν.

Rényi Divergence and the Rényi Order

- Given an experiment $P = (\Omega, P_0, P_1)$, a state θ, and parameter $t > 0$, the Rényi t-divergence of P under θ is

$$R_P^\theta(t) = R_t\left(P_\theta \| P_{1-\theta}\right).$$

- Intuitively, observing a sample realization for which the likelihood ratio $\mathrm{d}P_\theta/\mathrm{d}P_{1-\theta}$ is high constitutes evidence that favors state θ over $1-\theta$.
- The two Rényi divergences R_P^1 and R_P^0 of an experiment are related by the identity

$$R_P^1(t) = \frac{t}{1-t} R_P^0(1-t).$$

- Hence the values of $R_P^\theta(t)$ for $t \in [0, 1/2]$ are determined by the values of $R_P^{1-\theta}(t)$ on the interval $[1/2, 1]$. Thus, it suffices to consider values of t in $[1/2, \infty]$.

Rényi Divergence and the Rényi Order

Definition (2. Rényi Order)

An experiment P dominates an experiment Q in the **Rényi order** if it holds that for all $\theta \in \{0,1\}$ and all $t > 0$

$$R_P^\theta(t) > R_Q^\theta(t).$$

- The Rényi order is a extension of the (strict) Blackwell order.
- A simple calculation shows that if $P = S \otimes T$ is the product of two experiments, then for every state θ,

$$R_P^\theta = R_S^\theta + R_T^\theta.$$

- Additivity:

$$\prod_{n=1}^{N} \int \left(\frac{\mathrm{d}Q_n}{\mathrm{d}P_n} \right)^{1-\alpha} \mathrm{d}P_n = \int \left(\frac{\mathrm{d}\prod_{n=1}^{N} Q_n}{\mathrm{d}\prod_{n=1}^{N} P_n} \right)^{1-\alpha} \mathrm{d}\prod_{n=1}^{N} P_n$$

- Hence, the Rényi order compares experiments in terms of properties that are unaffected by the number of samples.

Characterization of Large Sample Order

- We say two bounded experiments P and Q form a **generic** pair if the essential maxima of the log-likelihood ratios $\log \frac{\mathrm{d}P_1}{\mathrm{d}P_0}$ and $\log \frac{\mathrm{d}Q_1}{\mathrm{d}Q_0}$ are different, and if their essential minima are also different.

Theorem (1.)

For a generic pair of bounded experiments P and Q, the following are equivalent:

- P dominates Q in large samples.
- P dominates Q in the Rényi order.

- A natural alternative definition of "Blackwell dominance in large samples" would require $P^{\otimes n} \succeq Q^{\otimes n}$ to hold for some n, but the resulting order is in fact equivalent under our genericity assumption. This is a consequence of Theorem 1, because $P^{\otimes n_0} \succeq Q^{\otimes n_0}$ for any n_0 implies P dominates Q in the Rényi order, which in turn implies $P^{\otimes n} \succeq Q^{\otimes n}$ for all large n.

Characterization of Large Sample Order

- We start with the observation that any two experiments, even if not ranked according to dominance in large samples, can be compared by applying different samples sizes.
- For example, suppose P and Q are not comparable, but $P^{\otimes 50}$ Blackwell dominates $Q^{\otimes 100}$. Then 50 samples from P are more informative than 100 from Q, and thus, in an intuitive sense, P is at least twice as informative as Q, for large enough samples.
- This leads to a well-defined measure of dominance, which we refer to as the dominance ratio
$$P/Q = \sup\left\{\frac{m}{n} : P^{\otimes n} \succeq Q^{\otimes m}\right\},$$
P will be at least m/n times as informative as Q in large samples.
- Thus, in large samples, each observation from P contributes at least as much as P/Q observations from Q.

Characterization of Large Sample Order

Proposition (3.)

Let P and Q be nontrivial, bounded experiments. Then

$$P/Q = \inf_{\substack{\theta \in 0,1\} \\ t>0}} \frac{R_P^\theta(t)}{R_Q^\theta(t)}.$$

Furthermore, the dominance ratio P/Q is always positive.

Blackwell Order With Additional Information

- Consider a decision maker choosing which of two experiments P and Q to conduct, on top of an independent source of information R.
- The decision maker chooses between two compound experiments $P \otimes R$ and $Q \otimes R$.
- If P dominates Q in the Blackwell order, then the same relation must hold between the two compound experiments.
- How about P and Q are incomparable?

Blackwell Order With Additional Information

Proposition (4.)

Let P and Q be a generic pair of bounded experiments. Then the following are equivalent:

- There exists a bounded experiment R such that $P \otimes R \succeq Q \otimes R$.
- P dominates Q in the Rényi order.

- The same holds for any experiment R' that is more informative than R, that is, $P \otimes R' \succeq Q \otimes R'$.
- Proposition 4 follows by combining the characterization in Theorem 1 together with the observation that if P dominates Q in the large sample order, then there exists an R such that $P \otimes R$ Blackwell dominates $Q \otimes R$ (from quantum information literature).

A Characterization of Additive Divergence

- We apply the characterization of Blackwell dominance in large samples to study **divergence**.
- Examples of divergences include total variation distance, the Hellinger distance, the Kullback–Leibler divergence, Rényi divergences, and more general f-divergences.
- Two key properties of Rényi divergences are additivity and data processing inequality.
- Given a Polish space Ω, we denote by $\mathcal{B}(\Omega)$ its Borel σ-algebra and by $\Delta(\Omega)$ the collection of Borel probability measures on $\mathcal{B}(\Omega)$.
- Given another Polish space Ξ, a measurable function $f : \Omega \to \Xi$ and a probability measure $\mu \in \Delta(\Omega)$, we denote by $f_*(\mu)$ the push-forward probability measure in $\Delta(\Xi)$ defined as

$$[f_*(\mu)](E) = \mu\left(f^{-1}(E)\right),$$

for all $E \in \mathcal{B}(\Xi)$.

A Characterization of Additive Divergence

- Consider, for each Ω, a map

$$D_\Omega : \Delta(\Omega) \times \Delta(\Omega) \to \mathbb{R}_+ \cup \{+\infty\},$$

and let $D = (D_\Omega)$ be the collection obtained by varying Ω. We say D is a divergence if $D_\Omega(\mu, \mu) = 0$ for all Ω and all $\mu \in \Delta(\Omega)$.

- A divergence satisfies the **data processing inequality** if for any measurable $f : \Omega \to \Xi$ it holds that

$$D_\Xi\left(f_*(\mu), f_*(\nu)\right) \leq D_\Omega(\mu, \nu).$$

The data processing inequality captures the idea that the distributions of two random variables X and Y are at least as dissimilar as those of $f(X)$ and $f(Y)$.

- We say that the divergence D is **additive** if

$$D_{\Omega\times\Xi}\left(\mu_1 \times \mu_2, \nu_1 \times \nu_2\right) = D_\Omega\left(\mu_1, \nu_1\right) + D_\Xi\left(\mu_2, \nu_2\right).$$

Representation Theorem

Theorem (2.)

Let D be an additive divergence that satisfies the data processing inequality and is finite on bounded experiments. Then there exist two finite Borel measures m_0, m_1 on $[1/2, \infty]$ such that for every bounded pair μ, ν it holds that

$$D(\mu, \nu) = \int_{[1/2,\infty]} R_t(\mu \| \nu) \mathrm{d}m_0(t) + \int_{[1/2,\infty]} R_t(\nu \| \mu) \mathrm{d}m_1(t)$$

with R_t given by

$$R_t(\mu \| \nu) = \frac{1}{t-1} \log \int_\Omega \left(\frac{\mathrm{d}\mu}{\mathrm{d}\nu}(\omega) \right)^{t-1} \mathrm{d}\mu(\omega)$$

$$R_1(\mu \| \nu) = \int_\Omega \log \left(\frac{\mathrm{d}\mu}{\mathrm{d}\nu}(\omega) \right) \mathrm{d}\mu(\omega).$$

Proof Sktech

- Regard D as a functional over experiments. When D is **additive**, the data processing inequality implies **monotonicity** with respect to the Blackwell order.
- The next crucial step is to leverage Theorem 1 to show that additivity renders D monotone in the Rényi order.
- Indeed, if (P_0, P_1) dominates (Q_0, Q_1) in the Rényi order, then, by Theorem 1, there exists a number n of repetitions such that (P_0^n, P_1^n) dominates (Q_0^n, Q_1^n) in the Blackwell order. Hence, by combining Blackwell monotonicity and additivity, we obtain that D must satisfy

$$nD\left(P_0, P_1\right) = D\left(P_0^n, P_1^n\right) \geq D\left(Q_0^n, Q_1^n\right) = nD\left(Q_0, Q_1\right)$$

  Hence, D is monotone in the Rényi order.
- We deduce from this that D is a monotone functional $F\left(R_P^0, R_P^1\right)$ of the Rényi divergences of the experiment. Additivity of D implies F is also additive. We then use tools from functional analysis to show that F extends to a positive linear functional, leading to the integral representation of Theorem 2.

Dominance in Large Samples Implies Dominance in the Rényi Order

- As is well known, each convex function v can be seen as the indirect utility induced by some decision problem. That is, for each convex v there exists a set of actions A and a utility function u defined on $A \times \{0,1\}$ such that $v(p)$ is the maximal expected payoff that a decision maker can obtain in such a decision problem given a belief p. Hence, $P \succeq Q$ if and only if in every decision problem, an agent can obtain a higher payoff by basing her action on the experiment P rather than on Q.
- Since Rényi divergences between two experiments is independent of the number of samples, it is sufficient to show that the Rényi order extends the strict Blackwell order.
- We do this by constructing decision problems with the property that higher expected payoff in these problems translates into higher Rényi divergences.

Dominance in Large Samples Implies Dominance in the Rényi Order

- For each $t > 1$, the function $v_1(p) = 2p^t(1-p)^{1-t}$ defined for $p \in (0,1)$ is strictly convex.
- Thus $v_1(p)$ is the indirect utility function induced by some decision problem. Moreover, we have that

$$\int_0^1 v_1(p)\mathrm{d}\pi(p) = \int_\Omega \left(\frac{\mathrm{d}P_1(\omega)}{\mathrm{d}P_0(\omega)}\right)^{t-1} \mathrm{d}P_1(\omega) = \mathrm{e}^{(t-1)R_p^1(t)},$$

where $\mathrm{d}\pi(p) = \frac{1}{2}\left(\mathrm{d}\pi_1(p) + \mathrm{d}\pi_0(p)\right)$ and $\mathrm{d}\pi_1(p) = \frac{p}{1-p}\,\mathrm{d}\pi_0(p)$.

- Thus $\mathrm{d}\pi(p) = \frac{1}{2p}\,\mathrm{d}\pi_1(p)$, which allows us to write

$$\int_0^1 v_1(p)\mathrm{d}\pi(p) = \int_0^1 2p^t(1-p)^{1-t}\cdot\frac{1}{2p}\,\mathrm{d}\pi_1(p) = \int_0^1 \left(\frac{p}{1-p}\right)^{t-1} \mathrm{d}\pi_1(p).$$

- Hence, experiment P yields higher expected payoff in this decision problem than Q if and only if $R_P^1(t) > R_Q^1(t)$.

Dominance in Large Samples Implies Dominance in the Rényi Order

- Similarly, for $t \in (0,1)$ we consider the indirect utility function $v_2(p) = -2p^t(1-p)^{1-t}$, which is now strictly convex due to the negative sign (its second derivative is $2t(1-t)\times p^{t-2}(1-p)^{-1-t}$). Then

$$\int_0^1 v_2(p)\mathrm{d}\pi(p) = -\mathrm{e}^{(t-1)R_P^1(t)}$$

  is again a monotone transformation of the Rényi divergence.

- For $t = 1$, we consider the indirect utility function $v_3(p) = 2p\log\left(\frac{p}{1-p}\right)$, which is strictly convex with a second derivative of $2p^{-1}(1-p)^{-2}$. We have

$$\int_0^1 v_3(p)\mathrm{d}\pi(p) = \int_0^1 \log\left(\frac{p}{1-p}\right)\mathrm{d}\pi_1(p) = \int_\Omega \log\left(\frac{\mathrm{d}P_1(\omega)}{\mathrm{d}P_0(\omega)}\right)\mathrm{d}P_1(\omega) = R_P^1(1)$$

  Thus P yields higher expected payoff in this problem if and only if $R_P^1(1) > R_Q^1(1)$.

- Summarizing, the above family of decision problems shows that P strictly Blackwell dominates Q only if $R_P^1(t) > R_Q^1(t)$ for all $t > 0$.

Repeated Experiments and log-Likelihood Ratios

- We turn to the proof that dominance in the Rényi order is (generically) sufficient for dominance in large samples.
- Recall that $P^{\otimes n}$ Blackwell dominates $Q^{\otimes n}$ if and only if the former induces a distribution over posterior beliefs that is a mean-preserving spread of the latter.
- However, the distribution over posteriors induced by a product experiment can be difficult to analyze directly.
- A more suitable approach consists in studying the distribution of the induced log-likelihood ratio

$$\log \frac{\mathrm{d}P_\theta}{\mathrm{d}P_{1-\theta}}.$$

- As is well known, given a repeated experiment $P^{\otimes n} = (\Omega^n, P_0^n, P_1^n)$, its log-likelihood ratio satisfies, for every realization $\omega = (\omega_1, \ldots, \omega_n)$ in Ω^n,

$$\log \frac{\mathrm{d}P_1^n}{\mathrm{d}P_0^n}(\omega) = \sum_{i=1}^{n} \log \frac{\mathrm{d}P_1}{\mathrm{d}P_0}(\omega_i).$$

Repeated Experiments and log-Likelihood Ratios

- Moreover, the random variables

$$X_i(\omega) = \log \frac{\mathrm{d}P_1}{\mathrm{d}P_0}(\omega_i) \quad i = 1, \ldots, n$$

are i.i.d. under P_θ^n, for $\theta \in \{0, 1\}$.

- Focusing on the distributions of log-likelihood ratios will allow us to transform the study of repeated experiments to the study of sums of i.i.d. random variables.

From Blackwell Dominance to FOSD

- We provide a novel characterization of the Blackwell order, expressed in terms of the **distributions of the log-likelihood ratios**.
- Given two experiments $P = (\Omega, P_0, P_1)$ and $Q = (\Xi, Q_0, Q_1\ )$ we denote by F_θ and G_θ, respectively, the cumulative distribution function of the log-likelihood ratios conditional on state θ. That is,

$$F_\theta(a) = P_\theta\left(\left\{\log \frac{\mathrm{d}P_\theta}{\mathrm{d}P_{1-\theta}} \leq a\right\}\right) \qquad \text{for all } a \in \mathbb{R}, \theta \in \{0,1\}.$$

  The c.d.f. G_θ is defined analogously using Q_θ.
- We associate to P a new quantity, which we call the perfected log-likelihood ratio of the experiment. Define

$$\tilde{L}_1 = \log \frac{\mathrm{d}P_1}{\mathrm{d}P_0} - E,$$

  where E is a random variable that, under P_1, is independent from $\log \frac{\mathrm{d}P_1}{\mathrm{d}P_0}$ and distributed according to an exponential distribution with support $\mathbb{R}_+$and cumulative distribution function $1 - \mathrm{e}^{-x}$ for all $x \geq 0$. Let $\tilde{F}_1$ denote C.D.F of $\tilde{L}_1$.

From Blackwell Dominance to FOSD

Theorem (3.)

Let P and Q be two experiments, and let $\tilde{F}_1$ and $\tilde{G}_1$, respectively, be the associated distributions of perfected log-likelihood ratios. Then

$$P \succeq Q \quad \text{if and only if} \quad \tilde{F}_1(a) \leq \tilde{G}_1(a) \text{ for all } a \in \mathbb{R}.$$

- The Blackwell order over experiments can be reduced to first-order stochastic dominance of the corresponding perfected log-likelihood ratios.
- As is well known, Blackwell dominance is equivalent to the requirement that π is a mean-preserving spread of τ. Equivalently the functions defined as

$$\Lambda_\pi(p) = \int_{[0,p]} (p-q)\mathrm{d}\pi(q) \quad \text{and} \quad \Lambda_\tau(p) = \int_{[0,p]} (p-q)\mathrm{d}\tau(q)$$

  must satisfy $\Lambda_\pi(p) \geq \Lambda_\tau(p)$ for every $p \in (0,1)$. (Not all convex functions, but $(x-t)^+$).

Large Deviations

- Large deviations theory studies low probability events, and in particular the odds with which an i.i.d. sum deviates from its expectation.
- The law of large numbers implies that for a random variable X, the probability of the event $\{X_1 + \cdots + X_n > na\}$ is low for $a > \mathbb{E}[X]$ and large n, where $X_1, \ldots, X_n$ are i.i.d. copies of X.
- A crucial insight due to Cramér (1938) is that the order of magnitude of the probability of this event is determined by the cumulant generating function of X, defined as

$$K_X(t) = \log \mathbb{E}\left[\mathrm{e}^{tX}\right].$$

- As is well known, K_X is strictly convex whenever X is not a constant. We denote by

$$K_X^*(a) = \sup_{t \in \mathbb{R}} t \cdot a - K_X(t) \quad a \in \mathbb{R},$$

  its Fenchel conjugate.
- Two facts we will repeatedly apply are that for every $a \in (\min[X], \max[X])$ the problem has a unique solution $t \in \mathbb{R}$, and such t is nonnegative if and only if $a \geq \mathbb{E}[X]$. Moreover, $K_X^* \geq 0 \cdot a - K_X(0) = 0$ is nonnegative.

Large Deviations

- In this paper, we are interested in comparing the probabilities of large deviations across different random variables.
- Consider, to this end, two random variables X and Y and a threshold a strictly greater than $\mathbb{E}[X]$ and $\mathbb{E}[Y]$. If

$$K_Y^*(a) > K_X^*(a),$$

then the probability of the event $\{X_1 + \cdots + X_n > na\}$ vanishes more slowly than the probability of the event $\{Y_1 + \cdots + Y_n > na\}$. Thus there exists n sufficiently large such that

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

Proposition (5.)

Let X and Y be random variables taking values in $[-b, b]$ and 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 $n \geq 4b^2(1 + \eta)\eta^{-3}$, it holds that

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

Application to the Rényi Order

- Now consider two experiments $P = (\Omega, P_0, P_1)$ and $Q = (\Xi, Q_0, Q_1)$. Denote the corresponding log-likelihood ratios

$$X^\theta = \log \frac{\mathrm{d}P_\theta}{\mathrm{d}P_{1-\theta}} \quad \text{and} \quad Y^\theta = \log \frac{\mathrm{d}Q_\theta}{\mathrm{d}Q_{1-\theta}}$$

  defined over the probability spaces (Ω, P_θ) and (Ξ, Q_θ), respectively.
- The cumulant generating function of the log-likelihood ratio is a simple transformation of the Rényi divergences

$$K_{X^\theta}(t) = t \cdot R_P^\theta(t+1).$$

- Hence, if P dominates Q in the Rényi order then the following relation must hold between the cumulant generating functions:

$$\begin{aligned} K_{X^\theta}(t) &> K_{Y^\theta}(t) \quad \text{for } t > 0 \\ K_{X^\theta}(t) &< K_{Y^\theta}(t) \quad \text{for } -1 < t < 0. \end{aligned}$$

Application to the Rényi Order

- The Fenchel conjugate is an order-reversing operation: From (15), we see that if $K_X \geq K_Y$ pointwise, then the corresponding conjugates satisfy $K_Y^* \geq K_X^*$ pointwise.
- The relation between K_Y^* and K_X^* implies the following ranking:

$$\begin{array}{ll} K_{Y^\theta}^*(a) > K_{X^\theta}^*(a) & \text{for } \mathbb{E}\left[X^\theta\right] \leq a \leq \max\left[Y^\theta\right] \\ K_{Y^\theta}^*(a) < K_{X^\theta}^*(a) & \text{for } 0 \leq a \leq \mathbb{E}\left[Y^\theta\right] \end{array}$$

Lemma (2.)

Suppose P and Q are a generic pair of bounded experiments such that P dominates Q in the Rényi order. Let $\left(X^\theta\right)$ and $\left(Y^\theta\right)$ be the corresponding log-likelihood ratios. Then there exists $\eta \in (0,1)$ such that in both states $\theta \in \{0,1\}$:

$$\begin{array}{ll} K_{Y^\theta}^*(a) - \eta > K_{X^\theta}^*(a+\eta) & \textit{for } \mathbb{E}\left[X^\theta\right] - \eta \leq a \leq \max\left[Y^\theta\right] \\ K_{Y^\theta}^*(a-\eta) < K_{X^\theta}^*(a) - \eta & \textit{for } 0 \leq a \leq \mathbb{E}\left[Y^\theta\right] + \eta \end{array}$$

Rényi Order Implies Large Sample Order

- We now complete the proof of Theorem 1 and show that if two experiments are ranked in the Rényi order then they are also ranked in the large sample order.
- By Theorem 3, we need to show that there exists a sample size n_0 such that for all $n \geq n_0$, the perfected log-likelihood ratios of n independent draws from P and Q are ordered in terms of first-order stochastic dominance.
- More concretely, consider the log-likelihood ratios X^θ and Y^θ (for a single sample) as defined above, with distributions F_θ and G_θ conditional on state θ. Let F_θ^{*n} be the nth convolution power of F_θ, which represents the distribution of log-likelihood ratios under the product experiment $P^{\otimes n}$; similarly define G_θ^{*n}. By Lemma 1, it suffices to show that for $n \geq n_0$ it holds that

$$F_1^{*n}(na) \leq G_1^{*n}(na) \quad \text{for all } a \geq 0$$

and

$$F_0^{*n}(na) \leq G_0^{*n}(na) \quad \text{for all } a \geq 0.$$

Rényi Order Implies Large Sample Order

- Let $X_1^1, \ldots, X_n^1$ be i.i.d. copies of X^1 and $Y_1^1, \ldots, Y_n^1$ be i.i.d. copies of Y^1. We can restate $F_1^{*n}(na) \leq G_1^{*n}(na)$ for all $a \geq 0$ as

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

- We will divide the prove into 4 cases.
- Case 1: $a \geq \max\left[Y^1\right]$. In this case the right-hand side of (22) is 1 , and hence the result follows trivially.
- Case 2: $\mathbb{E}\left[X^1\right] - \eta \leq a < \max\left[Y^1\right]$. From Lemma 2 , we have that

$$K_{Y^1}^*(a) - \eta > K_{X^1}^*(a + \eta).$$

As $a \geq \mathbb{E}\left[X^1\right] - \eta > \mathbb{E}\left[Y^1\right]$, we can directly apply Proposition 5 and conclude that it holds for all $n \geq 4b^2(1+\eta)\eta^{-3}$. Since $\eta < 1$, it holds for all $n \geq n_0 = 8b^2\eta^{-3}$.

Rényi Order Implies Large Sample Order

- Case 3: $\mathbb{E}\left[Y^1\right] + \eta \leq a < \mathbb{E}\left[X^1\right] - \eta$. By the Chebyshev inequality,

$$\mathbb{P}\left[X_1^1 + \cdots + X_n^1 \leq na\right] \leq \mathbb{P}\left[X_1^1 + \cdots + X_n^1 \leq n\left(\mathbb{E}\left[X^1\right] - \eta\right)\right] \leq \frac{\mathrm{Var}\left(X_1^1 + \cdots + X_n^1\right)}{n^2\eta^2}$$

Since $\mathrm{Var}\left(X_1^1 + \cdots + X_n^1\right) = n\,\mathrm{Var}\left(X^1\right) \leq nb^2$, we have that

$$\mathbb{P}\left[X_1^1 + \cdots + X_n^1 \leq na\right] \leq \frac{b^2}{n\eta^2}$$

By a similar argument,

$$\mathbb{P}\left[Y_1^1 + \cdots + Y_n^1 \leq na\right] \geq 1 - \frac{b^2}{n\eta^2}.$$

Hence for all $n \geq 2b^2\eta^{-2}$ we have

$$\mathbb{P}\left[X_1^1 + \cdots + X_n^1 \leq na\right] \leq \mathbb{P}\left[Y_1^1 + \cdots + Y_n^1 \leq na\right]$$

As $n_0 = 8b^2\eta^{-3}$ is bigger, (22) holds for $n \geq n_0$.

Rényi Order Implies Large Sample Order

- Case 4: $0 \le a < \mathbb{E}\left[Y^1\right] + \eta$. By Lemma 2, we have that

$$K^*_{X^1}(a) - \eta > K^*_{Y^1}(a - \eta).$$

For any random variable Z, we have
$K_{-Z}(t) = \log \mathbb{E}\left[\mathrm{e}^{t(-Z)}\right] = \log \mathbb{E}\left[\mathrm{e}^{(-t)Z}\right] = K_Z(-t)$, and
$K^*_{-Z}(a) = \sup_{t\in\mathbb{R}} t \cdot a - K_{-Z}(t) = \sup_{t\in\mathbb{R}}(-t)\cdot(-a) - K_Z(-t) = K^*_Z(-a)$.
Therefore,

$$K^*_{-X^1}(-a) - \eta > K^*_{-Y^1}(-a + \eta).$$

We can now apply Proposition 5 to the random variables $-Y^1$ and $-X^1$, and the threshold $-a > -\mathbb{E}\left[Y^1\right] - \eta > \mathbb{E}\left[-X^1\right]$. This yields

$$\mathbb{P}\left[-Y^1_1 - \cdots - Y^1_n > -na\right] \ge \mathbb{P}\left[-X^1_1 - \cdots - X^1_n > -na\right]$$

for all $n \ge 4b^2(1+\eta)\eta^{-3}$. Hence it holds for $n \ge n_0$.

Open Questions

- More than two states.
- Experiments with unbounded likelihood ratios.

Thanks!