THE VALUE AND COST OF INFORMATION

UNI BONN MINI COURSE LECTURE NOTES

OMER TAMUZ

Contents

1. Decision problems

A state of nature θ can take values in $\{0,1\}$. A decision maker has to choose an action out of a set A of actions. Her utility for taking action a when the state is θ is $u(a,\theta)$ for some $u\colon A\times\{0,1\}\to\mathbb{R}$. Suppose that the decision maker has a belief on $\{0,1\}$, assigning probability p to 1. Put together, $D=(A,u,p)$ is called a *decision problem*.

We assume that the decision maker is an expected utility maximizer. Then she would like to choose an action a that maximizes $pu(a,1)+(1-p)u(a,0)$. While in general such an action might not exist, we will only consider utility functions u for which there is such a maximizer for every p (see Exercise 1). Accordingly, define the *indirect utility* function $v\colon[0,1]\to\mathbb{R}$ by

$$v(p)=\max_{a\in A} pu(a,1)+(1-p)u(a,0).$$

Since v is the maximum of convex function (in fact, affine functions) it is convex. When A is finite, v is piecewise linear. For infinite A this is of course no longer true, but v is nevertheless always continuous (see Exercises 2 and 3).

Date: June 14, 2022.

A standard example is $A = \{0,1\}$ and $u(a,\theta) = \mathbb{1}_{\{a=\theta\}}$. In this case $v(p) = 1-p$ for $p \leq 1/2$, and $v(p) = p$ for $p \geq 1/2$. Another important example is $A = [0,1]$ and $u(a,\theta) = -(a-\theta)^2$. In this case $v(p) = 1/4 - (p-1/2)^2$.

2. Blackwell experiments

A *Blackwell experiment* is $\mu = (\Omega, \mu_0, \mu_1)$, where Ω is a set (equipped with a sigma-algebra, which we suppress) and μ_0, μ_1 are probability measures defined on it. The interpretation is that of a signal, taking values in Ω and distributed as μ_θ when the state is θ. Experiments are accordingly also called signals.

A common example is the symmetric binary experiment β for which $\Omega = \{0,1\}$, and $\beta_0(0) = \beta_1(1) = c$ for some $c \geq 1/2$. This is the signal that in each state equals the state with probability c. For example, suppose that there is an urn that has red and green balls. When $\theta = 0$ there are six red and four green balls. When $\theta = 1$ there are four red and six green balls. Then the symmetric binary experiment with $c = 6/10$ captures the physical experiment of choosing uniformly a ball from the urn.

Another example is the Gaussian experiment γ, for which $\Omega = \mathbb{R}$, $\gamma_0 = \mathrm{N}(0,\sigma^2)$ and $\gamma_1 = \mathrm{N}(1,\sigma^2)$. This can be interpreted as a measurement of the state with standard Gaussian noise.

To study decision making, it will be useful to think of a signal as a random variable s co-existing with the state in a probability space. That is, we will fix a probability space in which θ is a non-constant random variable taking values in $\{0,1\}$, s is a random variable taking values in Ω, and the distribution of s conditioned on θ is μ_θ. We will say that s realizes μ.

We will use $\mathbb{P}$ and $\mathbb{E}$ to denote probabilities and expectations in this space. This probability space can be interpreted as the subjective belief of the decision maker. As such, $p = \mathbb{P}[\theta = 1]$ is the prior belief of the decision maker that the state is 1. We will denote by $q = \mathbb{P}[\theta = 1|s]$ the posterior belief of the decision maker, after observing the signal. Note that this is a random variable, which is a function of the signal s. The martingale property of beliefs refers to the fact that the expectation of the posterior is equal to the prior: $\mathbb{E}[q] = p$. This is a consequence of the law of total expectations.

3. The value of information and the Blackwell order

Suppose that the decision maker observes a Blackwell experiment μ before choosing her action. That is, she chooses an action a^* that maximizes her expected utility conditioned on what she knows, which is a signal s that realizes μ:

$$a^* = \operatorname*{argmax}_{a \in A} \mathbb{E}[u(a,\theta)|s].$$

Recalling our notation $q = \mathbb{P}[\theta = 1|s]$, we note that given s, the choice of a^* depends only on q, so that

$$a^* = \operatorname*{argmax}_{a \in A} \mathbb{E}[u(a,\theta)|q].$$

Indeed, we can think of q itself as a signal, taking values in $[0,1]$ (see Exercise 4).

It follows that the decision maker's expected utility conditioned on q is $v(q)$, and so her unconditional expected utility is

$$\mathbb{E}\left[u(a^*,\theta)\right] = \mathbb{E}[v(q)].$$

We can think of this expected utility as the value of the experiment μ to a decision maker facing the decision problem $D = (A,u,p)$. We denote this by $V(\mu,D)$.

The Blackwell order is a notion of when one experiment contains more information than another. Given two experiment μ and μ', we say that μ *Blackwell dominates* μ' if $V(\mu,D) \geq V(\mu',D)$ for all decision problems D. This defines a partial order $\succeq$ on experiments (see Exercise 5). An equivalent formulation is that for every s and s' that realize μ and μ, and for every continuous and convex $v : [0,1] \to \mathbb{R}$, $\mathbb{E}[q] \geq \mathbb{E}\left[q'\right]$, where q and q' are the posteriors induced by s and s', respectively.

A different notion of informativeness is that of *garbling*. We say that μ' is a garbling of μ if there exist s and s' that realize Blackwell experiments μ and μ' such that there is a random variable r independent of θ and a measurable map f for which $s' = f(s,r)$. This captures the idea that it is possible to generate s' from s using no additional information about the state.

Blackwell's Theorem states that these notions are the same (see Exercise 6).

Theorem 3.1 (Blackwell [3, 4])**.** *Let μ,μ' be Blackwell experiments. Then $\mu \succeq \mu'$ if and only if μ' is a garbling of μ.*

As an example, the symmetric binary experiment β is a garbling of the Gaussian experiment γ as long as the former's precision c is small enough compared to the latter's precision $1/\sigma^2$. When $c = 1$, β dominates γ for any σ^2. But when $c < 1$, β never dominates γ (see Exercise 7).

4. Conditionally independent signals

Let $\mu = (\Omega, \mu_0, \mu_1)$ and $\eta = (\Xi, \eta_0, \eta_1)$ be Blackwell experiments. Denote the *product experiment* by

$$\mu \otimes \eta = (\Omega \times \Xi, \mu_0 \times \eta_0, \mu_1 \times \eta_1).$$

This experiment yields a signal that comprises a pair: one from Ω and one from Ξ. And these two are independent, conditioned on the state. This can be interpreted as combining two independent experiments: measuring a quantity twice with (unconditionally) independent measurement errors. If s_1 realizes μ and s_2 realized η then (s_1, s_2) realizes $\mu \times \eta$ if s_1 and s_2 are independent conditioned on θ.

We denote by $\mu^{\otimes n}$ the n-fold product of μ with itself. For example if μ is the experiment in which a ball is chosen from an urn, then $\mu^{\otimes n}$ is the experiment in which this is repeated n times, and where the ball is put back in the urn after each pick (see Exercise 8).

Bayesian updating of product experiments takes a simple, separable form. For this discussion, and henceforth, we will consider only experiments μ such that μ_0 and μ_1 are mutually absolutely continuous. That is, for any subset $A \subseteq \Omega$, $\mu_0(A) = 0$ if and only if $\mu_1(A) = 0$. This implies that the log-likelihood ratio $\ell_\mu := \log \frac{\mathrm{d}\mu_1}{\mathrm{d}\mu_0}$ exists and is finite; we write ℓ when μ can be inferred from the context. That is, there is a function $\ell_\mu \colon \Omega \to \mathbb{R}$ such that for any subset $A \subseteq \Omega$

$$\int_A \mathrm{d}\mu_1(\omega) = \int_A \mathrm{e}^{\ell(\omega)} \mathrm{d}\mu_0(\omega).$$

Suppose that $(s_1, \ldots, s_n)$ realize $\mu_1 \times \cdots \times \mu_n$. Then $\ell(s_i)$ are well-defined random variables. By Bayes' Law, $q = \mathbb{P}[\theta = 1 | s_1, \ldots, s_n]$ is given by

$$\log \frac{q}{1-q} = \log \frac{p}{1-p} + \ell(s_1) + \cdots + \ell(s_n).$$

Conditioned on θ, this is a sum of independent random variables, and so to understand it we can apply many classical tools from probability theory. In particular, if $(s_1, \ldots, s_n)$ realizes $\mu^{\otimes n}$ this is a sum of i.i.d. random variables, which are very well understood.

5. Asymptotic learning and the value of repeated experiments

Fix μ, and suppose $(s_1, s_2, \ldots)$ are a sequence of conditionally (on θ) i.i.d. random variables where each s_i realizes μ. Thus $(s_1, \ldots, s_n)$

realizes $\mu^{\otimes n}$. The expectation of $\ell(s_i)$ conditioned on $\theta = 1$ is

$$\mathbb{E}[\ell(s_i)|\theta=0] = \int \log \frac{\mathrm{d}\mu_1}{\mathrm{d}\mu_0}(\omega)\,\mathrm{d}\mu_0(\omega) = -\int \log \frac{\mathrm{d}\mu_0}{\mathrm{d}\mu_1}(\omega)\,\mathrm{d}\mu_0(\omega) = -D_{\mathrm{KL}}(\mu_0||\mu_1),$$

where $D_{\mathrm{KL}}(\cdot||\cdot)$ denotes the Kullback-Leibler divergence, which is positive unless the two measures are equal (see Exercise 9). Denote $q_n = \mathbb{P}[\theta = 1|s_1,\ldots,s_n]$. Then by the strong law of large numbers it almost surely holds that conditioned on $\theta = 0$

$$\lim_n \frac{1}{n} \log \frac{q_n}{1-q_n} = -D_{\mathrm{KL}}(\mu_0||\mu_1).$$

In particular, $\lim_n q_n = 0$.

The theory of large deviations can give us more nuanced estimates for the likelihood of *not* learning θ at time n. Let X be a random variable. The *cumulant generating function* $K_X : \mathbb{R} \to \mathbb{R}$ of X is

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

We assume that it is finite. This is a smooth convex function. It is equal to 0 at 0 and its derivative at 0 is equal to the expectation of X. Consider its Fenchel transform

$$K_X^\star(a) = \max_{t\geq 0}(ta - K_X(t)).$$

Since K_X is smooth, if $K_X^\star(a) = t$ then $a = K_X'(t)$. That is, $K_X^\star$ is the inverse of K_X'.

Consider a sequence of i.i.d. random variables $X_1, X_2, \ldots$ distributed like X.

Theorem 5.1 (Cramér [5])**.** *Suppose $a \geq \mathbb{E}[X]$. Then*

$$\mathbb{P}[X_1 + \cdots + X_n \geq na] = \exp\left(-nK_X^\star(a) + o(n)\right).$$

It follows that if $\mathbb{E}[X] \leq 0$ and if $\mathbb{P}[X > 0] > 0$ then for any constant $c \in \mathbb{R}$

$$\mathbb{P}[X_1 + \cdots + X_n \geq c] = \exp\left(-n\rho_X + o(n)\right), \tag{5.1}$$

where $\rho_X = \min_t K_X(t)$ (see Exercise 10).

Consider a decision problem with finitely many actions, and suppose that there is a unique action a_0 that is optimal when $\theta = 0$: $u(a,0) \geq u(a',0)$ for any $a' \in A$. Likewise, there is a unique a_1 that is optimal when $\theta = 1$. Then there is some threshold r_0 such that a decision maker will choose a_0 whenever their posterior is below r_0. Equivalently, there is some $c_0 \in \mathbb{R}$ such that a_0 is chosen whenever $\log \frac{q}{1-q} \leq c_0$.

Let

$$K^0_\mu(t) = \log \mathbb{E}\left[\mathrm{e}^{t\ell(s_1)} | \theta = 0\right] = \log \int_\Omega \mathrm{e}^{t \log \frac{\mathrm{d}\mu_1}{\mathrm{d}\mu_0}(\omega)}\,\mathrm{d}\mu_0(\omega)$$

be the cumulant generating function of the log-likelihood ratio, conditioned on $\theta = 0$. As observed by Moscarini and Smith [6], it is implied by (5.1) that

$$\mathbb{P}[q_n \le r_0 | \theta = 0] = \exp\left(-n\rho^0_\mu + o(n)\right),$$

where $\rho^0_\mu = \min_t K^0_\mu(t)$. Now (see Exercise 11),

$$K^1_\mu(t) = \log \mathbb{E}\left[\mathrm{e}^{-t\ell(s_1)} | \theta = 1\right] = K^0_\mu(1-t), \tag{5.2}$$

and so $\rho^1_\mu = \rho^0_\mu$. Thus the probability that a decision maker chooses a wrong action decays exponentially, and does so at the same rate in both states. This is independent of the choice of decision problem and prior, and depends only on the experiment μ. We thus denote it by ρ_μ. Another way of thinking of ρ is that if $\rho_\mu > \rho_\eta$ then for every decision problem D there is some N such that $V(\mu^{\otimes n}, D) > V(\eta^{\otimes n}, D)$ for all $n \ge N$. It follows that the order induced by ρ is a refinement of the Blackwell order: $\mu \succeq \eta$ implies that $\rho_\mu \ge \rho_\eta$.

6. Rényi divergences and dominance in large samples

We say that an experiment is bounded if the log-likelihood ratio that it induces is bounded. In this section we restrict ourselves to bounded experiments.

Blackwell [3] asked whether there exist experiments μ, η such that $\mu \not\succeq \eta$ but $\mu \times \mu \succeq \eta \times \eta$. This was answered positively by Stein [9]; see also Torgersen [10] and Azrieli [2]. More generally, one could ask: under which conditions on μ and η does it holds that $\mu^{\otimes n} \succeq \mu^{\otimes n}$ for all n large enough? In this case we say that μ dominates η in large samples.

It turns out that the answer is closely related to the discussion above. Recall that

$$K^0_\mu(t) = \log \mathbb{E}\left[\mathrm{e}^{t\ell(s_1)} | \theta = 0\right] = \log \int_\Omega \mathrm{e}^{t \log \frac{\mathrm{d}\mu_1}{\mathrm{d}\mu_0}(\omega)}\,\mathrm{d}\mu_0(\omega).$$

We can also write this as

$$K^0_\mu(t) = \log \int_\Omega \left(\frac{\mathrm{d}\mu_1}{\mathrm{d}\mu_0}(\omega)\right)^t \mathrm{d}\mu_0(\omega).$$

We reparametrize $t \mapsto 1-t$ and normalize to arrive at the definition of *Rényi divergence*:

$$R^0_\mu(t) = \frac{1}{t-1} K^0_\mu(1-t) = \frac{1}{t-1} \log \int_\Omega \left(\frac{\mathrm{d}\mu_0}{\mathrm{d}\mu_1}(\omega) \right)^{t-1} \mathrm{d}\mu_0(\omega),$$

with $R^1_\mu(t)$ defined analogously. Thus R^θ_μ is a family of divergences, indexed by $t \geq 0$ (and $\theta \in \{0,1\}$), which naturally generalizes Kullback-Leibler divergence. Indeed, $R^0_\mu(1) = D_{\mathrm{KL}}(\mu_0||\mu_1)$. We can also extend, by continuity, to $t = \infty$. Then $R^0_\mu(\infty)$ is the maximum of the support of the log-likelihood ratio $\log \frac{\mathrm{d}\mu_1}{\mathrm{d}\mu_0}$.

Like the cumulant generating function, Rényi divergences are additive for product experiments: $R^\theta_{\mu\otimes\eta} = R^\theta_\mu + R^\theta_\eta$. They are also monotone in the Blackwell order: $\mu \succeq \eta$ implies $R^\theta_\mu(t) \geq R^\theta_\nu(t)$ for all t (see Exercise 12). Interestingly, they determines the large sample order.

Theorem 6.1 (Mu et al. [7])**.** *Suppose $R^\theta_\mu(t) > R^\theta_\eta(t)$ for all $t \in (0,\infty]$ and $\theta \in \{0,1\}$. Then μ dominates η in large samples.*

The proof of this theorem is related to the following result. Recall that a random variable X *first order stochastically dominates* Y if for every $a \in \mathbb{R}$ it holds that $\mathbb{P}[X \leq a] \leq \mathbb{P}[Y \leq a]$.

Theorem 6.2 (Aubrun and Nechita [1])**.** *Let X, Y be bounded random variables. If $\frac{1}{t}K_X(t) > \frac{1}{t}K_Y(t)$ for all $t \in [-\infty,\infty]$ then for all n large enough a sum of n independent copies of X first-order stochastically dominates the sum of n independent copies of Y.*

The relation between first order stochastic dominance and Blackwell dominance is given by the following result.

Proposition 6.3 (Mu et al. [7])**.** *Suppose s_1, s_2 realize μ, η respectively. Let R be an independent r.v. distributed exponentially (i.e., with density $f(x) = \mathrm{e}^{-x}$ for $x \geq 0$). Then $\mu \succeq \eta$ if and only if, conditioned on $\theta = 1$, $\ell(s_1) - R$ first order stochastically dominates $\ell(s_2) - R$.*

7. The cost of information

Suppose that a firm produces bounded Blackwell experiments. A historically standard assumption for the cost of production is that it is additive for product experiments. For example, Wald [11] assumes that there is only one experiment μ that can be produced, and that the cost of producing $\mu^{\otimes n}$ is linear in n. Wilson [12] and many others assume that only Gaussian experiments can be produced, and that the cost is linear in the precision $1/\sigma^2$ (see Exercise 13); this implies that if γ is Gaussian then the cost of $\gamma^{\otimes n}$ is again linear in n.

Suppose that we allow for more flexible information acquisition. For example, suppose a firm can produce any bounded experiment. We would like to know how to think of a cost function $C\colon \mathscr{B} \to \mathbb{R}_+$, where $\mathscr{B}$ is the set of bounded experiments. A natural assumption is that C is Blackwell monotone: $\mu \succeq \eta$ implies that $C(\mu) \geq C(\eta)$. This is a "free disposal" assumption, since by garbling it is easy to turn a μ experiment into an η experiment by forgetting some information. If in addition we also assume that C is additive—i.e., $C(\mu \otimes \eta) = C(\mu) + C(\eta)$, what can we conclude? We have already seen that Rényi diverges are additive and monotone. Is there anything else?

The set of monotone additive cost functions is easily shown to be a convex cone. It can furthermore be shown to have a set of extreme rays that is compact (under pointwise convergence). Thus to understand it we must understand these rays.

Theorem 7.1 (Mu et al. [7]). *The extreme rays of the set of additive monotone cost functions are the Rényi divergences.*

In other words, every monotone additive cost function is a weighted sum (integral) of Rényi divergences.

An additional axiom was introduced by Pomatto et al. [8]. Given an experiment $\mu = (\Omega, \mu_0, \mu_1)$, and $\alpha \in (0,1)$, they consider the "diluted" experiment $\alpha \cdot \mu = (\Omega \cup \{e\}, \mu_0', \mu_1')$, where $\mu_\theta' = (1-\alpha)\delta_e + \alpha\mu_\theta$. This corresponds to an experiment that fails with probability one half, producing an error message e and no information, and succeeds and produces a μ signal otherwise. Their dilution axiom states that $C(\alpha \cdot \mu) = \alpha C(\mu)$.

Theorem 7.2 (Pomatto et al. [8]). *Every monotone additive cost function that satisfies the dilution axiom is of the form $C(\mu) = \beta_1 D_{\mathrm{KL}}(\mu_1||\mu_0) + \beta_0 D_{\mathrm{KL}}(\mu_0||\mu_1)$.*

This follows from Theorem 7.1, since none of the Rényi divergences satisfy the dilution axiom, except the Kullback-Leibler divergences. To see this, note that by the axioms,

$$C(\frac{1}{n} \cdot \mu^{\otimes n}) = C(\mu).$$

Now, if X is a random variable, and if Z_n is equal to X with probability $1/n$ and to 0 otherwise, then

$$K_{Z_n}(t) = \log \mathbb{E}\left[\mathrm{e}^{tZ}\right] = \log\left(\frac{1}{n}\mathbb{E}\left[\mathrm{e}^{tX}\right] + \frac{n-1}{n}\right) = \log\frac{n-1}{n} + \log\frac{\mathbb{E}\left[\mathrm{e}^{tX}\right]}{n-1}$$

and so

$$\lim_n K_{Z_n}(t) = 0$$

for all t. Since all Rényi divergences (except $R_\mu^0(1)$ and $R_\mu^1(1)$) are of this form, they do not satisfy the dilution axiom. An additional argument shows that this is still impossible for their convex combinations.

Appendix A. Exercises

(1) Let A be a set, and consider a map $u\colon A\times\{0,1\}\to\mathbb{R}$. Show that if the set $\{(u(a,1),u(a,0) : a\in A\}\subseteq\mathbb{R}^2$ is compact then for every $p\in[0,1]$ there exists $a\in A$ that maximizes $pu(a,1)+(1-p)u(a,0)$. Give an example showing that the converse is not true.

(2) Fix some $v\colon[0,1]\to\mathbb{R}$. Show that v is the indirect utility of some u if and only if v is continuous and convex.

(3) Suppose that if $a\neq a'$ then either $u(a,0)\neq u(a',0)$ or $u(a,1)\neq u(a,1)$. Show that there is a unique $a\in A$ such that $v(p)=pu(a,1)+(1-p)u(a,0)$ if and only if v is differentiable at p.

(4) Suppose that s is a signal and that $q=\mathbb{P}[\theta=1|s]$ is the induced posterior.
   (a) Show that $\mathbb{E}[q]=\mathbb{P}[\theta=1]$.
   (b) Show that $q=\mathbb{P}[\theta=1|q]$.
   (c) Denote by $\tilde{\mu}$ the distribution of q and by $\tilde{\mu}_0$ and $\tilde{\mu}_1$ its distributions conditional on $\theta=0$ and $\theta=1$ respectively. Show that

$$\frac{\mathrm{d}\tilde{\mu}_0}{\mathrm{d}\tilde{\mu}}(x)=\frac{x}{p}\quad\text{and}\quad\frac{\mathrm{d}\tilde{\mu}_1}{\mathrm{d}\tilde{\mu}}(x)=\frac{1-x}{1-p},$$

   where $p=\mathbb{P}[\theta=1]$ is the prior.

(5) Show that the Blackwell order $\succeq$ on experiments is a partial order, and that it is not a complete order.

(6) Show that if μ' is a garbling of μ then $\mu\succeq\mu'$.

(7) Calculate for which values of c and σ^2 the symmetric binary experiment with precision c is a garbling of the Gaussian experiment with precision $1/\sigma^2$. Show that a symmetric binary experiment with $c<1$ does not dominate any Gaussian experiment.

(8) Show that if $\mu_1\succeq\mu_2$ then for any η, $\mu_1\otimes\eta\succeq\mu_2\otimes\eta$. Conclude that if $\mu\succeq\eta$ then $\mu^{\otimes n}\succeq\eta^{\otimes n}$ for all $n\geq1$.

(9) Suppose that μ_1 and μ_0 are mutually absolutely continuous. Show that $D_{\mathrm{KL}}(\mu_1||\mu_0)\geq0$ with equality if and only if $\mu_1=\mu_0$.

(10) Prove (5.1).

(11) Prove (5.2). Conclude that $K^0_\mu(1)=K^1_\mu(1)=0$, but explain why this follows directly from Bayes' Law.

(12) Show that if $\mu\succeq\nu$ then $R^\theta_\mu(t)\geq R^\theta_\eta(t)$ for all $t\geq0$ and $\theta\in\{0,1\}$.

(13) Show that if γ,γ' are Gaussian experiments with precisions $1/\sigma^2,1/\tau^2$ then $\gamma\otimes\gamma'$ is Blackwell equivalent to a Gaussian experiment with precision $1/\sigma^2+1/\tau^2$.

REFERENCES

[1] Guillaume Aubrun and Ion Nechita, *Catalytic majorization and ℓ_p norms*, Communications in Mathematical Physics **278** (2008), no. 1, 133–144.

[2] Yaron Azrieli, *Comment on "the law of large demand for information"*, Econometrica **82** (2014), no. 1, 415–423.

[3] David Blackwell, *Comparison of experiments*, Proceedings of the second berkeley symposium on mathematical statistics and probability, 1951, pp. 93–102.

[4] ______, *Equivalent comparisons of experiments*, The Annals of Mathematical Statistics **24** (1953), no. 2, 265–272.

[5] Harald Cramér, *Sur un nouveau théoreme-limite de la théorie des probabilités*, Actual. Sci. Ind. **736** (1938), 5–23.

[6] Giuseppe Moscarini and Lones Smith, *The law of large demand for information*, Econometrica **70** (2002), no. 6, 2351–2366.

[7] Xiaosheng Mu, Luciano Pomatto, Philipp Strack, and Omer Tamuz, *From blackwell dominance in large samples to rényi divergences and back again*, 2020. Forthcoming in Econometrica.

[8] Luciano Pomatto, Philipp Strack, and Omer Tamuz, *The cost of information*, arXiv preprint arXiv:1812.04211 (2018).

[9] Charles Stein, *Notes on the comparison of experiments*, University of Chicago (1951).

[10] Erik Nikolai Torgersen, *Comparison of experiments when the parameter space is finite*, Probability Theory and Related Fields **16** (1970), no. 3, 219–249.

[11] Abraham Wald, *Sequential tests of statistical hypotheses*, The Annals of Mathematical Statistics **16** (1945), no. 2, 117–186.

[12] Robert Wilson, *Informational economies of scale*, The Bell Journal of Economics **6** (1975), no. 1, 184–195.