CONSUMER THEORY FOR CHEAP INFORMATION

Gary Baker

This version: 28 September 2020

[PRELIMINARY AND INCOMPLETE – PLEASE DO NOT CIRCULATE]

LINK TO THE MOST RECENT VERSION

ABSTRACT

In many natural scenarios, a decision-maker facing uncertainty must decide not only *how much* information to purchase, but also *from which* sources. Unfortunately, an understanding of the value of information for general information structures is notoriously elusive. I characterize tradeoffs between samples from different sources in a setting with large information purchases where the probability of a mistake is small and well-described by large deviations theory. More specifically, in environments with finitely many possible underlying states, I provide an approximation for the marginal rate of substitution for samples from distinct information sources, valid when samples are sufficiently cheap (or budgets sufficiently large). I then show marginal rate of substitution is given by a ratio of precision-like indices that depend only on properties of each information source and the relative proportions of each signal in the bundle. This formula naturally implies a particularly accurate approximation for information demand in constrained settings. Furthermore, because the precision of each signal does not depend on decision-maker characteristics, all decision-makers—independent of prior and payoff structure—agree on the relative proportions of each source in an asymptotically optimal bundle. Of particular note: in environments with more than two possible states, interior solutions arise when the signals differ in which pairs of states they struggle to distinguish most. In this case, the asymptotically optimal bundle either occurs at a corner or at one of a finite number of interior kink points where the worst-case pair of states switches. To illustrate these results, I consider a number of basic consumer theory exercises and discuss implications for information demand.

1 INTRODUCTION

Often a decision-maker may find herself in a position to acquire information prior to making a decision under an uncertain state of the world, and many cases, she must not only decide "how much" information to purchase, but also "from where" to acquire it. For example, a reader of the news must every day decide both how long and from which sites to read. Or a researcher studying the effects of a new drug on a disease might have multiple available tests of differing cost for the disease.

In these scenarios, how might one determine the optimal bundle of information? Are corner solutions always optimal? And if not, when?

Answering these questions ought to be little more than a principles exercise of setting marginal rate of substitution (MRS) equal to price ratio along a budget constraint. Unfortunately, such an exercise presupposes understanding of a utility function—no mean feat when the value for information is notoriously ill-behaved.[1]

I provide an answer to these questions for environments with finitely many possible underlying states, valid when information is sufficiently cheap (or budgets sufficiently large), by precisely characterizing asymptotic tradeoffs between conditionally independent samples from different information sources (Blackwell experiments) using methods from the theory of large deviations.

In particular, I develop a precision-like measure (the Chernoff precision) that quantifies the degree to which an information source can distinguish between a given pair of underlying states. I then show that, at large samples, indifference curves[2] are loci of sample bundles with equal precision of their worst-case dichotomy (the state pair for which the bundle is least precise). Furthermore, for any given state pair, the precision of a sample bundle is the sum of component precisions from each constituent information source, and the component precision of a given source is always weakly lower when an information source is a part of a heterogeneous bundle than it is by itself. Put more simply: for a single pair of states, composite information sources are weakly worse that the sum of their parts.

In a two-state environment, these results imply indifference curves eventually bow outward, and that corner solutions are thus optimal at low costs. In richer environments, however, because only the least-precision dichotomy matters at high samples, a source may complement another by covering for the other's weaknesses. In this case, indifference curves are outer contours of multiple outward bowing curves (one for each dichotomy), and thus inward-pointing kinks occur whenever the worst-case pair

1 Most famously, value of information is often non-concave, and thus marginal value of information is often rising. See, for example (Radner & Stiglitz 1984) or Chade & Schlee (2002).

2 Samples are a fundamentally discrete quantity, so non-singleton indifference sets typically don't exist. I will be more formal later.

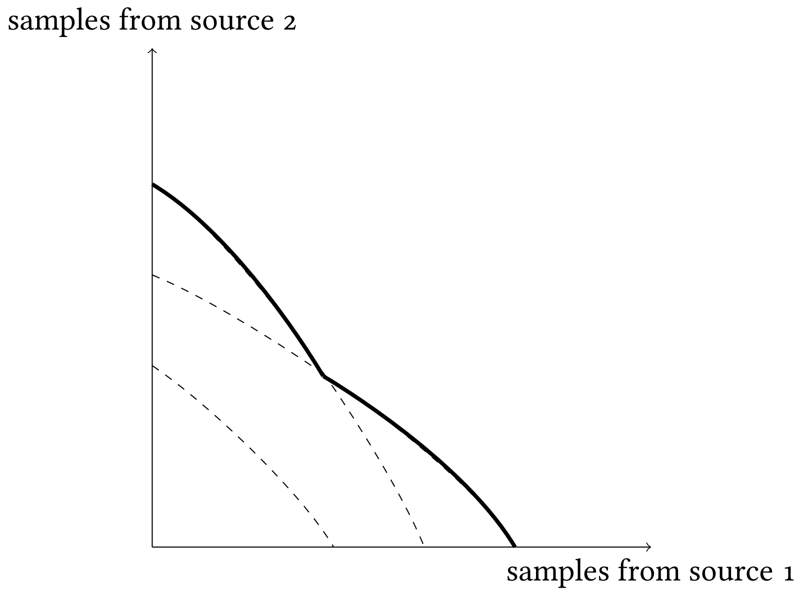

Figure 1: Illustration of iso-precision lines in a three state environment with two available information sources. Each dashed line represents the a locus of bundles with equal precision for a fixed dichotomy. The outer contour of these curves then represents the locus of bundles with equal *lowest* precision, and thus the approximate indifference curve. In this example, one of the dichotomies always has higher precision than the other two, and thus there is only a single interior kink point.

of states changes. The optimal bundle must then eventually lie near either a corner, or one of finitely many interior kink points. Fig. 1 stylistically illustrates this result graphically for an example environment with three states (and thus three dichotomies).

I then apply my formulae to derive asymptotically precise expressions for information demand given a budget constraint and cost vector. These results have certain testable consumer theoretic implications. For example, because interior solutions always lie near an indifference curve kink, for small price changes, information sources behave as perfect Marshallian complements with fixed bundle proportions.

In elasticity terms, except near finitely many discontinuities when the optimum jumps between kinks/corners, demand curves are inelastic, with finitely many small cost intervals of high elasticity when jumping between kinks/corners.

Furthermore, component precisions depend only on the relative proportions of each source in a bundle, so information value is approximately homothetic at large samples. Thus all information sources are normal goods—and more specifically, they have unit income elasticities—for low enough prices or high enough budgets.

Importantly, because the Chernoff precision only depends on characteristics of the information sources, and not on decision-maker specifics such as prior or payoff structure, (generically) all decision makers agree on tradeoffs between information sources at low enough price, and for a given price vector, all agree of the relative proportions of each information source in an asymptotically optimal bundle.

My results fit primarily into two different strains of the literature: one economic, and the other statistical.

First, this paper contributes to the statistical literature on asymptotic relative efficiency (ARE) which traditionally asks how many samples, n_1, from one test are required to perform as well as n_2 from another at large samples. In particular, my results build on those of Chernoff (1952) who showed that, in a two-state/two-action environment (e.g. null/alternative hypothesis testing), the ratio, n_2/n_1, of two equally performing sample sizes is given by the ratio of each test's respective Chernoff precision β_1/β_2. My results expand upon this in two ways: first, I generalize the problem to tradeoffs between arbitrary finite-action/finite-state decision scenarios, and second, I consider local tradeoffs between interior bundles, rather than just the extreme tradeoffs between corner bundles.

Second, this paper sits in the broader information economics literature, especially that on information demand. Of particular relevance—and crucial for my results—are the results of Moscarini & Smith (2002), henceforth MS02, who characterized the asymptotic marginal value of information using methods from the theory of large deviations to show that, in a two-state world, the probability of a mistake is falling exponentially, and that in a many-state environment, the only most likely mistake (that is, the state pair with the least precision) matters at large samples. They then use these results to calculate a particularly accurate approximation for information demand at low prices. Their demand approximation implies demand is always inelastic and thus linear pricing cannot be optimal for an information monopolist. I generalize these results to a world with multiple available information sources and show there is a rich consumer theory arising from the exercise.

Finally, the Chernoff precision fits in a smaller literature (somewhat adjacent to the information design literature) on quantifying the information content of signals. Pomatto et al. (2018) characterize a unique class of "linear" information measures based on weighted sums of the information source's various relative entropies of its conditional distributions. The linear nature of their measure lends itself to capturing costs of information *production*. By contrast the Chernoff precision is more useful on the *consumption* side of the equation, characterizing approximate tradeoffs between different information sources.

These results lend themselves to multiple applications. First and foremost, they provide a general approach for analysis of information markets. In particular, they imply that information can be treated with a fairly standard consumer theory toolkit, and suggest particular functional forms for information values and demand that may prove useful for modeling purposes in certain applied settings.

Additionally, this paper provides a prescriptive recommendation for optimal experiment design that accounts not only for statistical properties of the various design options, but also their costs. Although the finite-state assumption limits the immediate applicability of my results in most experimental contexts, the approach provides a basis for developing an economically-motivated design criterion.[3]

The remainder of the paper is structured as follows: Section 2 explains the model structure and assumptions. Section 3 introduces the relevant large deviations tools and develops the main approximation results, and Section 4 uses these results to build a consumer theory for information, and considers some implications for information demand. Section 5 explores the performance of the approximations both theoretically and numerically. Finally, Section 6 concludes and discusses areas for future research.

2 MODEL

A decision-maker (DM) must choose an action a from a finite set, A, under an uncertain decision-relevant state of the world θ drawn from a finite set, Θ. The DM has a full-support prior, $p \in \Delta\Theta$, and state-dependent (Bernoulli) utility function, $u(a, \theta)$. To avoid trivialities, assume no action is dominated, and for expositional simplicity assume that each state has a unique optimal action (so necessarily $|\mathrm{A}| \geq |\Theta|$).

Denote a particular information source—formally, a *Blackwell experiment*—as $\mathscr{E} = \langle \mathrm{F}(\cdot \mid \theta), \theta \in \Theta \rangle$. That is, each signal is a collection of state-dependent distributions over some arbitrary space of outcomes (or *realizations*), R. Purely for the sake of simplifying exposition, I will assume $\mathrm{R} = \mathbb{R}$ and that each state-dependent distribution has a density, $f(\cdot \mid \theta)$, but the results and proofs allow arbitrary realization space.

To avoid degenerate outcomes, assume all experiments are composed of distinct distributions—so signals are informative—and that those distributions are mutually absolutely continuous—so no signal realization completely rules in or rules out any strict subset of the states. Under the density assumption, this implies that for any state pair, $\theta\theta'$, and for all realizations (except perhaps a measure zero set) the likelihood ratio, $f(r \mid \theta')/f(r \mid \theta)$, is non-zero and finite.

The DM has access to a finite menu of such information sources, $\mathscr{E}_1, \dots, \mathscr{E}_\mathrm{J}$, from each of which she may purchase an arbitrary number of conditionally independent samples, $\mathrm{n} = [n_1, \dots, n_\mathrm{J}]$ at costs $\mathrm{c} = [c_1, \dots, c_\mathrm{J}]$ per sample. The DM then observes a realization from the joint signal $\bigotimes_{j=1}^{\mathrm{J}} \mathscr{E}_j^{n_j}$, where $\mathscr{E}^n$ denotes the compound signal

3 Experiment design is a large and varied literature within statistics going back to Fisher (and arguably Gauss). For a textbook treatment, see Montgomery (2012). I consider the experiment design implications more carefully in a separate working paper that generalizes these results to certain regression settings.

generated by n i.i.d. samples from $\mathscr{E}$, and $\mathscr{E}_1 \otimes \mathscr{E}_2$ is the experiment generated by one sample from each $\mathscr{E}_1$ and $\mathscr{E}_2$. The DM has a large but finite budget, Y, to spend on information.

After observing the realization, r, of the chosen signal bundle, the DM updates her beliefs according to Bayes's rule and the chooses an action to maximize her expected payoff. Putting this all together, we can define the value with information: the expected payoffs from acting after observing the signal realization.[4] That is,

$$v(\mathrm{n}) = \int_{r \in \mathrm{R}} \max_a \left\{ \sum_{\theta \in \Theta} u(a, \theta) \mathbb{P}(\theta \mid r) f_{\mathrm{n}}(r \mid \theta) \right\} \mathrm{d}r$$

where f_{n} is the conditional density of signal realizations from the chosen sample bundle.

The DM chooses n to maximize this value, subject to the usual budget constraint, $\mathrm{n} \cdot \mathrm{c} \leq \mathrm{Y}$.

Bayes's rule for probabilities is somewhat tricky to work with so we'll typically work with log-likelihood ratios (LLRs) where Bayes's rule becomes a sum:

$$\log\left(\frac{\mathbb{P}(\theta' \mid r)}{\mathbb{P}(\theta \mid r)}\right) = \log\left(\frac{p_{\theta'}}{p_\theta}\right) + \log\left(\frac{f(r \mid \theta')}{f(r \mid \theta)}\right)$$

We'll thus need some assumptions on the state-dependent distributions of LLRs for each signal:

First, for each signal, $\mathscr{E}_j$, for any pair of states, θ and θ', denote the moment generating function (MGF) of the LLR distribution in the denominator state—that is, the distribution of $\log(f_j(r \mid \theta')/f_j(r \mid \theta))$ in state θ—as

$$\begin{aligned} \mathrm{M}_i(t\,; \theta\theta') &\equiv \int_{r \in \mathrm{R}} e^{t \log\left(\frac{f_j(r \mid \theta')}{f_j(r \mid \theta)}\right)} f_j(r \mid \theta) \mathrm{d}r \\ &= \int_{r \in \mathrm{R}} f_j(r \mid \theta)^t f_j(r \mid \theta')^{1-t} \mathrm{d}r \end{aligned} \tag{1}$$

Assume that for any pair of states, this MGF is defined within an open interval containing zero, and thus that the LLR distribution has all its moments.[5] Finally, to simplify

4 Not to be confused with the value *of* information, which measures the expected *gain* in payoff from acting after observing the realization. Both values are ordinally equivalent (they differ only by a constant), so I use the value with information to simplify notation.

5 applying dominated convergence to Eq. 1 shows that the MGF is finite for $t \in [0, 1]$ for mutually absolutely continuous distributions.

the main proof, assume that for any pair of states and for any experiment, that the log-likelihood ratios of possible signal realizations do not form a lattice on the reals—that is, for $b, k \in \mathbb{R}$

$$\left\{\log\left(\frac{\mathrm{dF}_j(r\,|\,\theta')}{\mathrm{dF}_j(r\,|\,\theta)}\right) \,:\, r \in \mathrm{R}\right\} \not\subseteq \{b + kc \,:\, k \in \mathbb{Z}\}$$

In other words, the support of the log-likelihood ratio distribution does not exclusively consist of equally spaced points.

3 RESULTS

3.1 Background: large deviations

For the sake of illustrating the relevant tools, first consider the two-state/two-action world. I pose the problem as a classic statistical dichotomy: there are two states, $\Theta = \{\mathrm{H}_0, \mathrm{H}_1\}$, representing a null hypothesis and alternative hypothesis, respectively. The DM must choose whether to *reject*, $a = \mathscr{R}$, or *accept*, $a = \mathscr{A}$, the null hypothesis. Let $u(\mathscr{A}, \mathrm{H}_0) > u(\mathscr{R}, \mathrm{H}_0)$ and $u(\mathscr{R}, \mathrm{H}_1) > u(\mathscr{A}, \mathrm{H}_1)$, so rejecting the null is preferred when the alternative is true. Denote the prior probability that the alternative is true as p. Finally denote the cut-off belief—i.e. the one where the DM is exactly indifferent between the two actions—as $\bar{p}$.

For now, in order to introduce the necessary technical concepts, I review the single-source case examined by MS02. In this case, we can write the value after receiving n samples as

$$\begin{aligned} v(n) = (1-p)\big(&\alpha_{\mathrm{I}}(n)\, u(\mathscr{R}, \mathrm{H}_0) + (1-\alpha_{\mathrm{I}}(n))\, u(\mathscr{A}, \mathrm{H}_0)\big) \\ -p\ \big(&\alpha_{\mathrm{II}}(n)\, u(\mathscr{A}, \mathrm{H}_1) + (1-\alpha_{\mathrm{II}}(n))\, u(\mathscr{R}, \mathrm{H}_1)\big) \end{aligned} \tag{2}$$

where $\alpha_{\mathrm{I}}(n)$ and $\alpha_{\mathrm{II}}(n)$ denote the respective Type-I and Type-II error probabilities with the given number of samples from each experiment.

For the sake of algebraic simplicity, it will usually be convenient to work with the *full-information gap* (FIG): the difference between the value of a perfect information source, $\bar{v} = pu(\mathscr{R}, \mathrm{H}_1) + (1-p)u(\mathscr{A}, \mathrm{H}_0)$, and the value of the given information purchases.

$$\begin{aligned} \mathrm{FIG}(n) = (1-p)\,&\alpha_{\mathrm{I}}(n))\,(u(\mathscr{A}, \mathrm{H}_0) - u(\mathscr{R}, \mathrm{H}_0)) \\ &+p\,\alpha_{\mathrm{II}}(n))\,(u(\mathscr{R}, \mathrm{H}_1) - u(\mathscr{A}, \mathrm{H}_1)) \end{aligned} \tag{3}$$

Notice that the FIG goes to zero with the error probabilities, and that constrained *minimization* of the FIG is equivalent to constrained *maximization* of the value. Equally, the FIG can be seen as the expected loss for the loss function $\mathrm{L}(a, \theta) = u(a^*(\theta), \theta) - u(a, \theta)$.

We can now characterize the asymptotic behavior of α_{I} and α_{II} using a large-deviations approach:

First, as previously discussed, Bayes's rule is a sum for log-likelihood ratios, so we can write

$$\alpha_{\mathrm{I}}(n) = \mathbb{P}\left(l + \sum_{k=1}^{n} s_k > \bar{l} \,\middle|\, \theta = \mathrm{H}_0\right)$$

where $l \equiv \log(p/(1-p))$ is the prior log-likelihood ratio, $\bar{l} \equiv \log(\bar{p}/(1-\bar{p}))$ is the indifference belief log-likelihood ratio, and $s_k \equiv \log(\mathrm{dF}(r_k \,|\, \mathrm{H}_0)/\mathrm{dF}(r_k \,|\, \mathrm{H}_1))$ is the log-likelihood ratio of the realization from the k-th sample.

Because $\mathbb{E}(s_k \,|\, \theta = \mathrm{H}_0) < 0$, the posterior log-likelihood will stochastically drift towards negative infinity (certainty that the truth is H_0). Asymptotic approximation of the mistake probability thus falls into the realm of *large deviations theory* which considers the distribution of sample means far from the true mean.[6]

To attack such a problem, Cramér (1938) canonically proved that for a random variable X with negative expectation, the probability that the sum of n i.i.d. draws from X is positive is falling roughly exponentially with rate given by the minimized value of the MGF of X. Chernoff (1952) further developed this result to show that the ratio of two equally performing sample sizes from two statistical ratio tests on a binary state is given by the ratio of log minimized MGFs.[7]

Motivated by this approach, define $\rho \equiv \min_{t \in [0,1]} \mathrm{M}(t\,; \mathrm{H}_1, \mathrm{H}_0)$ as the *Chernoff number*[8] of the information source and τ as the minimizer. Because MGFs are always convex, this minimum must be unique, and, because the MGF is 1 when $t \in \{0, 1\}$ (see Eq. 1), we must have $\tau \in (0, 1)$.

Note that $\mathrm{M}(t\,; \mathrm{H}_1, \mathrm{H}_0) = \mathrm{M}(1 - t\,; \mathrm{H}_0, \mathrm{H}_1)$ so the minimized value is the same for both log-likelihood ratios and thus each test has a single Chernoff number in the two-state world. In the two-state world, I thus suppress the state-pair argument: $\mathrm{M}(t) \equiv \mathrm{M}(t; \mathrm{H}_1, \mathrm{H}_0)$.

6 In contrast to *small deviations* central limit theorems which consider the distribution of a sample mean near the true mean.

7 I grossly simplify Cramér's and Chernoff's by only considering implications for log-likelihood ratio tests. Both results are considerably more general than this. For a more thorough coverage of large deviations methods, see Dembo & Zeitouni (1998).

8 I follow Torgersen (1991) here. Moscarini & Smith (2002) call this the *efficiency index*.

The Chernoff number arises from a fairly abstract derivation so it's worth examining a transformation of it for the sake of intuition. We can consider $\beta \equiv -\log(\rho)$ a measure of a test's *precision*. For example, consider a (homoskedastic) Gaussian signal that is distributed $\mathcal{N}(0, 1/\gamma)$ when $\theta = \mathrm{H}_0$ and $\mathcal{N}(1, 1/\gamma)$ otherwise. In this case, γ is the signal's precision in the classical sense, and direct computation reveals that for this experiment we have $\beta = (1/8)\gamma$. In reference to this example, I will call β_j the *Chernoff precision* of the test. The following result establishes a number of other properties one might expect of a measure with such a name:

Proposition 1 (Properties of Chernoff precisions). *In an environment with two states, for any information source with mutually absolutely continuous state-dependent distributions, the Chernoff precision has the following properties:*

1. $\rho \in (0, 1)$, *and thus* $\beta > 0$*;*
2. *A test composed of* k *i.i.d. samples from* $\mathcal{E}$ *will have Chernoff precision* $k\beta$*; and,*
3. *If* $\mathcal{E}_1 \succsim \mathcal{E}_2$ *in the Blackwell sense, then* $\beta_1 \geq \beta_2$.

Proof. From Eq. 1, it is immediate that the MGF takes a value of 1 at the corners of the unit interval, and is strictly between 0 and 1 on the interior. The first property then follows. The second follows because the log-likelihood ratio MGF of k samples is M^k because the MGF of an i.i.d. sum is the product of MGFs. The last property follows from (Thm. 12, Blackwell 1951) which states that composites of Blackwell-dominant experiments dominate composites of dominated ones. Thus, Chernoff's result implies that dominant experiments must have lower Chernoff numbers and thus higher Chernoff precisions. $\square$

With these tools we can now state a version of MS02's main result:

Lemma 1 (MS02, Thm. 1). *The probability of a type-I error is proportional to*[9]

$$\alpha_{\mathrm{I}}(n) \propto \frac{\rho^n}{\sqrt{n}}\left(1 + \mathcal{O}\left(\frac{1}{\sqrt{n}}\right)\right) \tag{4}$$

9 $\mathcal{O}$ is the usual big-O Landau notation: a function $g(n)$ is $\mathcal{O}(f(n))$, if there exists some finite constant, such that $|g(n)| \leq |\mathrm{C}f(n)|$ for all n large enough. I will additionally use the little-o notation: $g(n)$ is $o(f(n))$ if the previous inequality holds *for any* $\mathrm{C} > 0$ for large enough n, or, equivalently, that $g(n)/f(n)$ goes to 0 as n grows large.

The type-II error probability is the same, up to the proportionality constant. For the appropriate constant,[10] B*, the FIG is thus given by*

$$\mathrm{FIG}(n) = \mathrm{B}\frac{\rho^n}{\sqrt{n}}\left(1 + \mathcal{O}\left(\frac{1}{\sqrt{n}}\right)\right) \tag{5}$$

To prove this, MS02 use a classic approach similar to Cramér (1938) and Chernoff (1952), namely, an exponential change of measure and Edgeworth series approximation. In this context, that method yields a conservative $\mathcal{O}(n^{-1/2})$ bound on the error. Later, I will state a version of this result with sharper bounds on the error using a saddlepoint approach.

From this result we can immediately derive the Chernoff's asymptotic relative efficiency for this setting:

Corollary 1 (Chernoff ARE). *If $\mathscr{E}_1$ and $\mathscr{E}_2$ are two information sources, then, at large samples, the minimum number of independent samples n_2 from $\mathscr{E}_2$ required to perform at least as well as n_1 from $\mathscr{E}_1$ satisfies*

$$\frac{n_2}{n_1} \simeq \frac{\beta_1}{\beta_2}$$

where β_j are the sources' respective Chernoff precisions.

Proof sketch. Take the log of the expression for FIG given by Eq. 5 for both information sources and choose n_1, n_2 to set them equal.

Note that these results do not depend on the binary actions I've considered here. Lem. 1 gives the asymptotic probability that the LLR sum exceeds some finite cutoff, so by taking differences, one can approximate the probability that the LLR falls in any interval whose lower bound is not $-\infty$. Thus any decision problem with finitely many actions will have a FIG that falls exponentially with the Chernoff number of the information source.

Notice that prior and payoff structure don't enter into this approximation (except through the proportionality constant). Given two information sources, for large enough samples any two DMs[11] eventually agree on which is superior at large samples!

10 MS02 give a precise form for the proportionality constant. The interested reader can find the precise form for the constant in Appendix A.4, but it doesn't affect my results, so I omit it from the discussion.

11 There is a subtlety here. For large enough samples, any two DM's agree, but there may not be any sample size such that *all* DM's agree. Azrieli (2014) calls the latter, stronger property *eventual Blackwell sufficiency* and provides a sufficient condition for it.

3.2 *The many state case*

Thus far I've relied heavily on LLRs to turn Bayes's rule into a sum and apply large deviations. However, with more than two states, we can no longer express beliefs in terms of a single LLR; we need an LLR for every state relative to some base state. Furthermore, the mistake probabilities become the probability that this vector of LLRs (whose entries are non-independent) falls in some non-interval subset.

Although at first glance, it seems we would need a non-trivial extension to Lem. 1, we can get some simplification out of the exponentially falling nature of the mistake probability.

First, define $a^*(\theta)$ as the optimal action in state θ, and $\alpha(n\,; a, \theta)$ as the probability of mistakenly choosing suboptimal action a in state θ. Then write the FIG as

$$\mathrm{FIG}(n) = \sum_{\theta} p_\theta \sum_{a \neq a(\theta)} \alpha(n\,; a, \theta)(u(a^*(\theta), \theta) - u(a, \theta))$$

So, the FIG generalizes to a similar form: the ex-ante expectation over all possible mistakes. Thus as before, approximating the FIG is a matter of approximating each error probability.

Now, for each dichotomy, $\theta\theta'$, we can define a MGF for the LLR of the signal realizations:

$$\mathrm{M}(t; \theta\theta') = \int f(r \mid \theta)^t f(r \mid \theta')^{1-t} \mathrm{d}r$$

and thus for each pair of states, we can define a Chernoff number, $\rho(\theta\theta')$, and precision, $\beta(\theta\theta')$.

Based on the two-state results, we might guess that each dichotomy FIG is roughly proportional to a term falling exponentially in that dichotomy's Chernoff number, and thus asymptotically all dichotomies except the one with the highest Chernoff number (lowest precision) are negligible for large samples. MS02's Thm. 4 makes this intuition rigorous:

Lemma 2 (MS02, Thm. 4). *Suppose there is a unique least-precision dichotomy,* $\theta\theta'$*, (generically true). Then for some* $\bar{\rho} < 1$*, we can write*

$$\mathrm{FIG}(n) = (p_\theta + p_{\theta'})\mathrm{FIG}^*_{\theta\theta'}(n)(1 + \mathcal{O}(\bar{\rho}^n))$$

where $\mathrm{FIG}^*_{\theta\theta'}(n)$ *is the FIG from observing* n *indpendent samples from the information source when the state is known to be either* θ *or* θ'*.*

Proof sketch. First, clearly $\mathrm{FIG}(n) \geq (p_\theta + p_{\theta'})\mathrm{FIG}^*_{\theta\theta'}(n)$ because the right-hand side is the FIG from observing n samples *plus* a signal that perfectly the reveals the state provided the state is neither θ nor θ' (more information implies lower FIG). The remainder of the proof consists of showing that each state-conditional mistake probability $\alpha(n; a, \theta)$ is $\mathcal{O}(\max_{\theta' \neq \theta} \rho(\theta\theta')^n)$, heuristically because each mistake probability requires a large deviation in at least one of the LLRs, and the most likely large deviation is the one for the dichotomy with the least precision. See MS02 for the full proof. □

Using this result plus Lem. 1, with appropriate choice of constant, B, we can rewrite

$$\mathrm{FIG}(n) = \mathrm{B}\frac{\rho(\theta\theta')^n}{\sqrt{n}}\left(1 + \mathcal{O}\left(\frac{1}{n}\right)\right)$$

where $\theta\theta'$ is the least-precision dichotomy.

Simply put, for large samples, an information sources asymptotic performance is entirely given by the pair of states which is most struggles to tell apart.

Thus, all our conclusions from the two-state case still hold: because the least-precision dichotomy is only a function of signal properties, all DMs eventually agree on which of two signals is optimal at large samples, and the Chernoff ARE is still given by the ratio of precisions—though now the ratio of each source's respective worst-case dichotomy precision.

Now if all we cared about was the ability to compare corner strategies, we could stop here—Chernoff's ARE tells us everything we need; however, Chernoff tells us nothing about tradeoffs between tests at interior bundles. In order to get a complete picture, we need to extend the theory to cover composite information sources.

3.3 Composite experiments

First, consider the two-state/two-action case again. Suppose our DM has access to a menu of various conditionally independent information sources, $\mathcal{E}_1, \dots, \mathcal{E}_\mathrm{J}$. If she purchases a bundle $\mathrm{n} = [n_1, \dots, n_\mathrm{J}]$ of samples, then the MGF of the LLR distribution of this composite experiment will simply be

$$\mathrm{M}_\mathrm{n}(t) = \prod_{j=1}^{\mathrm{J}} \mathrm{M}_j(t)^{n_j}$$

because the MGF of a sum of independent random variables is the product of the constituent MGFs. Define the *composite factor*, $\omega \equiv [n_1/\mathrm{N}, \dots, n_\mathrm{J}/\mathrm{N}]$ where $\mathrm{N} \equiv \sum n_j$ is the *total* sample size. We can then treat the composite signal as N samples from a single information source with LLR MGF given by $\mathrm{M}_\omega(t) \equiv \prod \mathrm{M}_j(t)^{\omega_j}$ and write the composite

MGF as[12]

$$M_{\mathrm{n}}(t) = M_\omega(t)^N$$

Not surprisingly given the previous section, we will be minimizing this object: define $\rho_\omega \equiv \min_t M_\omega(t)$ and $\beta_\omega \equiv -\log \rho_\omega$ as the *composite* Chernoff number and precision respectively, and $\tau_\omega \equiv \arg\min_t M_\omega(t)$.

The composite Chernoff number and precision can then be broken up by contribution from each information source by writing $\rho_{\omega i} \equiv M_j(\tau_\omega)$ and $\beta_{\omega j} \equiv -\log \rho_{\omega j}$. Call these the ω-component Chernoff numbers and precisions respectively. So we can then write the total Chernoff number and precision from an n bundle as

$$\rho_\omega^N = \prod_{j=1}^{J} \rho_{\omega j}^{n_j}$$

$$N\beta_\omega = \sum_{j=1}^{J} n_j \beta_{\omega j}$$

We can now state an augmented version of Lem. 1:

Proposition 2. *For a two-state/two-action decision problem, the probability of a type-I error given a bundle of* $\mathrm{n} = [n_1, \ldots, n_J]$ *conditionally independent samples from each information source is proportional to*

$$\alpha_I(\mathrm{n}) \propto \frac{\prod_{j=1}^{J} \rho_{\omega j}^{n_j}}{\sqrt{N}} \left(1 + \mathcal{O}\left(\frac{1}{N}\right)\right) \tag{6}$$

where N *is the sum of samples from each source. The type-II error is the same up to the proportionality constant. Merging terms into* B_ω*, the FIG can thus be written*

$$\mathrm{FIG}(\mathrm{n}) = B_\omega \frac{\prod_{j=1}^{J} \rho_{\omega j}^{n_j}}{\sqrt{N}} \left(1 + \mathcal{O}\left(\frac{1}{N}\right)\right) \tag{7}$$

Proof sketch. The result follows from a direct application of a *saddlepoint approximation* due to Lugannani & Rice (1980), building on a method due to Daniels (1954). Roughly, the idea is that, because the MGF is analytic, it's minimum along the real axis corresponds to a maximum along the perpindicular imaginary axis, and thus, the integrand

12 Note that M_ω will not generally be a valid MGF, but for our purposes it will suffice to treat it as such.

of the MGF (characteristic function) inversion formula becomes almost entirely determined by that maximum at high sample sizes. The full proof is a bit technical and thus relegated to the appendix.[13]

Besides the generalization to a multiple-source setting, Prop. 2 improves on Lem. 1 in two ways: first, it tightens the error bound to $\mathcal{O}(\mathrm{N}^{-1})$, and, second, the proof approach implies an asymptotic expansion that accounts for each source's contribution to the approximation error, and may be numerically useful. Because the expansion doesn't directly affect my results, I leave further discussion of it to the Appendix A.4.

We're primarily interested in *ordinal* properties of the information value, so we can simplify things further by taking a monotone transformation:

$$-\log \mathrm{FIG}(\mathrm{n}) \;=\; \sum_{j=1}^{\mathrm{J}} n_j \beta_{\omega j}\left(1+\mathcal{O}\left(\frac{\log(\mathrm{N})}{\mathrm{N}}\right)\right) \tag{8}$$

Then we have that constrained maximization of the above is equivalent to minimization of the FIG, and thus equivalent to maximization of the value, and thus that large sample bundles of equal precision, must have roughly equal information value for any decision maker.

From this it seems like tests must eventually be perfectly substitutable. But it's worse than that because we minimized the product of all the MGFs, rather than each of them individually, so we have

$$\begin{aligned} \min_t\{\mathrm{M}_\omega(t)\} &\geq \prod_{j=1}^{\mathrm{J}} \min_t\{\mathrm{M}_j(t)^{\omega_j}\}\} \\ \Leftrightarrow \qquad \beta_\omega &\leq \sum_{j=1}^{\mathrm{J}} \omega_j \beta_j \end{aligned} \tag{9}$$

with equality holding only for bundles of information sources all sharing the same minimizer. So, at least when it comes to distinguishing a single pair of states, all composite experiments are weakly worse than the sum of their parts in a large sample setting.

Given this, the optimal bundle for a constrained decision-maker follows immediately:

Corollary 2 (Corner bundles are (near) optimal for dichotomies). *Suppose a budget-constrained decision-maker has a collection of available information sources, $\mathscr{E}_1, \ldots, \mathscr{E}_\mathrm{J}$,*

13 Butler (2017) gives a relatively non-technical treatment of saddlepoint methods.

from each of which she may purchase an arbitrary number of samples at per-sample costs $c_1, \dots, c_J$ respectively. Suppose there is a unique source with the highest precision per dollar (true for generic costs), and without loss assume it's $\mathscr{E}_1$. Then for any (non-trivial) two-state decision problem, the composite factor of the optimal bundle approaches $\omega_1 = 1$ as the budget goes to infinity.

Proof. From Eq. 8, it suffices to show that the asymptotic payoff-per-dollar of the non-composite experiment is higher than that of any composite experiment. Ignoring lower-order terms, we want

$$\frac{\beta_1}{c_1} \geq \frac{\omega\beta_1(\omega) + (1-\omega)\beta_2(\omega)}{\omega c_1 + (1-\omega)c_2}$$

but this follows immediately from inequality 9. □

Note that this is not quite the same as corner bundles being precisely optimal. For example, the discrete nature of the problem often makes it optimal to consume some of a cheaper signal if only to make better use of the full budget[14] Nonetheless, because β_ω approaches the Chernoff precision of the non-composite experiment as ω approaches 0 or 1, if an interior bundle is optimal at large samples, it must have a composite factor arbitrarily close to the optimal corner.

Now, if we only cared about two-state decision problems, all of this effort would have been of little use: corners (or at least near corners) are always optimal at large samples, so Chernoff's original result would have been all we needed. But things get more interesting when we move to the many state world.

Recall from Section 3 that now an experiment is characterized by its collection of LLR MGFs, one for each state pair. We thus then have a Chernoff index/precision for each state pair: denote, $\rho_\omega(\theta\theta') \equiv \min M_\omega(t;\theta\theta')$ and $\beta_\omega(\theta\theta') \equiv -\log \rho_\omega(\theta\theta')$, and the ω-component parts, $\rho_{\omega j}(\theta\theta')$ and $\beta_{\omega j}(\theta\theta')$, similarly.

Then, using Lem. 2 (MS02's Thm. 4) we can generalize Prop. 2 to the many state case:

Proposition 3. *Assume that for all but a measure-zero set of composite factors, the least-precision dichotomy,*

$$D_\omega \equiv \underset{\theta,\theta' \text{ s.t. } \theta\neq\theta'}{\arg\min}\ \max_t \left\{ -\sum_{j=1}^{J} \omega_j \log M_j(t;\theta\theta') \right\} \equiv \underset{\theta,\theta' \text{ s.t. } \theta\neq\theta'}{\arg\min} \sum_{j=1}^{J} \omega_j \beta_{\omega j}(\theta\theta')$$

14 e.g. if $\mathscr{E}_1$ has a price of \$2 per sample and $\mathscr{E}_2$ a cost of \$1 per sample, it will often be optimal to purchase 1 sample from $\mathscr{E}_2$ simply to avoid leftover budget.

is unique (generically true). Then for a finite-state/finite-action decision problem, the FIG from $\mathrm{n} = [n_1, \ldots, n_\mathrm{J}]$ *conditionally independent samples from each information source is given by*

$$\mathrm{FIG}(\mathrm{n}) = \mathrm{B}_\omega \frac{\prod_{j=1}^{\mathrm{J}} \rho_{\omega j}(\mathrm{D}_\omega)^{n_j}}{\sqrt{\mathrm{N}}} \left(1 + \mathcal{O}\left(\frac{1}{\mathrm{N}}\right)\right)$$

where B_ω *is a uniformly bounded constant that depends only on decision-maker characteristics (prior and payoff structure) and the minimizer of the composite MGF for the least-precise dichotomy.*

Proof. Follows immediately from plugging the expression for the two-state FIG given by Prop. 2 into the many-state FIG equation given by Lem. 2. □

As before, we can monotonically transform this for simplicity letting D_ω continue to be the least-precision dichotomy:

$$-\log \mathrm{FIG}(\mathrm{n}) \;=\; \sum_{j=1}^{\mathrm{J}} n_j \beta_{\omega j}(\mathrm{D}_\omega) \left(1 + \mathcal{O}\left(\frac{\log(\mathrm{N})}{\mathrm{N}}\right)\right) \tag{10}$$

But unlike before, we can see the potential for complementarities between sources because

$$\min_{\theta,\theta' s.t. \theta \neq \theta'} \max_t \left\{ -\sum_{j=1}^{\mathrm{J}} \omega_j \log \mathrm{M}_j(t; \theta\theta') \right\}$$

is not necessarily smaller than

$$\sum_{j=1}^{\mathrm{J}} \min_{\theta,\theta' s.t. \theta \neq \theta'} \max_t \{-\log \mathrm{M}_j(t; \theta\theta')\}$$

Roughly, because the value of information at large samples only depends on the pair of states hardest to distinguish, information sources may complement each other by covering for each other's weaknesses. So even though the composite test might be less precise for every individual state pair, the *lowest* precision may actually be higher. Putting everything together, we have a natural criterion for finding a low-price optimal bundle:

Corollary 3 (Maxi-Min precision bundles are (near) optimal). *Let* $\omega^*(c)$ *be the composite factor that uniquely maximizes the mininum precision per dollar for a given set of* *information sources and cost vector,* c

$$\omega^*(c) = \arg\max_{\omega} \min_{\theta,\theta' s.t. \theta \neq \theta'} \frac{\beta_\omega(\theta\theta')}{\omega \cdot \mathrm{c}}$$

Then the composite factor of the optimal bundle approaches $\omega^*(c)$ *as the budget goes to infinity.*

Proof. The proof is the same as for Cor. 2, except for a minor technical complication, handled in Appendix A.2, arising from applying the approximation to bundles where the least-precision dichotomy is non-unique. □

In order to get a handle on when interior solutions exists and what form they typically take, it will be useful to think about the problem more geometrically. To that end, I'll now extend the notion of asymptotic relative efficiency to cover small tradeoffs between interior bundles by defining a notion of "marginal rate of substitution" appropriate for a discrete setting.

4 CONSUMER THEORY FOR CHEAP INFORMATION

4.1 "Marginal" Rate of Substitution

The previous section implies a natural maxi-min criterion for finding the low-price optimal bundle: for a given cost vector, find the relative proportion of samples from each bundle that maximizes the lowest precision per dollar. We can say a bit more about the nature of such solutions by applying some marginal analysis.

First, note that the definition and computation of Chernoff precisions for composite sources, doesn't depend at all on the samples being discrete. Thus, we can happily pretend for now that samples are a continuous quantity, and, thus, at any point where the least-precision dichotomy is unique, we can differentiate Eq. 10 to find the slope of an iso-least-precision line at a bundle with composite factor ω:

$$-\frac{\mathrm{d}n_2}{\mathrm{d}n_1} = \frac{\beta_{\omega 1}(\mathrm{D}_\omega)}{\beta_{\omega 2}(\mathrm{D}_\omega)}$$

where D_ω is the least precision dichotomy for that composite factor as before. Because, for fixed dichotomies we have that composite sources are less precise than the sum of their parts, we might expect the iso-precision lines to be outward bowing, and a quick computation confirms this: if the iso-least-precision line has a defined derivative, it must be *increasing* in magnitude with n_1. However, because we only care about the least-precision, there are finitely many inward pointing kink points where the least-precision dichotomy changes (recall Fig. 1 from Section 1).

With two available information sources, we clearly then must generically have that the maxi-min precision per dollar composite factor, $\omega^*(c)$, must be at either a corner or one of finitely many interior kink points where two dichotomies have the same precision. With more than two sources, we can apply this same geometric reasoning in higher dimensions to characterize the general maxi-min precision per dollar bundle:

Proposition 4 (Optimal bundle has as many sources as dichotomies of equal precision). *Suppose a decision-maker has a collection of available information sources, $\mathscr{E}_1, \dots, \mathscr{E}_J$, from each of which she may purchase an arbitrary number of samples at per-sample costs $c_1, \dots, c_J$ respectively. Then generically the relative proportions of each source in a bundle that maximizes the minimum precision per dollar generically must lie either at a corner (all but a single ω_j is zero), or must lie at a kink point where k dichotomies have the same precision, where k is the number of distinct information sources in the bundle, $|\{\mathscr{E}_j \mid \omega_j > 0\}|$.*

Proof sketch. Follows from a geometric reasoning: If there are k sources in the maxi-min precision per dollar bundle, that bundle lies at the tangency of a k-dimension hyperplane defined by the cost vector for those k sources and the iso-least-precision contour. Because the iso-least-precision contour is composed of multiple inward-bowing faces, the tangency must generically lie at an intersection of k faces. (Intuitively, if you drop a spiky ball, it generically lands on a point.)

Prop. 4 has an intuitive meaning: in a typical scenario, each test covers in the optimal bundle for a particular weakness of the other tests.
Alas, we cannot forever avoid the fact that samples are fundamentally discrete.[15] At the end of the day, we'd like to be able to draw meaningful conclusion about tradeoffs between discrete numbers of samples from the continuous curvature of the iso-precision lines, but in order to do so, we first need a formal discrete analog to the marginal rate of substitution:

$$k_2 = \min\{k \,:\, v(n_1 - k_1, n_2 + k, n_2, \dots, n_J) \geq v(n_1, n_2, n_2, \dots, n_J)\}$$

15 We could get around the issue by convexifying the choice space in some way. For example, we could consider lotteries over different sample bundles (which themselves define a new information source). One particularly elegant way to do this is to consider the *Poissonization* of an information source, where instead of getting n samples for sure, you get a random, Poisson-distributed number of samples. The Poisson parameter, a continuous object, then naturally takes the place of sample size (the sum of two Poisson draws is itself Poisson distributed). Although mathematically elegant, such an approach adds technical complication for little benefit.

That is, k_2 is the *minimum* number of additional samples of $\mathscr{E}_2$ required to at least compensate for a loss of k_1 from $\mathscr{E}_1$, all else held fixed.[16] We can then define a k_1-discrete rate of substitution as

$$\mathrm{DRS}_{12}(\mathrm{n}; k_1) \equiv \frac{k_2}{k_1}$$

Now, recall that the component precisions only depend on ω. Furthermore, the change in ω from the above substitution is eventually small—roughly $\mathcal{O}(k_1/\mathrm{N})$—so for large samples sources are *locally* perfectly substitutable, provided the substitution is small relative to total sample size and the least-precision dichotomy is unique at the bundle's composite factor. This suggests a natural definition for an asymptotic marginal rate of substitution:

Define the *asymptotic marginal rate of substitution* (AMRS) as

$$\mathrm{AMRS}_{12}(\omega) \equiv \lim_{\mathrm{N}\to\infty} \mathrm{DRS}_{12}(\omega_1\mathrm{N}, \ldots, \omega_J\mathrm{N}; k(\mathrm{N}))$$

Where the substitution size, $k(\mathrm{N})$, goes to infinity, but at a much smaller rate than N—i.e., $k(\mathrm{N}) = o(\mathrm{N})$.[17]

Putting aside notation, the AMRS tells us that for large enough total sample sizes, giving up k samples from $\mathscr{E}_1$ in exchange for more than $k \times \mathrm{AMRS}$ yields higher payoff. Symmetrically, receiving less than $k \times \mathrm{AMRS}$ in exchange yields a lower payoff. Graphically, it defines the slope of the boundary between upper and lower contour sets of a bundle with the given composite factor for large enough total sample size.

Using this definition, the AMRS between two information sources takes the form one might expect:

Proposition 5 (Substitutability of information sources). *Suppose the least-precision dichotomy is unique for composite factor ω, then for large enough N, the number of samples of $\mathscr{E}_2$ required to just compensate for a loss of k samples from $\mathscr{E}_1$ is*

$$k\mathrm{DRS}_{12}(\mathrm{n}; k) = \left\lceil k\frac{\beta_{\omega 1}(\mathrm{D}_\omega)}{\beta_{\omega 2}(\mathrm{D}_\omega)} \right\rceil$$

where D_ω is the least-precision dichotomy for composite factor ω, and $\lceil\cdot\rceil$ denotes the

16 k_1 can be negative.

17 Of course, we must start the limit sequence from high enough N so that the substitution is possible.

ceiling function. The AMRS is thus given by

$$\mathrm{AMRS}_{12}(\omega) = \frac{\beta_{\omega 1}(\mathrm{D}_\omega)}{\beta_{\omega 2}(\mathrm{D}_\omega)} \tag{11}$$

Proof sketch. Heuristically, we can solve for an indifference condition using Eq. 10, given a loss of k samples from $\mathscr{E}_1$, ignoring changes in the composite factor since they are eventually very small for a given substitution:

$$n_1\beta_{\omega 1}(\mathrm{D}_\omega) + n_2\beta_{\omega 2}(\mathrm{D}_\omega) = (n_1 - k)\beta_{\omega 1}(\mathrm{D}_\omega) + (n_2 + k\mathrm{AMRS})\beta_{\omega 2}(\mathrm{D}_\omega)$$

Rearranging terms to solve for AMRS gives the claimed formula. The formal proof is mostly a computational exercise and is thus relegated to the appendix.

Fig. 2 illustrates the AMRS numerically against a numerically computed lower boundary of an upper contour set. The top example illustrates a two-state environment where the AMRS always traces a bowed-out curve, and the bottom plot represents a three-state environment where one of the three dichotomies is never least-precision (similar to the example in Fig. 1). Section 5 will later discuss some typical properties of this approximation for typical information sources as sample sizes increase.

With this, we now have a formal basis for using the slope of iso-least-precision lines as an approximation for the rate at which samples from different sources may be exchanged at large samples. We thus have everything we need to characterize demand for information using standard consumer theory.

4.2 Demand under constrained optimization

With the results from the previous section, we can now easily derive some properties of information demand in this setting:

First, because the AMRS only depends on the relative proportions of each source (geometrically, the ray through the origin) for a bundle, preferences over information sources are homothetic by definition, and thus the optimal bundle composition is eventually income-independent. That is, when budgets are large enough, no source is ever either inferior or a luxury.

Proposition 6 (All sources are eventually normal goods). *For a generic cost vector—i.e. those that have a unique $\omega^*(c)$—the income elasticity of demand for all sources approaches unity as the budget grows large.*

Second, because optimal bundles must lie at one of finitely many kink points where the least-precision dichotomy is non-unique, we must have that demand for information behaves somewhat like the demand for perfect complements under Leonteif pref-

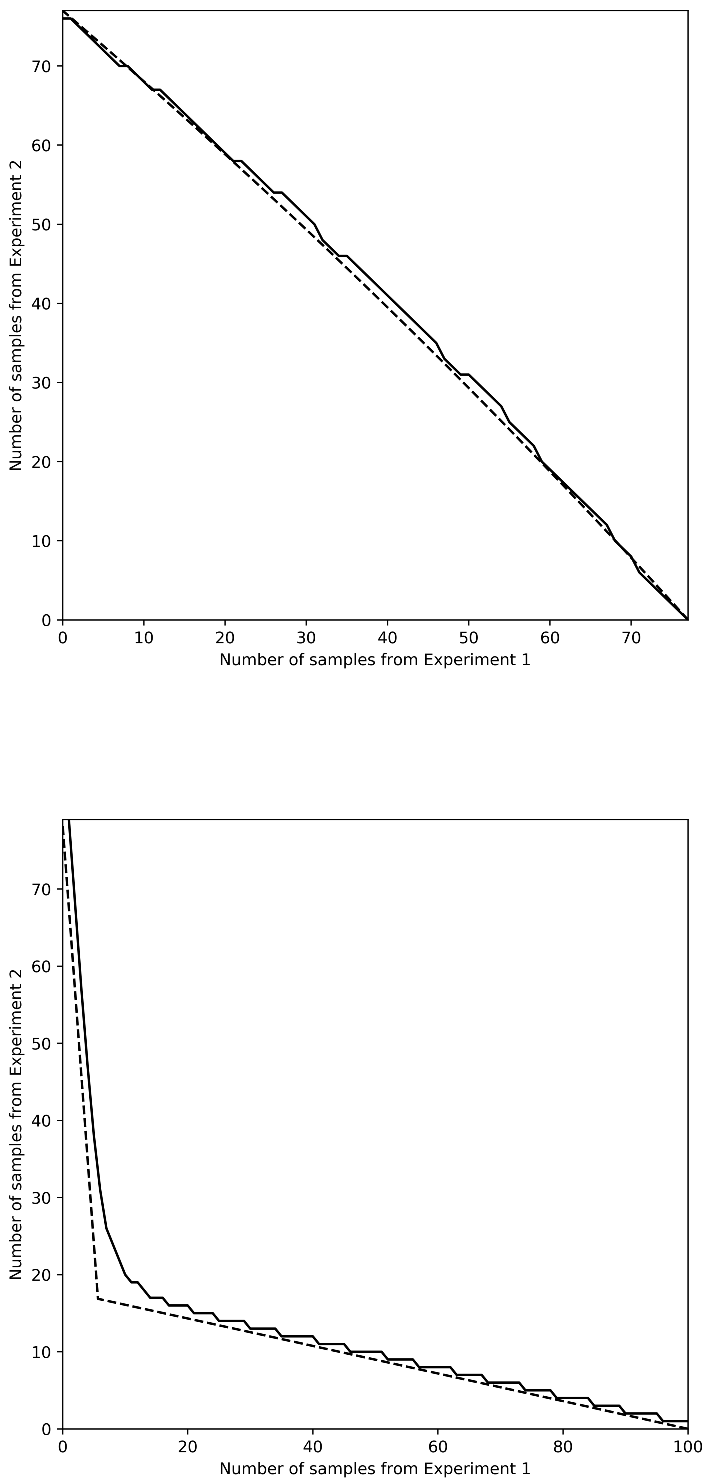

Figure 2: Top: Locus of bundles that perform *just* better than the lower right corner bundle (solid line), against the boundary predicted by Prop. 5 (dashed line) in a two-state decision problem. Bottom: The same plot for a three-state decision problem similar to the one illustrated by Fig. 1. The slope of the dashed line in both cases is by definition AMRS_{12}

erences at least locally. That is, for generic costs, the composition of the optimal bundle doesn't change with small changes in cost, but demand changes discontinuously around finitely many costs where the optimal kink changes.

We can thus write the demand elasticities as

Proposition 7 (Price elasticities). *For generic costs, the price elasticity of demand for all sources given a change in cost of source $\mathscr{E}_1$ is approaching*

$$\frac{-\omega_1^*(c)c_1}{\omega^*(c)\cdot c}$$

as the budget grows large. Holding all other costs fixed, there are finitely many costs around which elasticities explode.

Put another way, for small prices changes, information sources are all perfect Marshallian complements, and changes in demand are entirely attributable to income effects.

These results further reinforce the peculiar nature of information goods, especially when considering the implications for competition between information sellers. Of course in many cases information is non-rival (and often difficult to make excludable), so there was already little reason to think competition between firms with linear pricing was sustainable, but these results show that, even ignoring such issues, a market for information goods, at least in simple environments with relatively few possible states,[18] requires some special treatment.

For example, Prop. 7 implies that for any cost ratio with an interior optimal bundle, demand is inelastic, and thus, all else fixed, a monopolistically competitive firm can always locally improve its profits by raising its prices at least up to the point where the optimal bundle jumps to the next kink. Thus, in a fixed budget scenario, a monopolistic competition is likely to yield to either monopoly or some form of non-linear pricing such as a subscription scheme.

5 HOW GOOD IS THE APPROXIMATION?

It's worth considering whether any of this even matters. After all, in principle, the approximation only holds when many samples are purchased and error probabilities are vanishing. Nonetheless, I claim that these approximation are in fact useful. To that end, I briefly consider the performance of the considered approximations, both theoretically and numerically.

18 The approximations work best when there are relatively few states. See the next section for further discussion.

5.1 Theoretical performance

First, recall that the full-information gap approximation on which these results were based has a relative error roughly proportional to $1/N$, which by asymptotic approximation standards is respectable. By comparison, the standard central limit theorem approximation, a staple method for approximating standard errors in applied work, typically has an error proportional to $1/\sqrt{N}$. Although the approximations are not substitutes for one another, the central limit theorem is a useful reference point due to it's ubiquity in statistical approximations.

Second, from a practical standpoint, even though mistakes are astronomically unlikely, the relative nature approximation implies that the predicted optimal bundle eventually performs *much* better than any alternative:

Proposition 8 (The optimal policy is eventually much better than any other). *Let* n_Y^* *be the sample vector with composition closest to* $\omega^*(c)$ *feasible under budget* Y*. Further let* n_Y^ω *be a policy with composition factor* $\omega \neq \omega^*(c)$*. Then we have*

$$\lim_{Y\to\infty} \frac{\mathrm{FIG}(n_Y^*)}{\mathrm{FIG}(n_Y^\omega)} = 0$$

Proof. It suffices to show that $\log(\mathrm{FIG}(n_Y^\omega)) - \log(\mathrm{FIG}(n_Y^*))$ ratio goes to infinity for large budgets. But this follows immediately because Eq. 8 and Eq. 10 imply that $-\log(\mathrm{FIG})$ is the budget times the precision-per-dollar of the of the bundle. □

More succinctly, as budgets get large, the FIG from following the optimal policy is eventually arbitrarily smaller than the FIG from any other policy.

Put another way, suppose we were to flip the problem asking instead: what is the minimum budget required to achieve a certain performance? In this case, Prop. 8 implies that as the target FIG gets small, the budget required for the optimal strategy is arbitrarily smaller than the budget required for some other strategy, and thus that the budget necessary to achieve a fixed level of performance is extremely sensitive to the choice of bundle at large samples.

5.2 Numerical observations

Lastly, let's consider how the approximation works in actual practice.

- include tables and plots showing approximation error for
  - DRS
  - Demand in the constrained case (and elasticities)
  - Demand in the quasilinear case (and elasticities)

as sample sizes get large.

Finally, I note some qualitative properties of the approximations in practice, many of which are apparent in Fig. 2:

First, although indifference curves are definitively bowed outward for generic pairs of sources, in practice the outward bow is very subtle. The top plot of Fig. 2 was specifically selected to maximize the bow, and even then, the bow is barely perceptible. Given this, except in extraordinarily extreme cases, if two information sources differ in their least precision dichotomy, then interior solutions will exist for a non-empty set of price ratios.[19]

Second, the approximation performs relatively poorly very close a kink. Heuristically, there are two reasons for this: first, the approximation throws out all mistakes but the most likely one, but close to a kink, the second most likely mistake is very close to the most likely, so discarding it leads to a sizable overestimate of value, and second, because the kinks are inward pointing, the total sample size tends to be less than at the corners. For example, in Fig. 2, total sample size is roughly 30 at the kink, but around 100 at the corners.

Finally, because the true contour set boundary tends to be relatively smooth in the vicinity of a kink (although a very tight curve at large samples), the substitution effects will tend to be larger if there are more kinks. Since environments with more possible underlying states tend to have more kinks, it's reasonable to expect that contour sets would look more like the smooth curves of a textbook indifference curve. I thus caution against applying these approximations outside of simple environments with relatively few alternatives.

6 CONCLUSION AND FUTURE RESEARCH

This paper contributes to the literature on information values by applying previous results from the theory of large deviations to the evaluation of composite information sources. In particular, I've applied and extended the results of the results of Moscarini & Smith (2002) to develop an ordinal theory of tradeoffs for information sources.

I've shown that, for low costs/large budgets, maximizing information value is equivalent to maximizing the total Chernoff precision of the dichotomy that the information bundle most struggles to tell apart, and thus interior solutions can occur whenever two sources differ in their least-precision dichotomies.

19 The only way this can fail is if the iso-precision line for one dichotomy is sufficiently bowed so as to intersect another twice. This would lead to two kink points not on the convex hull of an upper contour set, and thus are never optimal.

Furthermore, I've explored implications for the standard consumer theory questions such as price, cross-price, and income elasticities for information demand. Specifically, in the "simple" (finite state) decision setting considered here, information goods will only be consumed in one of finitely many possible relative proportions, and thus for small changes in price, will behave like perfect complements.

These results naturally contribute to the broad literature in economics on information demand and suggest possible functional forms for information demand in applied settings such as media. On a more practical note, these approximations suggest a criterion (maxi-min precision per dollar) for the optimal design of experiments when the space of alternatives is small. Of course, the finite state assumption is somewhat restrictive, but the large deviation approach considered here should serve as a starting point for a general criterion for experiment design founded in Bayesian decision theory.[20]

20 In a separate project I show that Fisher information is the continuous-state analog to the Chernoff precision.

Appendix A omitted proofs

A.1 Proof of Prop. 2

First I prove the result for a single information source using a result of Lugannani & Rice (1980)—henceforth, LR80:

Lemma (LR80 CDF Saddlepoint approximation). *Suppose* Y *is a random variable with all finite moments, MGF* $M(t) = \mathbb{E}(e^{ty})$ *defined on an open interval containing the origin, and characteristic function* $\phi(t) \equiv \mathbb{E}(e^{ity})$*. Then, if* $\bar{Y}_n = \sum_i^n Y_i/n$ *is the sample average of n i.i.d. draws from* Y*, the probability that* $\bar{Y}_n$ *is greater than* $\xi > \mathbb{E}(Y)$ *can be written*

$$\mathbb{P}(\bar{Y}_n \geq \xi) = \frac{e^{n(K(\tau(\xi)) - \tau(\xi)\xi)}}{\tau(\xi)\sqrt{2\pi n K''(\tau(\xi))}} \left(1 + \mathcal{O}\left(\frac{1}{n}\right)\right) \tag{12}$$

where $K(t) \equiv \log(M(t))$ *is the cumulant generating function of* Y *and* $\tau(\xi)$ *is the minimizer of* $K(t) - tx$.

For the interested reader, in A.4, I provide a detailed proof sketch of the above result. For the full proof, see Lugannani & Rice (1980). We can now apply the above lemma to the case of LLR distributions:

In the two action case, the mistake probability is the probability that the average LLR exceeds $\xi_n \equiv (\bar{l} - l)/n$, where l is the prior LLR and $\bar{l}$ is the indifference LLR. Since the LLR has negative expectation, for n high enough, ξ_n exceeds the LLR expected value. Now define τ_n as the minimizer of $K(t) - t\xi_n$. Let τ be the minimizer of $K(t)$ (and thus the minimizer of the MGF). By applying Taylor's theorem and the FOC, $K'(\tau) = 0$, we can the write

$$\begin{aligned}
\tau_n &= \tau + \mathcal{O}(1/n) \\
K(\tau_n) &= K(\tau) + \frac{1}{2n^2}K''(\tau) + \mathcal{O}(1/n^3) \\
K''(\tau_n) &= K''(\tau) + \mathcal{O}(1/n)
\end{aligned}$$

Note that by definition of the Chernoff precision and number we have, $K(\tau) = -\beta$ so $e^{nK(\tau)} = \rho^n$. We can then plug each of these into Eq. 12 and apply Taylor's theorem again. Breaking it down into parts we have

$$\begin{aligned}
e^{n\mathrm{K}(\tau_n)} &= \rho^n(1+\mathcal{O}(1/n)) \\
e^{-n\tau_n\xi_n} &= e^{\tau(\bar{l}-l)}(1+\mathcal{O}(1/n)) \\
\tau_n\sqrt{2\pi n\mathrm{K}''(\tau_n)} &= (\tau+\mathcal{O}(1/n))\sqrt{2\pi n\mathrm{K}''(\tau)+\mathcal{O}(1)} \\
&= \tau\sqrt{2\pi n\mathrm{K}''(\tau)}(1+\mathcal{O}(1/n))
\end{aligned}$$

Plugging each of the above parts into Eq. 12 we have that the probability of a type-I error is

$$\alpha_{\mathrm{I}}(n) \propto \frac{\rho^n}{\sqrt{n^2}}\left(1+\mathcal{O}\left(\frac{1}{n}\right)\right)$$

as claimed. We can then easily generalize this result to multiple actions because the probability of an arbitrary mistake is the probability that the LLR falls in an interval whose lower bound is eventually higher than the LLR expected value.

To get the result for composite samples, one only need to treat the composite signal as $\mathrm{N} = n_1 + \ldots + n_j$ samples from a single source with LLR MGF given by $\mathrm{M}_\omega(t) = \prod \mathrm{M}_j(t)^{\omega_j}$ and minimizer τ_ω. Then we can use the fact that for a signal with composite factor ω, we can write the Chernoff number as[21]

$$\rho_\omega^{\mathrm{N}} = \prod_{j=1}^{\mathrm{J}} \rho_{\omega j}^{n_j}$$

and

$$\mathrm{N}\varsigma_\omega^2 = \sum_{j=1}^{\mathrm{J}} n_j\varsigma_{\omega j}^2$$

where $\rho_{\omega j} \equiv \mathrm{M}_j(\tau_\omega)$ and $\varsigma_{\omega j}^2 \equiv \mathrm{K}_j''(\tau_\omega)$. Plugging these into the previously derived expression for the single-source error probability yields the claimed expression.

Lastly, we need to show that that the proportionality constant is bounded and differentiable in the composite factor, ω. Plugging the appropriate terms from above into

21 This is a bit hand-wavy: M_ω generally isn't the MGF of a valid distribution. However, the saddlepoint approximation proof in A.4 doesn't require M_ω to actually be a valid MGF. Besides the tighter bound on the error, this is a key advantage of the saddlepoint method over the method used by MS02, which doesn't obviously work if M_ω isn't a valid MGF.

Eq. 12 gives the proportionality constant, A, for the type-I error approximation.

$$A = \frac{e^{\tau(\bar{l}-l)}}{\sqrt{2\pi\tau^2\varsigma^2}}$$

The proportionality constant for the type-II error can be computed similarly. Recall that flipping the LLR simply reflects the MGF around $t = 0.5$ (see Eq. 1) and thus does not affect ρ or ς^2. One must only change $\bar{l} - l$ to their appropriate flipped LLR, and replace τ by $1 - \tau$.

Recall that, in the case of composite experiments, $\varsigma_\omega^2 = \sum \omega_i \varsigma_{i\omega}^2$, so we could have equally written the composite approximation as

$$\alpha_I(n) \propto \frac{\prod \rho_{i\omega}^{n_i}}{\sum n_i \varsigma_{i\omega}^2}\left(1 + \mathcal{O}\left(\frac{1}{N}\right)\right)$$

To show the constant is bounded, it then suffices to show that both τ is bounded and ς^2 is bounded away from zero. As previously mentioned, $\tau \in (0, 1)$ so that's bounded. Now, recall that the MGF is *strictly* convex in t under our assumptions, and thus $\varsigma^2 > 0$. Finally, since $M_\omega(t) \equiv \prod M_i(t)^{\omega_i}$. Thus we must, so $\varsigma_\omega^2 \geq \min\{K_i''(\tau_\omega)\}$. This implies ς_ω^2 is bounded away from zero, provided that τ_ω is bounded away from zero and one, which it is since τ_ω must lie between the min and max τ_i. This follows from strict convexity: $K_\omega'(\min\{\tau_j\}) = \sum \omega_i K_i'(\min\{\tau_j\}) > 0$ because every K_i' at least weakly positive at that point. Similar logic shows that τ_ω is smaller than the biggest τ_i.

Finally, differentiability follows immediately from the fact that M_ω is infinitely differentiable in all ω_i. Since M_ω always has a unique minimum and varies differentiably in ω, we must haveτ and ς^2 differentiable in ω as well. □

Asymptotic expansion of the error: Although it doesn't affect my results, it's worth noting that the LR80 approximation admits a fuller asymptotic expansion. In particular, following Daniels (1987), we can write the relative error (the $\mathcal{O}$ term) as

$$\frac{1}{NK''(\tau)}\left\{\frac{1}{8}\frac{K^{(4)}(\tau)}{K''(\tau)} - \frac{5}{24}\frac{K^{(3)}(\tau)^2}{K''(\tau)^2} - \frac{1}{2\tau}\frac{K^{(3)}(\tau)}{K''(\tau)} - \frac{1}{\tau}\right\} + \mathcal{O}(N^{-2})$$

Further expansion is given by LR80. Importantly, this expansion allows some exploration of how each information source contributes to the error. In particular, recall that $K''(\tau) \equiv \varsigma^2$ and that for a composite experiment we have $\varsigma_\omega^2 = \sum \omega_i \varsigma_{i\omega}^2$. We thus could have written the error as

$$\frac{1}{\sum n_i \varsigma_{i\omega}^2} \mathrm{A}_\omega + \mathcal{O}(\mathrm{N}^{-2})$$

implying that signals with high curvature (high ς^2) have a larger contribution to reducing error. Furthermore, because $\mathrm{M}(0) = \mathrm{M}(1) = 1$ for any information source, but better sources have lower minima, better sources *tend* to have higher curvature. So what is meant by "asymptotic" depends on the quality of the sources considered, with more informative sources tending to require fewer samples to be well-approximated by the saddlepoint formula.

A.2 Proof of Cor. 3

To show that the composition of the optimal bundle approaches $\omega^*(c)$ as the budget goes to infinity, it suffices to show that a bundle with composition factor as close as possible to $\omega^*(c)$ eventually has a higher value than one ϵ away from $\omega^*(c)$ for Y high enough.

Now, the relative error of the $-\log$ FIG approximation at a given composite factor is smaller in magnitude than $A_\omega \log(N)N^{-1}$, for some constant A_ω. Given this, we can immediately say that the bundle nearest to $\omega^*(c)$ eventually beats any bundle with fixed ω, but to say that there is some budget where the bundle nearest to $\omega^*(c)$ beats *all* composite factors outside an ϵ region, we need to find an upper bound on A_ω that holds for all ω.

This is *almost* trivial since the space of composite factors (the simplex) is compact, except that Lem. 2 requires that the least-precision dichotomy be unique, so the space of compositions where the approximation holds possibly has holes and is thus might be non-compact.

In order to remedy this, I show that, although the approximation for the FIG fails, the approximation for the log FIG holds, thus filling the holes left by Lem. 2.

First, following MS02's reasoning, I show that the probability of mistakenly taking action a under state θ, $\alpha(n; a, \theta)$, is $\mathcal{O}(\rho(\theta h(\theta))^n)$ where $h(\theta)$ is the state hardest to distinguish from θ—i.e., $h(\theta) \in \arg\max_{\theta'} \rho(\theta\theta')$. Note that $h(\theta)$ need not be unique.

First, note that a necessary condition for a mistake is that the posterior expected payoff to mistakenly choosing action a must exceed that of choosing $a^*(\theta)$. Thus, an upper bound on the mistake probability is

$$\mathbb{P}\left(\sum_{\theta'\neq\theta} \frac{q_{\theta'}}{q_\theta} b_{\theta'} > -b_\theta | \theta\right)$$

where the $q_{\theta'}$ are the posterior beliefs and $b_{\theta'} \equiv u(a\,|\,\theta') - u(a^*(\theta)\,|\,\theta')$. Note that $-b_\theta > 0$ since $a^*(\theta)$ is assumed the optimal action in state θ. Now suppose $\beta_{\theta'} > 0$ for $\theta' \in \bar{\Theta}$, then we must have

$$\begin{aligned}
\mathbb{P}\left(\sum_{\theta'\neq\theta} \frac{q_{\theta'}}{q_\theta} b_{\theta'} > -b_\theta | \theta\right) &\leq \mathbb{P}\left(\sum_{\theta'\in\bar{\Theta}} \frac{q_{\theta'}}{q_\theta} b_{\theta'} > -b_\theta | \theta\right) \\
&\leq \sum_{\theta'\in\bar{\Theta}} \mathbb{P}\left(\frac{q_{\theta'}}{q_\theta} > \frac{-b_\theta}{|\bar{\Theta}| b_{\theta'}}\right) \\
&= \sum_{\theta'\in\bar{\Theta}} \mathbb{P}\left(\log\left(\frac{q_{\theta'}}{q_\theta}\right) > \log\left(\frac{-b_\theta}{|\bar{\Theta}| b_{\theta'}}\right)\right)
\end{aligned}$$

where the second line follows from the fact that at least one of the scaled likelihood-ratios must exceed $-b_\theta/\bar{\Theta}$ in order for the sum to exceed $-b_\theta$.

What remains then is a sum of probabilities only in terms of individual LLRs, and thus the usual large-deviation approximation holds. Thus we have that the mistake probability in state θ is $\mathcal{O}(\rho(\theta h(\theta))^n n^{-1/2})$.

Finally, we still have that $\mathrm{FIG}(n) \geq (p_\theta + p_{\theta'})\mathrm{FIG}^*_{\theta\theta'}(n)$ because the RHS is the FIG after additionally observing a signal that perfectly reveals the state unless the truth is either θ or θ′. Letting θθ′ be among the least-precision dichotomies, then we can write

$$\mathrm{B}_1\frac{\rho(\theta\theta')^n}{\sqrt{n}}(1+\mathcal{O}(n^{-1})) \leq \mathrm{FIG}(n) \leq \mathrm{B}_2\frac{\rho(\theta\theta')^n}{\sqrt{n}}(1+\mathcal{O}(n^{-1}))$$

which, by squeezing, implies that

$$-\log(\mathrm{FIG}(n)) = n\beta(\theta\theta')(1+\mathcal{O}(\log(n)n^{-1}))$$

as claimed. □

A.3 Proof of Prop. 5

For the sake of keeping notation manageable, I prove the result for a two source case. The logic easily generalizes by replacing derivatives with gradients where appropriate.

Let $N = n_1 + n_2$. Then consider a $k(N) = o(N)$ loss from $\mathscr{E}_1$. Define $d\omega$ as the change in ω from substituting $k(N)$ samples of $\mathscr{E}_1$ with DRS $\times$ $k(N)$ samples from $\mathscr{E}_2$. Note that $d\omega = \mathcal{O}(k(N)/N)$. Further, for the sake of notation, define the *additive* error of the approximation $-\log(\mathrm{FIG}(n_1, n_2))$ as $\epsilon(n_1, n_2)$. That is, by taking the log of the FIG approximation from Prop. 3, we have

$$\epsilon(n_1, n_2) = \log(B_\omega) - \frac{1}{2}\log(n_1 + n_2) + \mathcal{O}\left(\frac{1}{n_1 + n_2}\right)$$

Further recall that B_ω only depends on ω through a differentiable function of τ_ω and is thus itself differentiable in ω.

By assumption, the least precision dichotomy is unique at ω, so for notational compactness, all precisions are for the least-precision dichotomy at ω.

By definition, $k(N)\mathrm{DRS}$ is the smallest number of samples from $\mathscr{E}_2$, that just compensates for the loss of $k(N)$ samples of $\mathscr{E}_1$ so we have

$$\begin{aligned} n_1\beta_{1\omega} + n_2\beta_{2\omega} + &\epsilon(n_1, n_2) - \epsilon(n_1 - k(N), n_2 + k(N)\mathrm{DRS}) \\ &\leq (n_1 - k(N))\beta_{1(\omega+d\omega)} + (n_2 + k(N)\mathrm{DRS})\beta_{2(\omega+d\omega)} \end{aligned} \tag{13}$$

Now we can apply Taylor's theorem to write

$$\begin{aligned} &(n_1 - k(N))\beta_{1(\omega+d\omega)} + (n_2 + k(N)\mathrm{DRS})\beta_{2(\omega+d\omega)} \\ &\quad = (n_1 - k(N))\beta_{1\omega} + (n_2 + k(N)\mathrm{DRS})\beta_{2\omega} + d\omega(n_1\frac{d}{d\omega}\beta_{1\omega} + n_2\frac{d}{d\omega}\beta_{2\omega}) + \mathcal{O}(k(N)/N) \end{aligned}$$

Now notice that by the definition of β_j we have $(d/d\omega)\beta_{j\omega} = -(d/d\omega)\rho_{j\omega}/\rho_{j\omega}$. This plus the FOC for ρ_ω implies

$$n_1\frac{d}{d\omega}\beta_{1\omega} + n_2\frac{d}{d\omega}\beta_{2\omega} = 0$$

Now we can apply Taylor's theorem again to approximate the change in ϵ:

$$\epsilon(n_1, n_2) - \epsilon(n_1 - k(N), n_2 + k(N)\mathrm{DRS}) = \mathcal{O}\left(\frac{k(N)}{N}\right)$$

Finally, substituting this all back into (13) and solving for DRS gives

$$\text{DRS} \geq \frac{\beta_{1\omega}}{\beta_{2\omega}} + \mathcal{O}\left(\frac{1}{N}\right)$$

By repeating the process for the substitution of $k(N)$ samples of $\mathscr{E}_1$ for $k(N)\text{DRS}-1$ samples of $\mathscr{E}_2$ we get

$$\text{DRS} \leq \frac{\beta_{1\omega}}{\beta_{2\omega}} + \frac{1}{k(N)} + \mathcal{O}\left(\frac{1}{N}\right)$$

Taking the limit in N and squeezing completes the proof. $\square$

A.4 Proof sketch for the Lugannani & Rice (1980) saddlepoint approximation

Lugannani & Rice (1980) approximate the CDF by applying a saddlepoint approximation to a particular characteristic function inversion formula. The full formal proof relies a fair bit on some advanced complex analysis, and so I provide a sketch of the approach that attempts to minimize reliance on complex analysis beyond the standard Cauchy-Riemann conditions for analytic functions. The skeptical reader should see Lugannani & Rice (1980) for a formal treatment.

Recall that the characteristic function of a random variable, X is $\phi(t) = \mathbb{E}(e^{itX})$. Equivalently we can think of the characteristic function as a rotation of the MGF in the complex plane, $\phi(t) = M(it)$, so it shares with the MGF the property that the characteristic function of an independent sum of random variables is the product of each characteristic function.

The following lemma establishes a useful property about the characteristic function in the context of LLR distributions:

Lemma (Differentiability of the characteristic function). *Suppose ϕ is the characteristic function for a LLR distribution whose MGF is finite in an open interval containing the origin. Then if $\Im(t) \in (-1, 0)$, then ϕ is analytic at t, and thus is infinitely differentiable there.*[22]

Proof. Assuming ϕ is differentiable, we can use Leibniz rule to write

$$\phi^{(k)}(t) = \int \left(\log \frac{dF(r \mid \theta')}{dF(r \mid \theta)}\right)^k dF(r \mid \theta')^{it} dF(r \mid \theta)^{1-it}$$

Thus, so long as this integral is finite, we have differentiability. For t such that $\Im(t) \in (-1, 0)$ this follows immediately from dominated convergence and the fact that the LLR distribution has all its moments (because the MGF is defined on an open interval containing the origin). ▽

In particular, the previous lemma also implies that in a sufficiently small interval around any t in the analytic strip, ϕ is well approximated by its Taylor series—a tool that will prove useful later.

Now, because characteristic functions uniquely define their distribution, we can invert it to get the distribution function. Typically, one would use the standard Fourier inversion formula to get the density of the distribution; however, we don't generally have a density, so we instead must use a less commonly known inversion formula:

22 $\Re(t)$ and $\Im(t)$ respectively denote the real and imaginary parts of t

Lemma (Characteristic function inversion). *Let* F *be the CDF for some distributon* Y *on* $\mathbb{R}$, *with characteristic function,* ϕ. *If* ξ *is a continuity point of* F *then*

$$1 - F(\xi) = \frac{1}{2\pi} \lim_{L\to\infty} \int_{-L-ci}^{L-ci} \frac{e^{-it\xi}\phi(t)}{it} dt \tag{14}$$

for arbitrary real constant $c > 0$ *such that the MGF of* F *is differentiable at* c.

Lugannani & Rice (1980) state this formula without proof or reference. For the sake of completeness, I prove it in Appendix A.5.

Now, recall that the distribution in question is an i.i.d. sum of n samples, and we want to know when the sample average exceeds ξ (so when the sum exceeds $n\xi$). Thus we can write the characteristic function as $\phi(t) = M(it)^n = e^{nK(it)}$. Using this and changing variables to $T = it$, rewrite the inversion formula as

$$1 - F(\xi) = \frac{1}{2\pi} \lim_{L\to\infty} \int_{-iL+c}^{iL+c} \frac{e^{n(K(T)-T\xi)}}{iT} dT$$

Now, let $c = \tau(\xi)$ where $\tau(\xi)$ is the (real) argmin of $K(t) - t\xi$, so the path of integration crosses the real line perpindicular to the minimum (along the real line) of $K(t)-t\xi$. Recall that K is analytic, and thus by the Cauchy-Riemann equations, $\tau(\xi)$ is a saddlepoint. So, although $\tau(\xi)$ minimizes $K(t) - t\xi$ when traveling along the real axis, but it *maximizes* it when traveling along the perpindicular complex axis.

Notice, the integrand in the above equation is complex, but the right-hand side is real, so it will be useful to separate the integrand into its complex part (which must integrate to zero) and its real part. First do another change of variables:

$$1 - F(\xi) = \frac{1}{2\pi} \lim_{L\to\infty} \int_{-L}^{L} \frac{e^{n(K(\tau(\xi)+ix)-(\tau(\xi)+ix)\xi)}}{\tau(\xi) + ix} dx$$

Then split into real and complex parts:

$$\begin{aligned}
1 - F(\xi) &= \frac{1}{2\pi} \lim_{L\to\infty} \int_{-L}^{L} \frac{e^{n\Re(g(\tau(\xi)+ix))}(\cos(n\Im(g(\tau(\xi) + ix))) + i\sin(n\Im(g(\tau(\xi) + ix))))}{\tau(\xi) + ix} dx \\
&= \frac{1}{2\pi} \lim_{L\to\infty} \int_{-L}^{L} \frac{e^{n\Re(g(\tau(\xi)+ix))}(\tau(\xi)\cos(n\Im(g(\tau(\xi) + ix))) + x\sin(n\Im(g(\tau(\xi) + ix))))}{\tau(\xi)^2 - x^2} dx \\
&\equiv \frac{1}{2\pi} \lim_{L\to\infty} \int_{-L}^{L} e^{n\hat{g}(x)} h(x) dx
\end{aligned}$$

where $g(T) \equiv K(T) - T\xi$. The first line follows from Euler's formula, the second follows from multiplying numerator and denominator by $\tau(\xi) - ix$ and then canceling any imaginary terms since the complex part of the integrand must integrate to zero, and the third follows from appropriately defining $\hat{g}$ and h to simplify notation.

The integrand of the last line is now mapping reals into reals and can thus be handled with standard real analysis tools. In particular, we can approximate the integral using a classical Laplace approximation. Roughly, the idea is that $\hat{g}(x)$ takes it's max at $x = 0$, and on account of the exponentiation, the integrand will get almost all its mass from a vanishing interval around that max as n gets large.

Now write

$$\int_{-\infty}^{\infty} e^{n\hat{g}(x)} h(x) \mathrm{d}x = e^{n\hat{g}(0)} \int_{-\infty}^{\infty} e^{n(\hat{g}(x) - \hat{g}(0))} h(x) \mathrm{d}x$$

Using a second-order Taylor approximation, we can then write

$$e^{n\hat{g}(0)} \int_{-\infty}^{\infty} e^{n\hat{g}(x) - \hat{g}(0)} h(x) \mathrm{d}x \approx e^{n\hat{g}(0)} \int_{-\infty}^{\infty} e^{\frac{1}{2} n\hat{g}''(0) x^2} h(x) \mathrm{d}x$$

The LHS of the above now is a familiar form: a Gaussian integral. Thus we can use the Taylor approximation for the expectation of $h(x)$ when x is Gaussian with small variance

$$\begin{aligned} e^{n\hat{g}(0)} \int_{-\infty}^{\infty} e^{\frac{1}{2} n\hat{g}''(0) x^2} h(x) \mathrm{d}x &\approx e^{n\hat{g}(0)} \sqrt{\frac{2\pi}{-n\hat{g}''(0)}} (h(0) - (n\hat{g}''(0))^{-1} h''(0)) \\ &= e^{n\hat{g}(0)} \sqrt{\frac{2\pi}{-n\hat{g}''(0)}} h(0)(1 + \mathcal{O}(n^{-1})) \end{aligned}$$

Now, plugging things in and using the Cauchy-Riemann equations, we have

$$\begin{aligned} \hat{g}(0) &= \mathrm{K}(\tau(\xi)) - \tau(\xi)\xi \\ -\hat{g}''(0) &= \mathrm{K}''(\tau(\xi)) \\ h(0) &= \frac{1}{\tau(\xi)} \end{aligned}$$

Plugging all this into our original equations, we have

$$1 - \mathrm{F}(\xi) = \frac{e^{n(\mathrm{K}(\tau(\xi)) - \tau(\xi)\xi)}}{\tau(\xi)\sqrt{2\pi n \mathrm{K}''(\tau(\xi))}} \left(1 + \mathcal{O}\left(\frac{1}{n}\right)\right)$$

as claimed. $\square$

A.5 Proof of the characteristic function inversion formula

First, I show that

$$\lim_{L\to\infty}\int_{-L}^{L}\frac{e^{-itA}}{it}dt = 2\pi\mathbb{I}_{A>0}$$

provided that the path of integration sweeps *below* the origin to avoid the singularity there. To see this, consider an integration path that goes from $-L$ to $-\epsilon$, then below the origin along a half circle of radius ϵ, and finally from ϵ to L:

$$\int_{-L}^{L}\frac{e^{-itA}}{it}dt = \left[\int_{-L}^{-\epsilon}+\int_{\epsilon}^{L}\right]\frac{e^{-itA}}{it}dt + \int_{\pi}^{2\pi}\exp(i\epsilon e^{i\theta}A)d\theta$$

where the second term follows from a change to radial coordinates along the path around origin. Following CITATION NEEDED, the second term equals $\pi + 2\text{Si}(\epsilon A)$ where $\text{Si}(x) = \int_0^x \sin(t)t^{-1}dt$ is the sine integral. Note that $\text{Si}(x)$ vanishes as x goes to zero.

Turning to the first term, we can use Euler's formula—plus the fact that $\sin(t)/t$ and $\cos(t)/t$ are even and odd functions respectively—to write

$$\left[\int_{-L}^{-\epsilon}+\int_{\epsilon}^{L}\right]\frac{e^{-itA}}{it}dt = 2\int_{\epsilon}^{L}\frac{\sin(At)}{t}dt$$

Now sending ϵ to zero, we can write

$$\int_{-L}^{L}\frac{e^{-itA}}{it}dt = 2\int_{0}^{L}\frac{\sin(At)}{t}dt + \pi$$

Finally, sending L to infinity, we have (see Jeffreys & Jeffreys 1956, p. 471, eq. 18)

$$\lim_{L\to\infty}\int_{-L}^{L}\frac{e^{-itA}}{it}dt = \pi(1+sgn(A)) = 2\pi\mathbb{I}_{A>0}$$

Applying Fubini's theorem completes the proof of the inversion formula for $c = 0$ (with a detour around the origin).

The $c > 0$ case follows from Cauchy's integral theorem: provided that the integrand is analytic in the region between two curves with the same termina, the integrals across those curves are equal. First note ϕ is analytic along the entire strip with $\{\Re(t) \in [0, c]\}$ because $M(t)$ is differentiable for $t \in [0, c]$ and because $|e^t| = e^{\Re(t)}$. So instead of integrating straight from $-L$ to L, we can integrate across a staple-shaped path that first goes from $-L$ to $-L - ci$, then across to $L - ci$, then up to L. Since the norm of the numerator of the integrand along one of the two tines of the path is at most $e^{|c|}$, we must have that the integrand along those tines is vanishing as $L \to \infty$. $\square$

BIBLIOGRAPHY

Azrieli, Y. (2014), 'Comment on "The Law of Large Demand for Information"', *Econometrica* **82**(1), 415–423.

Blackwell, D. (1951), Comparison of experiments, *in* 'Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability', University of California Press, pp. 93–102.

Butler, R. W. (2017), *Saddlepoint Approximations with Applications*, Cambridge University Press.

Chade, H. & Schlee, E. E. (2002), 'Another look at the Radner-Stiglitz nonconcavity in the value of information', *Journal of Economic Theory* **107**(2), 421–452.

Chernoff, H. (1952), 'A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations', *The Annals of Mathematical Statistics* **23**(4), 493–507.

Cramér, H. (1938), 'Sur un nouvaeu théorème-limite de la théorie des probabilités', *Actualités Scientifiques et Industrielles* (736).

Daniels, H. E. (1954), 'Saddlepoint approximations in statistics', *The Annals of Mathematical Statistics* **25**(4), 631–650.

Daniels, H. E. (1987), 'Tail probability approximations', *International Statistical Review / Revue Internationale de Statistique* **55**(1), 37.

Dembo, A. & Zeitouni, O. (1998), *Large Deviations Techniques and Applications*, 2nd edn, Springer Berlin Heidelberg.

Jeffreys, H. & Jeffreys, B. (1956), *Methods of Mathematical Physics*, 3rd edn, Cambridge University Press.

Lugannani, R. & Rice, S. (1980), 'Saddle point approximation for the distribution of the sum of independent random variables', *Advances in Applied Probability* **12**(2), 475–490.

Montgomery, D. C. (2012), *Design and Analysis of Experiments*, 8th edn, John Wiley & Sons.

Moscarini, G. & Smith, L. (2002), 'The law of large demand for information', *Econometrica* **70**(6), 2351–2366.

Pomatto, L., Strack, P. & Tamuz, O. (2018), 'The cost of information'.

Radner, R. & Stiglitz, J. E. (1984), 'A nonconcavity in the value of information'.

Torgersen, E. (1991), *Comparison of Statistical Experiments*, Cambridge University Press.