Choosing Sample Sizes∗

Sanket Patil and Yuval Salant†

April 20, 2021

Abstract

How much data about an unknown parameter would a designer provide to a decision maker (DM) in order to convince the DM that the parameter value is sufficiently high? We study this question for DMs who are unbiased or Bayesian statisticians and for data which are Bernoulli experiments governed by the parameter value. We establish that in many environments the designer’s optimal sample size is the largest one satisfying that one or more — or a simple majority of — favorable data realizations would convince the DM that the parameter value is sufficiently high.

Keywords: Sampling, statistical inference, sample size, experimental design.

JEL classification: D81, D83, C90.

∗We thank seminar participants at Columbia University, the Econometric Society World Congress, the joint economic theory workshop to Hebrew University and Tel-Aviv University, Northwestern University, University of Michigan, Washington University in St. Louis, and Western Economic Association International Conference for helpful comments and discussions.

†Both authors are at the Kellogg School of Management, Northwestern University. Emails: sanket.patil@kellogg.northwestern.edu and y-salant@kellogg.northwestern.edu.
1 Introduction

Decision makers (DMs) often rely on data and statistical inference to estimate payoff-relevant parameters and make choices. For example, buyers experiment with products before making purchase decisions in order to estimate their value. Likewise, politicians use polls and public opinion surveys to inform their decision making. In both cases, a strategic designer may control the size of the data obtained by the DM but not necessarily its distribution. In the context of product experimentation, it is the seller who decides how much experimentation to allow potential buyers, and the buyers conduct the experiments themselves by interacting with the product. And in the context of polls and surveys, some strategic party (for example, lobbyist or think tank) often decides how extensive a poll to commission, and exogenous factors such as the politician’s own interests dictate the question and the subject population. The goal of this paper is to study the designer’s choice of the data sizes in such settings.

We consider a designer who decides how much data about a payoff-relevant parameter a DM can obtain. The DM is a statistician who uses the data realizations, that is, the sample, and statistical inference to estimate the parameter value. The DM takes an action, which is desirable to the designer, if the estimated value is larger than the DM’s outside option.

The data are Bernoulli experiments governed by the parameter value. For example, in surveys, the parameter may be the proportion of individuals supporting a particular issue and each Bernoulli experiment may correspond to the opinion of a single survey respondent. A similar setting is meetings where the DM relies on the opinions of participants to arrive at a decision. Bernoulli experiments also fit product experimentation, for example, when the product is a platform with many products, the parameter is the fraction of products that fit the buyer’s taste, and each experiment corresponds to the DM experimenting with a single product. Of course, there are other settings where an experiment can convey more information about the parameter value.
As for statistical inference, there are roughly speaking two classes of inference procedures: those that rely only on the sample (frequentist inference), and those that start with a prior belief on the parameter value and Bayes-update it based on the sample (Bayesian inference). The advantages of the frequentist approach are that it is prior-free, does not require knowledge of conditional probabilities, and uses only objective data.\footnote{The statistics literature has debated the merits of the two approaches for over a century. Efron (2005) is a nice recent discussion.} As such, the frequentist approach fits environments in which the DM is less experienced, less knowledgeable about fundamentals, or does not want prior beliefs to influence decision making.

Among frequentist inference procedures, an important subclass is the class of unbiased inference procedures whereby the expected value of an estimate (which we allow to be a distribution over possible parameter values) is identical to the proportion of successes in the sample. Leading examples include maximum likelihood estimation, maximum entropy estimation, and Beta estimation.

We first consider environments in which the DM uses an unbiased inference procedure. The focus on unbiased inference enables us to identify the forces at play that do not relate to prior beliefs the DM may have. It creates, however, a potential tension because the designer has a prior belief about the parameter value and the DM does not make strategic inferences about this belief from the sample size choice. This tension is alleviated throughout the analysis in two ways. First, we extend our results to cases in which the designer’s sample size choice does not reveal information about the designer’s prior. Second, we establish analogous results for environments with Bayesian inference in which the DM and the designer have the same prior.

Our first result is about cases in which the designer’s prior is decreasing. We establish that the optimal sample size is the largest one satisfying that even a single favorable realization (henceforth, success) would trigger the DM to take the action. In other words, any positive integer $n$ is the uniquely optimal sample size when the DM’s outside option is between $1/(n+1)$ and $1/n$. 
This result has two implications. First, as the DM’s outside option becomes more attractive, so that the DM’s tendency to take the action decreases for any sample size, the designer’s optimal sample size decreases. Second, an outsider who observes the DM’s choice behavior as a function of data realizations may interpret this behavior as exhibiting “strong” preference for taking the action. This is because the DM chooses to take the action unless there is unanimity against doing so. To be sure, the DM is an unbiased statistician whose preference for taking the action is neither strong nor weak. But the designer’s choices result in behavior that exhibits what may seem like a strong preference for taking the action.

Our second result is about symmetric or increasing and concave priors when the value of the DM’s outside option is not small (formally, larger than 1/2). For symmetric priors, the optimal sample sizes are all odd sample sizes satisfying that a simple majority of successes would trigger the DM to take the action. And for increasing and concave priors, the largest among these sample sizes is uniquely optimal. An important implication of this result is that the DM’s choice behavior as a function of data realizations endogenously follows the simple-majority rule.

These two results extend to settings in which the designer is concerned that the DM may be able to strategically infer something about the designer’s prior from the sample size choice and use this information when making choices. We establish that when the value of the DM’s outside option is not small, a designer with such a concern would choose some odd sample size satisfying that a simple majority of successes would trigger the DM to take the action, thus shutting down the channel for strategic inference.

We also consider Bayesian environments in which the designer and the DM share a Beta(α, β) prior and the DM uses Bayesian inference. Beta(α, β) priors are a common way of modeling prior knowledge in environments with Bernoulli experimentation. These priors have many different shapes as α and β change, and they facilitate comparative statics on how the prior strength α + β affects the designer’s optimal
The focus of our analysis is on the intermediate case in which the value of the outside option is larger than the expected value of the prior but is not too large, so that the candidates for optimality for unbiased and Bayesian inference are identical. We establish that “compressed” versions of the results about unbiased inference carry over to Bayesian inference. Figure 1 provides a graphical illustration for symmetric priors. The optimal sample sizes for unbiased inference are also optimal for Bayesian inference from a $Beta(\alpha, \alpha)$ prior although the range of the outside option (denoted...
by \( t \) in the figure) for which they are optimal shifts toward the prior mean and shrinks. As the strength \( 2\alpha \) of the prior diminishes and the DM puts increasingly larger weight on the sample, the corresponding range of the outside option for Bayesian inference converges to that of unbiased inference. The reason for this convergence is that as the DM puts increasingly large weight on the sample relative to the prior, the DM becomes closer and closer to being an unbiased statistician who relies only on the sample.

Our paper is related to several literatures. The first is the literature that incorporates sampling (Osborne and Rubinstein (1998, 2003)) and statistical inference (Salant and Cherry (2020)) into games. In Osborne and Rubinstein (1998)’s \( S(k) \)-equilibrium, players who do not know the mapping from own actions to payoffs sample the payoff of each action \( k \) times and choose the action with the highest sampled payoff. Osborne and Rubinstein (2003) is a subsequent contribution in which players sample other players’ actions instead of own payoffs and best respond to sample averages. Salant and Cherry (2020) enrich this framework by considering players who use statistical inference to estimate other players’ actions and best respond to their estimates. Sethi (2000, 2020) and Mantilla, Sethi and Cárdenas (2020) study the dynamic stability properties of \( S(1) \)-equilibria, and Spiegler (2006\textit{a,b}) studies competition between firms that face consumers who sample a payoff-relevant parameter once. This literature treats players’ sample size as a primitive of their decision making procedure or a component of the solution concept. And with the exception of Salant and Cherry (2020), the literature solves models with players who obtain very small samples. The focus of the current paper is on an orthogonal question. We treat the sample size as a design parameter and solve for the optimal sample size.

As such, our paper also contributes to the small literature on the design of experiments in strategic environments. Di Tillio, Ottaviani and Sørensen (2021) consider a setting in which the DM decides (1) the minimal sample size the designer will collect on behalf of the DM, and (2) whether to allow the designer to collect a larger
sample at a cost and provide the DM with a non-random selection from it. We consider a complementary setting in which the designer decides the sample size without any constraints imposed by the DM, and the DM obtains all the data realizations. Other contributions to this literature include Chassang, Padró i Miquel and Snowberg (2012) who propose selective trials to improve the external validity of experiments with strategic participants and Banerjee et al. (2020) who provide a decision-theoretic foundation for randomization in controlled trials.

A third related literature is the literature on sender-optimal persuasion (Kamenica and Gentzkow (2011)). We depart from this literature in two ways. First, the set of tools available to the designer in our setup is limited. Rather than choosing the data-generating process, the designer only chooses how much data the DM obtains. As discussed above, this assumption seems plausible when the experiments are conducted by the DM (product experimentation) or a third party, or when the experimental question and subject population cannot be easily changed (polls and surveys). Second, a considerable part of our analysis focuses on frequentist inference and on comparing frequentist inference to Bayesian inference. In this respect, our paper is related to contributions on persuasion with non-Bayesian DMs who, for example, have limited ability to find persuasive messages (Glazer and Rubinstein (2012)), fail to account for the correlation between different information sources (Levy, de Barreda and Razin (2018)), focus only on a subset of a multi-dimensional signal (Eliaz, Spiegler and Thysen (2021)), or reject their prior in favor of a new one when obtaining evidence that is inconsistent with the prior (Galperti (2019)).

We proceed as follows. Section 2 presents the model. Section 3 analyzes the optimal sample size for unbiased inference. Section 4 analyzes the optimal sample size for Bayesian inference and compares the results to those of Section 3. Section 5 concludes. Appendix A contains proofs that do not appear in the main text. Appendix B contains lemmas that are used in the proofs.

---

2For receiver-optimal persuasion, see for example Glazer and Rubinstein (2004).
2 Model

A decision maker (DM) has to decide whether to take an action or keep the status quo. The value $t \in (0, 1]$ of keeping the status quo is known to the DM and the value $q \in [0, 1]$ of taking the action is not. To make a decision, the DM estimates $q$ using data and statistical inference, and takes the action if the estimated value of $q$ is weakly larger than $t$.

The data are independent Bernoulli experiments with success probability $q$. A successful experimental realization, or simply a success, is interpreted as a data point in favor of taking the action.

The size of the data $n \in \mathbb{N} \cup \{\infty\}$ is decided by a designer who obtains a payoff of 1 if the DM takes the action and 0 otherwise. That is, the designer wishes the DM to take the action regardless of the values of $t$ and $q$. The designer knows $t$ and has a prior $f$ on $q$.\footnote{Throughout the analysis, we assume the prior $f$ has some positive mass on $(0, 1)$.} To avoid technical issues, we assume that the designer can choose any data size or $\infty$, which is interpreted as fully revealing the value of $q$. We consider lower and upper bounds on the data size in Section 3.4.

After the designer decides the data size, the Bernoulli experiments are carried out by the DM or a third party, and the DM obtains their realizations. The DM’s sample is the pair $(n, k)$ where $n$ is the number of experiments and $k$ is the number of successes. We will refer to $n$ as the sample size and to $k/n$ as the sample mean.

2.1 Statistical inference

An inference procedure describes how the DM makes inferences from samples.

Definition 1. (Salant and Cherry (2020)) An inference procedure $G = \{G_{n,k}\}$ assigns a cumulative distribution function $G_{n,k}$, called an estimate, to every sample $(n, k)$ such that:
(i) the estimate $G_{n,k'}$ first-order stochastically dominates the estimate $G_{n,k}$ when $k' > k$, and

(ii) the estimate $G_{n,n}$ strictly first-order stochastically dominates the estimate $G_{n,0}$.\(^4\)

An inference procedure is the analogue of an estimator in the statistics literature. It can be used to describe many forms of statistical inference. A focal example is Bayesian inference.

**Example 1** (Bayesian Inference). The DM has a non-degenerate prior on $q$ and uses Bayes rule to update this prior based on the sample. A common modeling assumption in Bayesian environments with Bernoulli experimentation is that the prior is a $\text{Beta}(\alpha, \beta)$ distribution with a density function

$$ g(\alpha, \beta) = \frac{q^{\alpha-1}(1-q)^{\beta-1}}{\int_0^1 q^{\alpha-1}(1-q)^{\beta-1} dq}. $$

The two parameters $\alpha$ and $\beta$ are positive real numbers interpreted as the ex-ante “number of successes” and “number of failures” respectively. The strength of the prior is the ex-ante “total number of experiments” $\alpha + \beta$, and its bias $\frac{\alpha}{\alpha + \beta}$ is the prior mean, or the proportion of successes in the prior.

The DM’s posterior after obtaining the sample $(n, k)$ is the $\text{Beta}(\alpha + k, \beta + (n - k))$ distribution because Beta is a conjugate prior for the Bernoulli distribution. The first parameter in the posterior $\alpha + k$ is the sum of successes in the prior and sample, and the second is the corresponding sum of failures. The mean of the posterior

$$ \frac{\alpha + \beta}{\alpha + \beta + n} \cdot \frac{\alpha}{\alpha + \beta} + \frac{n}{\alpha + \beta + n} \cdot \frac{k}{n} $$

is a weighted average of the prior- and sample-means where the weights are the prior-

\(^4\)This definition is weaker than in Salant and Cherry (2020) as we require strict dominance only for $G_{n,n}$ and $G_{n,0}$.
and sample-strengths relative to the total number of experiments $\alpha + \beta + n$. As we will see in Section 4, Beta($\alpha, \beta$) priors facilitate comparative statics on how changes in the prior’s strength and bias affect the optimal sample size.

In addition to Bayesian inference, an inference procedure can also be used to model frequentist inference procedures that do not rely on a prior belief. Here are a few examples.

**Example 2** (Maximum Likelihood Estimation (MLE)). The DM calculates the most likely parameter $q$ to have generated the sample. It is easy to verify that this parameter is the sample mean. Thus, the resulting estimate is

$$G_{n,k}(q) = \begin{cases} 0 & \text{if } q < k/n \\ 1 & \text{if } q \geq k/n \end{cases}.$$  

**Example 3** (Beta Estimation). The DM wishes to conduct Bayesian updating relying as little as possible on prior beliefs. The DM starts with Haldane’s “prior” (Haldane (1932)), which is the limit of the Beta($\epsilon, \epsilon$) distribution as $\epsilon \to 0$. Note that the limit is not a proper prior. The DM’s posterior belief after obtaining the sample $(n, k)$ is the limit of the corresponding posteriors Beta($\epsilon + k, \epsilon + (n - k)$), i.e., it is the Beta($k, n - k$) distribution. In other words, the DM puts no weight on the prior and bases the estimate only on the sample.$^5$

**Example 4** (Maximum Entropy). The DM follows the Principle of Maximum Entropy. From among all distributions with expected value equal to the sample mean, the DM searches for the one with the maximal uncertainty in terms of entropy. By Conrad (2004), when the sample mean is in $(0,1)$, this distribution is a truncated exponential distribution with density $g_{n,k}(q) = C(n, k)e^{\alpha(n,k)q}$. The values $C(n, k)$ and $\alpha(n,k)$ are determined uniquely by the constraints that (1) $g_{n,k}$ is a density function

\footnote{When the sample includes only failures or successes, the DM concentrates the estimate on 0 or 1 respectively.}
(i.e., \(\int_0^1 g_{n,k}(q)\,dq = 1\)), and (2) the expected value of the estimate is equal to the sample mean (i.e., \(\int_0^1 qg_{n,k}(q)\,dq = k/n\)). For example, when \(k/n = 1/2\), we have that \(C(n,k) = 1\) and \(\alpha(n,k) = 0\), so the resulting estimate is the uniform distribution over \([0, 1]\).\(^6\)

**Example 5** (Dogmatic Views). The DM believes \(q\) is distributed either according to the CDF \(F_0\) or \(F_1\) that strictly first-order stochastically dominates \(F_0\). He uses the sample to decide which distribution should be used in decision making. If \(k(n)\) or more realizations are successes, he uses \(F_1\). Otherwise he uses \(F_0\).

Frequentist inference procedures can be classified according to whether they are unbiased.

**Definition 2.** An inference procedure \(G\) is *unbiased* if the expected value \(\int_0^1 q\,dG_{n,k}\) of any estimate \(G_{n,k}\) is equal to the sample mean \(k/n\).

The MLE, Beta Estimation, and Maximum Entropy procedures are unbiased. The dogmatic views procedure is not. Bayesian inference is also not unbiased because the expected value of an estimate depends on both the prior and sample means whereas unbiasedness requires that the expected value depends only on the sample.

Following the estimation of \(q\), the DM uses the estimate to calculate the expected value of \(q\), and takes the action if this value is weakly larger than \(t\).

### 2.2 The designer’s objective

Fix the DM’s inference procedure \(G\), a sample size \(n\), and the values of \(t\) and \(q\). Let

\[
b(n, j, q) = \binom{n}{j} q^j (1-q)^{n-j}
\]

denote the probability of obtaining \(j\) successes in \(n\) Bernoulli experiments governed by \(q\). By the first-order stochastic dominance property of an inference procedure, if

\(^6\)When the sample mean is 0 or 1, the unique distribution with expected value equal to the sample mean puts a unit mass on the sample mean.
the DM takes the action after obtaining \( j \) successes in \( n \) Bernoulli experiments, then
the DM also takes the action after obtaining more than \( j \) successes in \( n \) experiments. Thus the probability the DM takes the action is

\[
P(n, k, q) = \sum_{j=k}^{n} b(n, j, q) = \sum_{j=k}^{n} \binom{n}{j} q^j (1 - q)^{n-j}
\]

where \( k = k(n, G, t) \) denotes the smallest number of successes after which the DM takes the action.\(^7\) The derivative of \( P(n, k, q) \) with respect to \( q \) is

\[
P'(n, k, q) = k \binom{n}{k} q^{k-1} (1 - q)^{(n-k)}.
\]

The objective of the designer is to maximize the expected value, according to \( f \), of the probability of taking the action, i.e. solve

\[
\arg \max_{n \in \mathbb{N} \cup \{ \infty \}} \int_0^1 P(n, k(n, G, t), q) f(q) dq.
\]

For unbiased inference procedures, the focus of our analysis in Section 3, the smallest number of successes that trigger the DM to take the action is \( k(n, t) = \lceil tn \rceil \). Because this number does not depend on the specifics of the inference procedure, we can write the probability of taking the action as:

\[
P(n, \lceil tn \rceil, q) = \sum_{j=\lceil tn \rceil}^{n} b(n, j, q)
\]

for any unbiased inference procedure.

\(^7\)It is possible that there are sample sizes \( n \) for which such \( k(n, G, t) \leq n \) does not exist. Such sample sizes are never optimal, and so we ignore them.
3 Unbiased inference

This section studies the designer’s optimal sample size for unbiased inference procedures. Section 3.1 presents two useful properties of $P(n, k, q)$ and derives a general lower bound on the optimal sample size. Section 3.2 analyzes the optimal sample size when the designer’s prior is degenerate, i.e., assigns probability 1 to a particular $q$. The objective of analyzing degenerate priors is to highlight the forces that would determine the optimal sample size when the designer’s prior is non-degenerate. Section 3.3 studies the designer’s optimal sample size for non-degenerate priors that are either monotone or symmetric. Section 3.4 incorporates into the analysis lower and upper bounds on the feasible sample sizes.

3.1 Two properties of the probability $P(n, k, q)$

There are two properties of the probability $P(n, k, q)$ that are frequently used in the analysis. The first is sample-size monotonicity, which says that for a fixed number of successes, the probability of taking the action increases in the sample size.

Property 1 (Sample-size monotonicity). The probability $P(n, k, q)$ increases in $n$.

Proof. Fix $k \leq n$ and $q$. Then, $P(n+1, k, q) = P(n, k, q) + qb(n, k-1, q) > P(n, k, q)$ where the left equality holds because $b(m, l, q) = qb(m-1, l-1, q) + (1-q)b(m-1, l, q)$ for any two positive integers $m$ and $l \leq m$.

An immediate implication of sample-size monotonicity is that:

Observation 1. For any $t$ and any prior on $q$, the optimal sample size is weakly larger than the largest sample size satisfying that even a single success would trigger the DM to take the action.

In other words, for any integer $n$ and any prior on $q$, the optimal sample size is weakly larger than $n$ when the value $t$ of the status quo is in the interval $(0, 1/n]$.
As an illustration, consider the case in which \( t < 1/2 \). No matter what prior the designer has, it is never optimal to provide the least amount of information about \( q \), i.e., a single Bernoulli experiment, to the DM. This is because a DM who knows more about \( q \) in the sense of obtaining a two-experiment sample takes the action with probability \( P(2, \lceil 2t \rceil, q) = q^2 + 2q(1 - q) \) which is larger, for any \( q \), than the probability \( P(1, \lceil t \rceil, q) = q \) of taking the action for a single-experiment sample.

The second property is single crossing. It says that for a smaller sample size with a smaller number of successes the probability \( P(n, k, q) \) dominates the corresponding probability for a larger sample size with a larger number of successes if \( q \) is below some cutoff, and that a reverse ranking holds above the cutoff.

**Property 2** (Single crossing). Consider two pairs \((n', k')\) and \((n, k)\) such that \( n' < n \) and \( k' < k \). Then, there exists a cutoff probability \( q^* > 0 \) such that \( P(n', k', q) > P(n, k, q) \) for \( q \in (0, q^*) \). If \( q^* < 1 \) then \( P(n, k, q) > P(n', k', q) \) for \( q \in (q^*, 1) \).

In particular, a smaller sample size will dominate a larger one if (1) the value of the outside option is such that taking the action requires fewer successes for the smaller sample size, and (2) the designer’s prior puts enough weight on small \( q \)’s.

### 3.2 Degenerate priors

When the designer assigns probability 1 to some fixed \( q \), full revelation is optimal if and only if \( q \geq t \). Indeed, with full revelation, the DM takes the action with probability 1 when \( q \geq t \) and with probability 0 when \( q < t \) whereas the probability of taking the action is strictly between 0 and 1 for any finite sample size.

The more interesting case is \( q < t \). Observation 1 implies that the optimal sample size is weakly larger than \( \lfloor t^{-1} \rfloor \) in this case. To see that larger sample sizes have the potential be be optimal, let us consider sample sizes 1 and 3. When \( t \leq 1/3 \), sample-size monotonicity implies that sample size 3 dominates sample size 1. When \( t \) exceeds the “critical point” \( 1/3 \), the probability of taking the action for sample
size 3 "loses" the term $b(3, 1, q)$ in the sum on the right-hand side of Equation (1) because a single success no longer triggers the DM to take the action. This leads to a discrete drop in the probability of taking the action. However, the marginal increase in the probability of taking the action as $q$ increases is larger for sample size 3 than for sample size 1. For $q < 1/2$, sample size 1, which did not lose a term, dominates sample size 3. At the cutoff probability $q^* = 1/2$, sample size 3 catches up with sample size 1, and it dominates sample size 1 from that point until the next critical point at 2/3, i.e., for $1/2 < q < t \leq 2/3$. When $t$ exceeds 2/3, the probability of taking the action for sample size 3 loses another term, $b(3, 2, q)$, and sample size 1 dominates it again. The left panel in Figure 2 provides a graphical illustration of this phenomenon. The right panel in Figure 2 illustrates a similar phenomenon for sample sizes 3 and 5.

More generally, the discrete drop in $P(n, \lceil tn \rceil, q)$ at critical points of the form $k/n$ followed by a continuous increase of $P(n, \lceil tn \rceil, q)$ as $q$ moves toward $t$ implies that larger sample sizes than those identified in Observation 1 may be optimal. Theorem 1 establishes that this does not happen when $q$ is not too close to $t$, and Theorem 2 establishes that it does when $q$ is close to $t$. Analogous theorems, entitled Theorems
3 and 4 respectively, hold for monotone priors, and so we postpone the discussion of the economic implications of the four theorems to Section 3.3 and focus here on stating and proving Theorems 1 and 2.

**Theorem 1.** For any $t$ and $q \in (0, (1 + |t^{-1}|)^{-1})$, the optimal sample size is the largest one satisfying that even a single success would trigger the DM to take the action.

Put differently, for any integer $n$, sample size $n$ is optimal when $q < 1/(n + 1) < t \leq 1/n$. The shaded rectangles in Figure 3 provide a graphical illustration.

The proof intuition is as follows. By sample-size monotonicity, for any number of successes $k$, the largest sample size $n(k)$ such that $k/n(k) \geq t$ dominates smaller sample sizes that require at least $k$ successes for taking the action. The optimization problem thus reduces to

$$\max_{k \geq 1} P(n(k), k, q)$$

where $n(k)$ increases in $k$. By single crossing, if $q$ is not too large, in other words,
if successes are not very likely, the probability $P(n(1), 1, q)$ dominates $P(n(k), k, q)$ for $k > 1$. We establish in the proof that $q$ is not too large if it is smaller than $1/(n(1) + 1)$.

**Proof of Theorem 1.** Fix an integer $n'$ and $q < 1/(n' + 1) < t \leq 1/n'$. By Observation 1, it suffices to show that sample size $n'$ dominates larger sample sizes. To show this, we first observe that if sample size $n'$ is optimal for some $t \in \left(\frac{1}{n'+1}, \frac{1}{n'}\right]$ then it is also optimal for larger $t$’s in this range. This is because as $t$ increases, the smallest number of successes required for taking the action remains the same for sample size $n'$ while it may increase for larger sample sizes implying a decrease in the probability of taking the action for these sample sizes. It thus suffices to establish optimality for $t$’s just above $1/(n' + 1)$.

For every $k > 1$, let $n(k)$ be the largest integer such that $k/n(k) > t$. As $t$ converges to $1/(n' + 1)$ from above, we have that $n(k) = (n' + 1)k - 1$. By sample-size monotonicity, the probability of taking the action for sample size $n(k)$ is larger than the probability of taking the action for any other sample size $n$ such that the smallest number of successes required for taking the action is $k$. It thus suffices to show that $P(n', 1, q) > P(n(k), k, q)$ for $q < 1/(n' + 1)$.

Because $k > 1$ and $n(k) > n'$, single crossing implies that the desired inequality holds if the cutoff probability $q^*$ for $P(n', 1, q)$ and $P(n(k), k, q)$ is weakly larger than $1/(n' + 1)$. To prove this, it suffices to show that $P(n', 1, 1/(n' + 1)) \geq P(n(k), k, 1/(n' + 1))$ for every $k > 1$ which we do in Lemma 1 in Appendix B. □

Theorem 2 studies the optimal sample size when the value $q$ of taking the action is close to the value $t$ of the status quo.

**Theorem 2.** For $\frac{1}{2} < q < t$, the optimal sample size is the largest odd sample size satisfying that a simple majority of successes would trigger the DM to take the action.

In other words, for any integer $m$, sample size $2m - 1$ is optimal when $t \in$
and \( \frac{1}{2} < q < t \). The dotted trapezoids in Figure 3 provide a graphical illustration.

The proof intuition is as follows. To see that sample size \( 2m - 1 \) dominates smaller sample sizes, we first observe that for any smaller sample size, the DM takes the action after obtaining a simple majority of successes. Since simple majority is achieved by the same number of successes for an even sample size and the odd one just above it, sample-size monotonicity implies that even sample sizes are dominated. As for odd sample sizes \( \leq 2m - 1 \), the symmetry of the binomial distribution around 1/2 when \( q = 1/2 \) implies that the cutoff probability for every two odd sample sizes is 1/2. Thus, by single crossing, the largest among them, that is, sample size \( 2m - 1 \), dominates all smaller ones for \( q > 1/2 \). To see that sample size \( 2m - 1 \) dominates larger sample sizes, we observe that the smallest number of successes that trigger the DM to take the action for larger sample sizes is larger than \( m \). Thus, by single crossing, establishing the dominance of sample size \( 2m - 1 \) over larger sample sizes requires showing that the cutoff probability is \( \geq t \), which is proved in Appendix A.

Note that Theorems 1 and 2 solve for the optimal sample size when \( q < t \) with the exception of the parameter range \( (1 + |t^{-1}|)^{-1} \leq q < t < 1/2 \). Characterizing the optimal sample size in this case is challenging. This is because, unlike in Theorem 2, the cutoff probability \( q^* \) is not well-behaved as the sample sizes change, and so identifying the \((t, q)\) boundaries at which the optimal sample size changes is not straightforward.

### 3.3 Non-degenerate priors

We now extend the analysis to non-degenerate monotone or symmetric priors.

Our first result establishes that the optimal sample size identified in Theorem 1 is also optimal for decreasing priors where a prior is decreasing (increasing) if it weakly decreases (increases) in \( q \) on \([0, 1]\) and differs from the uniform prior on a non-zero measure.
Theorem 3. For any decreasing prior and any $t$, the optimal sample size is the largest one satisfying that even a single success would trigger the DM to take the action.

The proof intuition is as follows. Similarly to Theorem 1, we can restrict attention to sample sizes of the form $n(k)$. For all these sample sizes, the expected probability of taking the action is identical under the uniform prior. We can think about the transition from a uniform prior to a decreasing one as an operation that “shifts mass” from higher values of $q$ to lower values of $q$. By single crossing, mass is thus shifted to values of $q$ in which smaller sample sizes dominate larger ones, and so the smallest among these sample sizes is uniquely optimal.

Proof of Theorem 3. We first consider finite sample sizes. Similarly to Theorem 1, it suffices to show that the expected probability of taking the action for sample size $n' = \lfloor t^{-1} \rfloor$ dominates that for sample sizes of the form $n(k) = k(n' + 1) - 1$ where $k \geq 2$, or, more specifically, that the expected probability $P(n', 1) = \int_{0}^{1} P(n', 1, q)f(q) dq$ is strictly larger than $P(n(k), k) = \int_{0}^{1} P(n(k), k, q)f(q) dq$ where $f$ is the designer’s prior.

By single crossing, $P(n', 1, q)$ crosses $P(n(k), k, q)$ from above at some $q^* > 0$. Suppose $f(q^*) = 0$. Then, because $f$ is decreasing, its support is nested in $[0, q^*]$ and $P(n', 1, q)$ dominates $P(n(k), k, q)$ on the entire support of $f$. The result follows.

Suppose $f(q^*) > 0$. We establish the result in two steps. The first step shows that it suffices to examine the expected probabilities with respect to the uniform prior.

Step 1. If $P(n', 1) \geq P(n(k), k)$ with respect to the uniform prior, then $P(n', 1) > P(n(k), k)$ with respect to $f$.

Proof. Consider the constant function $h(q) = f(q^*)$. Because $f$ is decreasing, the function $h$ either reduces the mass on $q$’s smaller than $q^*$ relative to $f$, i.e., $\int_{0}^{q^*} f(q) dq > \int_{0}^{q^*} h(q) dq$, or $h$ increases the mass on $q$’s larger than $q^*$. Thus, by single crossing, if $P(n', 1) \geq P(n(k), k)$ where the expectation is taken with respect to $h$ then a strict
inequality holds with respect to \( f \). Since \( h \) is a re-scaling of the uniform distribution, 
\[ P(n', 1) \geq P(n(k), k) \]
with respect to \( h \) if and only if 
\[ P(n', 1) \geq P(n(k), k) \]
with respect to the uniform prior. The result follows.

The second step examines the expected probabilities according to the uniform prior.

**Step 2.** For the uniform prior, 
\[ P(n', 1) = P(n(k), k). \]

**Proof.** For the uniform prior, 
\[ P(n, k) = 1 - \frac{k}{n+1} \]
because (i) \( P(n, k, q) \) is equal to the incomplete regularized Beta function \( I_q(k, n - k + 1) \) and (ii) \( I_q(k, n - k + 1) \) is the CDF of the \( Beta(k, n - k + 1) \) distribution with mean \( \frac{k}{n+1} \). Thus,
\[
P(n(k), k) = 1 - \frac{k}{(n(k) + 1)} = 1 - \frac{1}{(n' + 1)} = P(n', 1).
\]

As for full revelation, by using the reasoning of Step 1 above, it suffices to show that sample size \( n' \) dominates full revelation for the uniform prior. This is true because the expected value of full revelation in this case is \( 1 - t \) and \( t > 1/(n' + 1) \).

Theorems 1 and 3 have several implications. First, for any integer, there is a non-trivial interval of status quo values for which this integer is the optimal sample size. Clearly, providing the least amount of information to the DM need not be optimal. Second, fixing the designer’s prior, as the status quo becomes more attractive, the optimal sample size gradually decreases. In other words, the designer chooses to provide less information to the DM as the DM’s ex-ante tendency to take the action decreases. Third, consider an analyst who observes the DM’s choices as a function of the data obtained by the DM. The analyst may conclude that the DM’s choice behavior is consistent with a strong preference for taking the action: The DM chooses to take the action unless all realizations are in favor of keeping the status quo.
We proceed to analyze increasing priors. The intuition developed for Theorem 3 suggests that the largest $k$ and the corresponding sample size $n(k)$ satisfying $k/n(k) \geq t$ would dominate smaller sample sizes. This is because an increasing prior shifts mass, relative to the uniform prior, from lower values of $q$ where smaller sample sizes of the form $n(k)$ dominate to larger values of $q$ where larger sample sizes of the form $n(k)$ dominate. Building on this intuition, we first identify a tighter lower bound on the optimal sample size for increasing priors. We then show this bound is optimal when the prior is also concave.

For a fixed $t$, let $n' = \lfloor t^{-1} \rfloor$ and let $a_k = k/n(k)$ where $n(k) = k(n' + 1) - 1$ and $k \geq 1$. Because the sequence $\{a_k\}$ decreases in $k$ and converges to $1/(n' + 1) < t$ as $k$ tends to infinity, there exists a unique $k'$ such that $a_{k' + 1} < t \leq a_{k'}$.

**Proposition 1.** For any increasing prior and any $t$, the optimal sample size is weakly larger than $n(k')$ where $n' = \lfloor t^{-1} \rfloor$ and $k'$ is the unique integer satisfying $a_{k' + 1} < t \leq a_{k'}$.

Note that the lower bound identified in Proposition 1 is identical to the one identified in Observation 1 for $t \in (1/(n' + 1), 1/n')$. However, for $t \in (1/(n' + 1), 1/(n' + 1/2)]$, the lower bound of Proposition 1 is tighter. For example, if $t > 1/2$, then the bound of Observation 1 implies that any sample size can be optimal whereas Proposition 1 implies that the smallest candidate for optimality is the largest sample size satisfying that a simple majority of successes would trigger the DM to take the action. The following result shows that this candidate is indeed optimal when the prior is also concave.

**Theorem 4.** For any increasing and concave prior and any $t > 1/2$, the optimal sample size is the largest one satisfying that a simple majority of successes would trigger the DM to take the action.

Thus, similarly to Theorem 2, any odd sample size is optimal for some interval of status quo values, and the optimal sample size decreases as the value of the status
quo increases. In addition, an analyst who observes the DM’s choices may conclude that the DM’s choice procedure is consistent with simple majority: the DM takes the action when a majority of the realizations are in favor of doing so.

Theorem 4 extends to the class of priors that increase in $q$ and decrease in the ratio $f(q)/q$. Intuitively, when the ratio $f(q)/q$ decreases in $q$, the prior assigns sufficient mass to small values of $q$ relative to large ones that simple majority continues to be optimal. When this is not the case, very large sample sizes or full revelation may be optimal. Indeed, fix some $q'$ and consider a prior that assigns probability 0 to all $q$ below $q'$. For this increasing prior, full revelation is optimal when $t \leq q'$ because the probability of taking the action for full revelation is 1 whereas it is strictly smaller than 1 for any finite sample size.

The last class of priors that we consider is the class of symmetric priors.

Theorem 5. For any symmetric prior with a non-zero mass around $1/2$ and any $t > 1/2$, a sample size is optimal if and only if it is an odd sample size satisfying that a simple majority of successes would trigger the DM to take the action.

Thus, similarly to Theorem 4, simple majority arises endogenously in the setting with symmetric priors.

Figure 4 summarizes the optimal sample sizes for monotone or symmetric priors. Decreasing priors provide the lower envelope of the optimal sample sizes. Increasing and concave priors provide the upper envelope for $t > 1/2$, and symmetric priors span the range of odd sample sizes between the lower and upper envelopes for $t > 1/2$. The figure implies that the designer does not need to have an exact prior in order to identify the optimal sample size. Rather, knowledge of more general properties such as monotonicity, concavity, or symmetry suffices.

Another implication of Figure 4 relates to cases in which the designer’s prior is decreasing, increasing and concave, or symmetric, and the designer is concerned that the sample size choice would reveal information about the prior to the DM. For

---

See the proof of Theorem 4 in Appendix A for details.
example, if the designer chooses sample sizes as in Figure 4, the DM may be able to infer that the prior is not decreasing when the chosen sample size is 3 and \( t > \frac{1}{2} \). One way the designer may proceed in such cases is to choose some distribution over sample sizes (possibly a degenerate one) that is independent of the prior. Our previous results imply the following robust prediction regarding the support of this distribution.

**Corollary 1.** For any \( t > \frac{1}{2} \), if sample size \( n \) is chosen by the designer with positive probability then it is an odd sample size satisfying that a simple majority of successes would trigger the DM to take the action.

In other words, only sample sizes in the set \( \mathcal{D} = \{1, 3, \ldots, 2k' - 1\} \), where \( k' \) is the largest integer such that \( k'/(2k' - 1) \geq t \), are candidates for optimality. To verify this, it suffices to show that any sample size \( n \notin \mathcal{D} \) is dominated by some sample size in \( \mathcal{D} \) irrespectively of whether the prior is decreasing, increasing and concave, or symmetric. This holds because (1) if \( n \) is an even integer smaller than \( 2k' - 1 \) then by sample-size monotonicity it is dominated by the odd sample size just above it,
which is in $\mathcal{D}$, and (2) if $n$ is larger than $2k' - 1$ or $n = \infty$ then it is dominated by sample size $2k' - 1$ as the proofs of Theorems 3, 4 and 5 indicate.

### 3.4 Bounds on the sample size

The analysis so far characterized the designer’s optimal sample size without any limitation on the sample size. It is possible, however, that the designer faces constraints when choosing the sample size.

One relevant constraint is an upper bound on the sample size. For example, in an experimental setting or a public opinion survey, the pool of potential participants may be small. Incorporating such an upper bound is relatively straightforward. We demonstrate for the setting of Theorem 3.

**Observation 2.** For any decreasing prior and any upper bound $\bar{n}$ on the sample size, sample size $\bar{n}$ is optimal for $t \leq 1/\bar{n}$, and the sample size identified in Theorem 3 is optimal otherwise.

Clearly, for $t > 1/\bar{n}$, the optimal sample size identified in Theorem 3 is feasible and therefore optimal. For $t \leq 1/\bar{n}$ and sample size $n \leq \bar{n}$, even a single success would trigger the DM to take the action, and thus by sample-size monotonicity $\bar{n}$ is optimal.

Another relevant constraint is a lower bound $\underline{n}$ on the sample size. For example, the Food and Drug Administration requires a minimal number of participants in clinical trials when a pharmaceutical company applies for drug approval. Incorporating such a bound into the analysis is more involved. We demonstrate again for the setting of Theorem 3.

**Proposition 2.** For any decreasing prior, $t > 1/2$, and an odd lower bound $\underline{n}$ on the sample size, if simple majority triggers the DM to take the action for sample size $\underline{n}$, then this sample size is optimal. Otherwise, the optimal sample size is weakly smaller than $2\underline{n} + 1$. 

24
4 Bayesian inference

This section studies the designer’s optimal sample size for Bayesian inference from a $Beta(\alpha, \beta)$ prior, and compares the results to the ones obtained for unbiased inference.

$Beta(\alpha, \beta)$ priors are commonly used to model prior knowledge in Bayesian environments with Bernoulli experimentation and are especially suitable in the current context for three reasons. First, the class of $Beta(\alpha, \beta)$ priors is rich. It has a wide variety of shapes as a function of $\alpha$ and $\beta$ making it suitable for studying distributions with shapes we already analyzed in Section 3. Second, $Beta(\alpha, \beta)$ priors facilitate interesting comparative statics on how the strength of the prior ($\alpha + \beta$) and its bias $\frac{\alpha}{\alpha + \beta}$ affect the optimal sample size. Third, $Beta(\alpha, \beta)$ priors are analytically tractable because the $Beta(\alpha, \beta)$ distribution is a conjugate prior for the Bernoulli distribution implying that the DM’s posterior is also a $Beta$ distribution.

Consider an environment in which the DM and the designer share a $Beta(\alpha, \beta)$ prior on $q$ and where the DM Bayes-updates this prior based on the sample in order to estimate $q$. Thus, the DM believes the expected value of $q$ is $\frac{\alpha}{\alpha + \beta}$ prior to obtaining any data and $\frac{\alpha + k}{\alpha + \beta + n}$ after obtaining the sample $(n, k)$.

The introduction of a prior has two immediate implications for the analysis. First, for $t \leq \frac{\alpha}{\alpha + \beta}$, the DM will take the action without obtaining any additional data, and thus the designer will not provide any data to the DM. Second, for $t > \frac{\alpha + n}{\alpha + \beta + n}$, fewer than $n + 1$ experiments will not trigger the DM to take the action even if all realizations are successes. Thus, for $t > \frac{\alpha + 1}{\alpha + \beta + 1}$ the set of candidates for optimality with Bayesian inference is smaller than with unbiased inference, and it shrinks as $t$ increases. For the intermediate case in which $t \in \left( \frac{\alpha}{\alpha + \beta}, \frac{\alpha + 1}{\alpha + \beta + 1} \right]$, the set of candidates for optimality is identical to that of unbiased inference.

We now analyze three classes of $Beta(\alpha, \beta)$ priors for $t \in \left( \frac{\alpha}{\alpha + \beta}, \frac{\alpha + 1}{\alpha + \beta + 1} \right]$. The first is the class of decreasing $Beta(\alpha, \beta)$ priors. This class is parameterized by $\alpha < 1 < \beta$. The second is the class of increasing and concave priors, which is parameterized by $\beta = 1 < \alpha \leq 2$. The third is the class of symmetric priors, which is parameterized
by \( \alpha = \beta \). In all cases, we obtain results that are “compressed” versions of the ones established for unbiased inference.

For the class of decreasing priors, Theorem 3B, where B stands for Bayesian, establishes that for \( t \)’s close to the prior’s mean, the optimal sample size is the largest one satisfying that even a single success would trigger the DM to take the action.

**Theorem 3B.** For a decreasing Beta(\( \alpha, \beta \)) prior, the largest sample size \( n' \) satisfying that even a single success would trigger the DM to take the action for some \( t > \frac{\alpha}{\alpha + \beta} \) is optimal when \( t \in \left( \frac{\alpha}{\alpha + \beta}, \frac{\alpha + 1}{\alpha + \beta + n'} \right) \).

Theorem 3B is a compressed version of Theorem 3 in the sense that the interval of \( t \)’s for which sample size \( n' \) is optimal is nested in the corresponding interval \( \left( \frac{1}{n' + 1}, \frac{1}{n'} \right) \) of \( t \)’s for unbiased inference.\(^9\) The nestedness of the intervals illustrates how the prior affects the analysis. Providing no data is optimal for \( t \)’s smaller than the prior mean thus eliminating the interval \( \left( \frac{1}{n' + 1}, \frac{\alpha}{\alpha + \beta} \right) \) of \( t \)’s for which sample size \( n' \) was optimal for unbiased inference. And because the posterior belief is a weighted average of the prior mean and the sample mean, the sample \((n', 1)\) will not trigger the DM to take the action for \( t \)’s between the posterior mean and the sample mean thus eliminating the interval \( \left( \frac{\alpha + 1}{\alpha + \beta + n'}, \frac{1}{n'} \right) \) of \( t \)’s as well.

An interesting comparative statics relates to how the designer’s optimal sample size changes when the prior’s bias decreases. To isolate the effect of a decrease in the bias, we fix the strength of the prior and reduce its bias by reducing \( \alpha \). As \( \alpha \) decreases, the largest sample size satisfying the conditions of Theorem 3B increases, and thus the optimal sample size increases for \( t \)’s just below the original prior mean.

---

\(^9\)To see the nestedness, observe that by definition sample size \( n' \) has to satisfy the inequality

\[
\frac{\alpha + 1}{\alpha + \beta + n' + 1} \leq \frac{\alpha}{\alpha + \beta} < \frac{\alpha + 1}{\alpha + \beta + n'}
\]

and thus the inequality

\[
\frac{1}{n' + 1} \leq \frac{\alpha}{\alpha + \beta} < \frac{1}{n'}.
\]

The expression \( \frac{\alpha + 1}{\alpha + \beta + n'} \) is a weighted average of the prior mean and the sample mean \( \frac{1}{n'} \) and thus lies between them.
Proof of Theorem 3B. Let \( n' \) be the largest sample size satisfying the conditions in the statement of the theorem. By definition, we have \( \frac{\alpha + 1}{\alpha + \beta + n' + 1} \leq \frac{\alpha}{\alpha + \beta} \), implying that
\[
\frac{\alpha + \beta}{\alpha} \leq n' + 1.
\] (2)
Sample size \( n' \) dominates full revelation by reasoning analogous to that in the proof of Theorem 3 because \( t > \frac{\alpha}{\alpha + \beta} \geq \frac{1}{n' + 1} \) where the right inequality is implied by Equation (2). For finite sample sizes, recall that for any decreasing prior, any \( k \geq 2 \), and \( n(k) = k(n' + 1) - 1 \), we have that \( P(n', 1) > P(n(k), k) \) by the proof of Theorem 3. Thus, to conclude the proof, it suffices to show that any sample \( (n, k) \), where \( n \) is any integer and \( k \) is the smallest number of successes triggering the DM to take the action for sample size \( n \), is dominated by the sample \( (n(k), k) \). And to show this dominance, it suffices to prove — by sample-size monotonicity — that \( n(k) \geq n \).

The inequality \( n(k) \geq n \) holds because the choice of \( k \) implies that \( \frac{\alpha + k}{\alpha + \beta + n} > \frac{\alpha}{\alpha + \beta} \) and thus
\[ n < \frac{k(\alpha + \beta)}{\alpha} \leq k(n' + 1) \]
where the right inequality is by Equation (2). Thus \( n \leq k(n' + 1) - 1 = n(k) \) as required.

Our second result is about increasing and concave priors, and is analogous to Theorem 4. We establish that for \( t \)'s close to the prior’s mean, the optimal sample size is the largest one satisfying that a simple majority of successes would trigger the DM to take the action.

Theorem 4B. For any increasing and concave Beta(\( \alpha, \beta \)) prior, the largest odd sample size \( 2m - 1 \) satisfying that a simple majority of successes would trigger the DM to take the action for some \( t > \frac{\alpha}{\alpha + \beta} \) is optimal when \( t \in \left( \frac{\alpha}{\alpha + \beta}, \frac{\alpha + m}{\alpha + \beta + 2m - 1} \right) \).

Similarly to Theorem 3B, the interval of \( t \)'s for which a given odd sample size is optimal is nested in the corresponding interval for unbiased inference illustrating again how the prior affects the analysis.
Beta(α, β) priors are increasing and concave for β = 1 < α ≤ 2. Fixing β = 1 and decreasing α in this range, we obtain a comparative static on how a reduction in the bias and strength of the prior affect the optimal sample size. Similarly to above, the optimal sample size increases for t’s just below the original prior mean. Our third result is about symmetric Beta(α, β) priors. We establish that the entire structure of the optimal sample sizes identified in Theorem 5 continues to hold for a smaller interval of t’s.

**Theorem 5B.** For any Beta(α, α) prior and \( t \in \left( \frac{1}{2}, \frac{\alpha+1}{2\alpha+1} \right) \), a sample size is optimal if and only if it is an odd sample size satisfying that a simple majority of successes would trigger the DM to take the action.

Figure 1 provides a graphical illustration of how the optimal sample sizes compare to those of unbiased inference. For \( t < 1/2 \), the optimal sample size for unbiased inference is larger than for Bayesian inference and for \( t > \frac{\alpha+1}{2\alpha+1} \), a reverse ranking of the optimal sample sizes holds for \( \alpha \leq 1 \). For intermediate t’s, the optimal sample sizes for Bayesian inference are a compressed version of those for unbiased inference.

Reducing the strength of the symmetric prior by reducing \( \alpha \) increases the interval of t’s for which Theorem 5B holds. The upper bound of the interval converges to 1 as \( \alpha \) tends to 0 implying that the optimal sample sizes for Bayesian inference converge to those of unbiased inference as the DM’s prior strength converges to 0 and the DM puts all the weight on the sample.

The analysis of the optimal sample size for \( t > \frac{\alpha+1}{2\alpha+1} \) is more involved. As the strength of the prior increases, there are two aspects of the problem that change: the shape of the prior and the candidates for optimality. These two aspects may operate in opposite directions making the analysis challenging. A focal case of interest is the uniform prior, which is the Beta(1, 1) distribution. Because we can derive closed form expressions for the expected probability of taking the action for the uniform prior, we can solve for the optimal sample size directly.
Figure 5: Optimal sample sizes for Bayesian inference from a uniform prior and $t \in [2/3, 4/5)$

**Proposition 3.** For the uniform prior, any integer $n$, and $t \in \left(\frac{1+n-1}{2+n-1}, \frac{1+n}{2+n}\right)$, sample size $n'$ is optimal if and only if (i) $n'$ is congruent modulo $(n+1)$ to $n$, and (ii) a $\frac{n}{n+1}$ super-majority of successes would trigger the DM to take the action.\(^{10}\)

For example, for $n = 1$, the relevant interval of $t$’s is $\left(\frac{1}{2}, \frac{2}{3}\right]$, and the optimal sample sizes are all odd integers (i.e., congruent to 1 modulo 2) satisfying that a simple majority (i.e., one-half super-majority) of successes would trigger the DM to take the action. For $n = 2$, the required super-majority is two thirds, and for $n = 3$ it is three quarters. Figure 5 depicts the collection of optimal sample sizes in these cases.

## 5 Conclusion

This paper considered a designer who wishes to convince a DM to take an action and controls how much data about a payoff-relevant parameter the DM sees prior to making a choice. The DM is a statistician who uses the data and statistical inference to estimate the parameter and decide whether to take the action.

---

\(^{10}\)For sample size $n'$ and some $t$, we say that a $\frac{n}{n+1}$ super-majority of successes triggers the DM to take the action if $k'/n' > t$ where $k'$ is the smallest number of successes such that $k'/n' > n/(n+1)$.
A large part of the analysis focused on classical frequentist inference that does not rely on prior beliefs. In addition to analytical convenience, incorporating prior-free statistical inference into economic theory and the study of strategic interactions is relevant because of the increasing use of prior-free estimation methods and machine learning techniques in real-life strategic settings. Earlier contributions that incorporate the frequentist approach into economic theory include Al-Najjar (2009) in the context of individual decision making and Salant and Cherry (2020) in the context of multi-person decision making.

Our main results established that the designer’s optimal choice of sample size exhibits two regularities. The designer either tends to choose sample sizes satisfying that even a single success would trigger the DM to take the action, or that a simple majority of successes would trigger the DM to take the action. As a result, the DM’s choice behavior as a function of data realizations either exhibits a “strong” preference against the status quo or it follows the simple-majority rule. These regularities hold for both unbiased inference and Bayesian inference from a Beta prior although in the latter case the regularities are “compressed” for a smaller interval of parameter values because the prior introduces inertia toward its mean.

A Proofs

**Proof of Property 2.** Consider the ratio function

\[ r(q) = \frac{P'(n, k, q)}{P'(n', k', q)} = \frac{k^n}{k'^{n'}} q^{k-k'} (1-q)^{(n-k-n'+k')} \]

where \( P' \) is the derivative of \( P \). The function \( r(q) \) is continuous and approaches 0 from above as \( q \) approaches 0 implying that \( P(n', k', q) \) is above \( P(n, k, q) \) in a neighborhood of 0.

To verify that \( P(n, k, q) \) crosses \( P(n', k', q) \) at most once in \((0, 1)\), assume to the contrary that there are two or more such crossings. Then, the difference function
$\Delta(q) = P(n', k', q) - P(n, k, q)$ has at least three interior extremum points with $\Delta'(q) = 0$ because $\Delta(0) = \Delta(1) = 0$. Because $\Delta'(q) = 0$ implies $r(q) = 1$ for $q \in (0, 1)$, the ratio function $r(q)$ thus has at least three interior points in which it is equal to 1. But as we will momentarily show, $r(q)$ has at most two such points implying that $P(n, k, q)$ crosses $P(n', k', q)$ at most once in $(0, 1)$.

To observe that $r(q)$ has at most two points in which it is equal to 1, we consider two cases. First, if $(n - k - n' + k') \leq 0$ then $r(q)$ strictly increases in $q$ and thus has at most one point in which it is equal to 1. Second, if $(n - k - n' + k') > 0$ then $r(q)$ increases up to $\frac{k - k'}{n - n'} < 1$ and then decreases. Thus, there are at most two interior points in which $r(q)$ is equal to 1.

\[\Box\]

**Proof of Theorem 2.** Fix an integer $m \geq 1$. By the discussion following the statement of Theorem 2, it suffices to show that sample size $2m - 1$ dominates larger sample sizes for $t \in \left(\frac{m+1}{2m+1}, \frac{m}{2m-1}\right)$. For sample size $n \geq 2m - 1$, let

$$k[n] = \left\{ \left\lceil nt \right\rceil \mid t \in \left(\frac{m+1}{2m+1}, \frac{m}{2m-1}\right) \right\} \quad (1)$$

denote the set of all integers $k$ satisfying that there exists $t$ in the relevant interval for which $k$ is the smallest integer such that the DM takes the action after obtaining the sample $(n, k)$. Let $\kappa(n)$ denote the smallest integer in $k[n]$. By definition, $k[2m - 1]$ is a singleton with $\kappa(2m - 1) = m$, and $\kappa(2m) = m + 1$.

Figure 6 illustrates construction of the set $k[n]$ for $n = 3, \ldots, 13$ and $m = 2$. For $m = 2$, the relevant interval of $t$’s is $(3/5, 2/3)$. Dots in the figure (solid and striped) correspond to $k/n$’s and they are circled when $k \in k[n]$. For example, $k[8] = \{5, 6\}$ and therefore the corresponding dots are circled, while $7 \notin k[9]$ and therefore the corresponding dot is not circled. Circled striped dots correspond to $\frac{\kappa(n)}{n}$. For example, $\kappa(8) = 5$ and therefore the corresponding dot is striped.

Fix a sample size $n > 2m - 1$ and $k \in k[n]$. Let $\Delta_{n,k}(q) = P(2m - 1, m, q) - P(n, k, q)$. We need to show that $\Delta_{n,k}(q) > 0$ for $q \in (0, z(n,k)]$ where $z(n,k) = \left(\frac{m+1}{2m+1}, \frac{m}{2m-1}\right)$.
min \( \left\{ \frac{k}{n}, \frac{m}{2m-1} \right\} \). By the single-crossing property, it suffices to show that the cutoff probability \( q^* \) of \( P(2m-1, m, q) \) and \( P(n, k, q) \) is weakly larger than \( z(n, k) \), and to do so, it suffices to show that \( \Delta(n, k, z(n, k)) > 0 \).

Let \( \mathcal{D} \) be a set of all sample sizes \( n \) satisfying that \( \kappa(n) = \frac{n+1}{l} \) for every \( n_1 \) with \( 2m - 1 \leq n_1 < n \). In Figure 6, sample sizes 3, 8 and 13 belong to set \( \mathcal{D} \) and the corresponding dots are circumscribed by diamonds. We first prove that for any sample size \( n \in \mathcal{D} \setminus \{2m-1\} \), \( \Delta(n, \kappa(n))(z(n, \kappa(n))) > 0 \). This domination is shown by solid arrows in Figure 6. We then prove that \( \Delta(n, k) > 0 \) for all other pairs \( (n, k) \) with \( n > 2m - 1 \) and \( k \in k[n] \).

Fix a sample size \( n \in \mathcal{D} \setminus \{2m-1\} \). Lemma 2 implies that \( n = n(l) = (2m+1)l - 2 \) for some \( l \geq 2 \). By definition, \( \kappa(n) = (m+1)l - 1 \). Since \( z(n, \kappa(n)) = \kappa(n)/n \) in this case, we need to prove that \( \Delta(l) = \Delta(n(l), \kappa(n(l)))(\kappa(n(l))/n(l)) > 0 \) for any \( l \geq 2 \). Lemma 4 establishes this inequality for \( l = 2 \), and Lemma 5 for \( l \geq 3 \).

Fix any of the remaining pairs \( (n, k) \) with \( n > 2m - 1 \) and \( k \in k[n] \). By Lemma
there exists \( n_1 \in \mathcal{D} \) that satisfies \( n_1 \leq n, n_1 - \kappa(n_1) \leq n - k, \) and \( \frac{\kappa(n_1)}{n_1} \leq \frac{k}{n} \). The following series of inequalities implies that \( \Delta_{n,k}(z(n,k)) > 0 \):

\[
P(n, k; \frac{k}{n}) < P\left(n_1, \kappa(n_1), \frac{\kappa(n_1)}{n_1}\right) < P\left(2m - 1, m, \frac{\kappa(n_1)}{n_1}\right) < P\left(2m - 1, m, \frac{k}{n}\right).
\]

The left inequality holds by the Lemma below, the middle one was proved in the previous paragraph, and the right one follows from the monotonicity of \( P(\cdot, \cdot, q) \) in \( q \).

**Lemma.** Suppose \( n_1 < n, n_1 - k_1 \leq n - k, \) and \( \frac{k_1}{n_1} \leq \frac{k}{n} \). Then \( P(n, k; \frac{k}{n}) < P(n_1, k_1; \frac{k_1}{n_1}) \).

**Proof.** The required inequality follows from:

\[
P(n, k; \frac{k}{n}) \leq P\left(n_1 + k - k_1, k, \frac{k}{n_1 + k - k_1}\right) < P\left(n_1, k_1; \frac{k_1}{n_1}\right).
\]

The left inequality is obtained by applying \( (n - n_1) - (k - k_1) \) times the inequality \( P(n, k; \frac{k}{n}) \leq P(n - 1, k; \frac{k}{n - 1}) \) (Anderson and Samuels (1967), Theorem 2.3). The right inequality is obtained by applying \( k - k_1 \geq 0 \) times the inequality \( P(n, k; \frac{k}{n}) < P(n - 1, k - 1, \frac{k - 1}{n - 1}) \) (Anderson and Samuels (1967), Theorem 2.2).

**Proof of Proposition 1.** By definition, after obtaining the sample \((n(k'), k')\), the DM takes the action. For a sample size \( n < n(k') \), the smallest number of successes that trigger the DM to take the action is \( k \leq k' \). By sample-size monotonicity, sample size \( n \) is a candidate for optimality only if it has the form \( n = n(k) \). Thus, to complete the proof, it suffices to show that \( P(n(k'), k') > P(n(k), k) \) for any increasing prior \( f \) and \( k < k' \).

For \( k < k' \) and \( n(k) < n(k') \), single crossing implies that \( P(n(k), k, q) \) crosses \( P(n(k'), k', q) \) from above at some \( q^* > 0 \). Because \( P(n(k), k) = P(n(k'), k') \) for the uniform prior, as established in Step 2 in Theorem 3, we also have that \( q^* < 1 \). The reminder of the proof is analogous to the proof of Theorem 3.
Proof of Theorem 4. Fix an integer \( m \geq 1 \) and let \( t \in \left( \frac{m+1}{2m+1}, \frac{m}{2m-1} \right) \). The largest sample size such that a simple majority of successes triggers the DM to take the action is \( 2m - 1 \).

We first consider finite sample sizes. By Proposition 1, it suffices to show that sample size \( 2m - 1 \) dominates larger sample sizes. So fix a sample size \( n > 2m - 1 \) and \( k \in k[n] \), where \( k[n] \) is defined as in Equation (1). The following two steps establish that \( P(2m - 1, m) > P(n, k) \).

The first step establishes that it suffices to examine the expected probabilities with respect to the linear prior \( h(q) = 2q \).

**Step 1.** If \( P(2m - 1, m) > P(n, k) \) for the linear prior then \( P(2m - 1, m) > P(n, k) \) for an increasing and concave prior.

**Proof.** By the Single-crossing Property, \( P(2m - 1, m, q) > P(n, k, q) \) for \( q \in (0, q^*) \) and a reverse inequality holds for \( q \in (q^*, 1) \) if \( q^* < 1 \). Consider the linear function \( f_{h}(q) = \frac{f(q^*)}{q}q \) created by extending the secant between \((0, 0)\) and \((q^*, f(q^*))\) until \( q = 1 \). This function either reduces the mass on \( q < q^* \) or increases the mass on \( q > q^* \) implying that if \( P(2m - 1, m) > P(n, k) \) for \( f_{h} \) (which is not necessarily a density), then \( P(2m - 1, m) > P(n, k) \) for \( f \). Because \( f_{h}(q) = \frac{f(q^*)}{2q}h(q) \), it suffices to prove the inequality for \( h \).

The second step establishes the desired inequality with respect to \( h \).

**Step 2.** For the linear prior, \( P(2m - 1, m) > P(n, k) \).

**Proof.** For any \( n \) and \( k \), we have:

\[
P(n, k) = \left[ P(n, k, q)q^2 \right]_{q=0}^{1} - \int_{0}^{1} P'(n, k, q)q^2 dq
\]

\[
= 1 - \frac{k \binom{n}{k}}{(k+2)(n+2)} \int_{0}^{1} P'(n + 2, k + 2, q) dq = 1 - \frac{k(k+1)}{(n+1)(n+2)}
\]

where the first equality follows from integration by parts, and the second equality follows from \( P'(n, k, q) = k \binom{n}{k}q^{k-1}(1-q)^{(n-k)} \).
To show that $P(2m - 1, m) > P(n, k)$, we thus need to show that 

$$\frac{k(k+1)}{2(n+1)(n+2)} > \frac{(m+1)}{2(2m+1)}$$

which holds if (i) $\frac{k+1}{n+1} > \frac{m+1}{2m+1}$, and (ii) $\frac{k}{n+2} \geq \frac{1}{2}$.

Inequality (i) holds because $\frac{k+1}{n+1} \geq \frac{k}{n} \geq \frac{1}{2}$ and $t > \frac{m+1}{2m+1}$. To prove inequality (ii), it suffices to consider whether (ii) holds for $\kappa(n)$, and to do that, we define the set $\mathcal{D}$ as in Theorem 2. If $n \in \mathcal{D}\{2m - 1\}$, Lemma 2 implies that $n = (2m + 1)l - 2$ for some $l \geq 2$. By definition, $\kappa(n) = (m + 1)l - 1$. Thus, $\kappa(n)/(n + 2) \geq 1/2$.

If $n \notin \mathcal{D}$, Lemma 3 implies that there exists $n_1 \in \mathcal{D}$ that satisfies $n_1 \leq n$ and $\frac{\kappa(n_1)}{n_1} \leq \frac{\kappa(n)}{n}$. These two inequalities imply that $\kappa(n) \geq \kappa(n_1)$, which in turn implies that $\kappa(n)/(n + 2) > \kappa(n_1)/(n_1 + 2) \geq 1/2$. 

As for the full revelation, by Step 1, it suffices to show that sample size $2m - 1$ dominates full revelation for the linear prior. This holds because the expected value of the full revelation is $1 - t^2$ and $t > \frac{m+1}{2m+1}$. 

**Proof of Theorem 5.** Fix an integer $m \geq 1$. For $t \in \left(\frac{m+1}{2m+1}, \frac{m}{2m-1}\right)$, we have to show that sample sizes in the set $\mathcal{D} = \{1, 3, ..., 2m - 1\}$ are optimal.

For a finite sample size $n$, the sample $(n, k)$ triggers the DM to take the action for $t$ slightly above $1/2$ if $\frac{k}{n} > \frac{1}{2}$. Thus, the DM needs to obtain at least a simple majority of successes in order to take the action in this case.

For an even sample size $n$ and $t$ in the relevant interval, $P(n, \lceil nt \rceil)$ is weakly smaller than $P(n, \frac{n+2}{2})$ because $\frac{n+2}{2} \leq \lceil nt \rceil$, and $P(n, \frac{n+2}{2})$ is strictly smaller than $P(n+1, \frac{n+2}{2})$ by sample-size monotonicity. Thus, the expected probability $P(n, \lceil nt \rceil)$ of any even sample size $n$ is dominated by the expected probability $P(n+1, \frac{n+2}{2})$ of the odd sample size just above it.

For an odd sample size $n > 2m - 1$ and $t$ in the relevant interval, the expected probability of taking the action $P(n, \lceil nt \rceil)$ is strictly smaller than $P(n, \frac{n+1}{2})$ because $\frac{n+1}{2} < \lceil nt \rceil$ by the choice of $n$ and $t$. Thus, to show that sample sizes in $\mathcal{D}$ are optimal, it suffices to show that $P(n, \frac{n+1}{2})$ is equal for all odd integers. This holds
because for any \( n = 2k - 1 \) and any symmetric distribution \( f \), we have that

\[
P(2k - 1, k) = \int_0^{1/2} P(2k - 1, k, q) f(q) dq + \int_{1/2}^1 (1 - P(2k - 1, k, 1 - q)) f(q) dq
\]

\[
= \int_0^{1/2} P(2k - 1, k) f(q) dq + \int_{1/2}^{1/2} (1 - P(2k - 1, k)) f(q) dq
\]

\[
= \int_0^{1/2} f(q) dq = F(1/2)
\]

where the first equality holds because \( P(2k - 1, k, q) = 1 - P(2k - 1, k, 1 - q) \) and the second equality holds by the symmetry of \( f \).

Full revelation is also dominated because (i) the expected probability of taking the action for full revelation is

\[
\int_t^1 f(q) dq = \int_0^{1-t} f(q) dq = F(1-t),
\]

and (ii) \( F(1-t) < F(1/2) \) because \( t > 1/2 \) and the prior has a positive mass in a neighborhood of \( 1/2 \).

Proof of Proposition 2. For sample size \( n \), a simple majority of successes triggers the DM to take the action if and only if \( 1/2 < t \leq \frac{n+1/2}{n} \). Reasoning analogous to the proof of Theorem 3 (replacing \( (n', 1) \) with \( (n, (n+1)/2) \) in that proof) establishes the optimality of sample size \( n \) for \( 1/2 < t \leq \frac{n+1/2}{n} \).

For \( t > \frac{n+1}{2}/n \), it suffices to show that any sample size \( n \geq 2n + 1 \) is dominated by a sample size, which is “about half of it”, in order to establish the upper bound on the optimal sample size. The following table identifies candidates for domination. For every possible combination of even and odd \( n \) and \( k = \lceil tn \rceil \), the table lists a sample size \( n' \) (on left) and an integer \( k' \) (on right) such that

(i) \( n' \) is about half of \( n \) and \( n' \geq n \) (i.e., \( n' \) is feasible), and

(ii) \( k'/n' \geq k/n \) (i.e., \( k' \) successes trigger the DM to take the action for sample size \( n' \)):

| \( n \) even | \( n/2, k/2 \) | \( n/2 - 1, (k + 1)/2 - 1 \) |
| \( n \) odd (\( k/n > 2/3 \)) | \( (n - 1)/2 - 1, k/2 - 1 \) | \( (n + 1)/2, (k + 1)/2 \) |
| \( n \) odd (\( k/n \leq 2/3 \)) | \( (n - 1)/2 + 2, k/2 + 1 \) | \( (n + 1)/2, (k + 1)/2 \) |
For the uniform prior, it is straightforward to verify that \( P(n', k') = 1 - k'/(n' + 1) \geq 1 - k/(n + 1) = P(n, k) \). Hence, by reasoning analogous to Step 1 in Theorem 3, \( P(n', k') > P(n, k) \) for any decreasing prior. \( \square \)

**Proof of Theorem 4B.** Let \( 2m - 1 \) be the largest sample size satisfying the conditions in the statement of the theorem. By definition, we have \( \frac{\alpha + m + 1}{\alpha + \beta + 2m + 1} \leq \frac{\alpha}{\alpha + \beta} < \frac{\alpha + m}{\alpha + \beta + 2m - 1} \) implying that

\[
\frac{m + 1}{2m + 1} \leq \frac{\alpha}{\alpha + \beta} < \frac{m}{2m - 1}.
\]  \( (2) \)

Thus the interval of \( t \)'s for which sample size \( 2m - 1 \) is conjectured to be optimal is nested in the corresponding interval in Theorem 4. Hence, for any sample size \( n \), the collection of smallest \( k \)'s that trigger the DM to take the action for some \( t \) in the relevant interval is nested in the set \( k[n] \) identified in Equation (1). It follows by the proof of Theorem 4 that sample size \( 2m - 1 \) dominates any finite sample size.

Sample size \( 2m - 1 \) also dominates full revelation by reasoning analogous to the proof of Theorem 4 and observing that \( t > \frac{\alpha}{\alpha + \beta} \geq \frac{m + 1}{2m + 1} > \frac{1}{2} \). \( \square \)

**Proof of Theorem 5B.** For a finite sample size \( n \), the sample \( (n, k) \) triggers the DM to take the action for \( t \) slightly above \( 1/2 \) if \( \frac{\alpha + \beta}{2 \alpha + n} > \frac{1}{2} \). This condition reduces to \( \frac{k}{n} > \frac{1}{2} \) and implies that the DM needs to obtain at least a simple majority of successes in order to take the action. The set of candidates for optimality is therefore similar to that of Theorem 5 and thus the result follows. \( \square \)

**Proof of Proposition 3.** Fix an integer \( n \). A sample size \( n(l) \) is congruent to \( n \) modulo \( (n + 1) \) if \( n(l) = l(n + 1) + n \) and \( l \) is a non-negative integer. For \( n(l) \), the number of successes that achieves \( \frac{n}{n + 1} \) super-majority is \( k(l) = n(l + 1) \). Fix \( t \in \left( \frac{l + n - 1}{2 + n - 1}, \frac{l + n}{2 + n} \right) \). By the statement of Proposition 3, we need to show that optimal sample sizes are of the form \( n(l) \) where \( l \in \{0, 1, \ldots, \left[ \frac{n + 1 - t(n + 2)}{t(n + 1) - n} \right] \} \).

By Step 2 in the proof of Theorem 3, \( P(n(l), k(l)) = 1 - \frac{k(l)}{n(l) + 1} = 1 - \frac{n}{n + 1} \). It is larger than the corresponding value \( 1 - t \) for full revelation because \( t > n/(n + 1) \).
All sample sizes smaller than $n$ are also dominated because the DM will not take the action even when all realizations are successes.

For larger sample sizes, fix two integers $m \geq n$ and $l \geq 0$. The DM takes the action after obtaining the sample $(m, m-l)$ if $\frac{n}{n+1} < t \leq \frac{m-l+1}{m+2}$, i.e. if $m > l(n+1)+n-1$. Because $m$ is an integer, the smallest $m$ that satisfies this condition is $m = n(l) = l(n+1)+n$, which is the candidate for optimality identified above, and $P(m, m-l) = P(n(l), k(l)) = 1 - n/(n+1)$. For larger $m$’s, sample size $m$ is dominated by $n(l)$ because $\frac{m-l}{m+1} > \frac{n}{n+1}$ for $m > n(l)$.

\[ \]  

B Additional lemmas used in the proofs

**Lemma 1.** For $k \geq 2$, $P(n, 1, \frac{1}{n+1}) \geq P(k(n + 1) - 1, k, \frac{1}{n+1})$.

**Proof.** Let $B(n, k, q)$ denote the probability of obtaining up to $k$ successes in $n$ Bernoulli experiments with success probability $q$. Then $B(n, k, q) = 1 - P(n, k+1, q)$. Let $B(n, k) = B(n, k, \bar{q})$ where $\bar{q} = \frac{1}{n+1}$. To prove the lemma, it suffices to show that $B(k(n + 1) - 1, k - 1) \geq B(n, 0)$.

Suppose $k = 2$. Then,

$$B(2(n+1) - 1, 1) - B(n, 0) = (1 - \bar{q})^n \left( \frac{3n + 1}{n+1} (1 - \bar{q})^n - 1 \right)$$

is non-negative if and only if $\frac{3n+1}{n+1} \geq (1 + \frac{1}{n})^n$. This inequality can be verified numerically for $n \leq 6$. And for $n \geq 7$, we observe that $\frac{3n+1}{n+1} > e$ together with the combinatorial inequality $e \geq (1 + \frac{1}{n})^n$ imply the desired inequality.

Suppose $k \geq 3$. By adding and subtracting terms of the form $B((n + 1)i + n, i)$ for $i \in \{1, 2, \cdots, k-1\}$, we obtain:

$$B(k(n + 1) - 1, k - 1) - B(n, 0) = \sum_{i=1}^{k-1} B((n+1)i + n, i) - B((n+1)(i-1) + n, i-1)$$

$$= \sum_{i=1}^{k-1} \Delta_i.$$
We can further expand each $\Delta_i$ by adding and subtracting $n$ expressions of the form $B((n+1)i + n - j - 1, i)$ as follows:

$$\Delta_i = \sum_{j=0}^{n} \left( B((n+1)i + n - j, i) - B((n+1)i + n - j - 1, i) \right)$$

$$\quad + B((n+1)i - 1, i) - B((n+1)(i-1) + n, i-1)$$

$$\quad = \sum_{j=0}^{n} \delta_j + \alpha \quad \text{(3)}$$

To complete the proof, it suffices to show that $\frac{\alpha}{n+1} \geq -\delta_j$ for every $j$. Standard combinatorial identities imply that $-\delta_j = \bar{q}((n+1)i+n-j-1)(1-\bar{q})^{(mi+n-j-1)}$, and $\alpha = ((n+1)i-1)(1-\bar{q})^{(ni-1)}$ by definition. We thus need to show that

$$((n+1)i-1) \geq ((n+1)i+n-j-1) \left( \frac{n}{1+n} \right)^{(n-j)}$$

which holds because the LHS is a constant whereas the RHS increases in $j$ and is equal to the LHS for $j = n$. \qed

**Lemma 2.** Fix an integer $m \geq 1$. If $n \in \mathcal{D}$ then $n = n(l) = (2m+1)l - 2$ for some $l \geq 1$.

**Proof.** Fix $l \geq 1$ and consider sample sizes $n_1 = (2m+1)l - 2$ and $n_2 = (2m+1)(l+1) - 2$. By definition, $\kappa(n_1) = (m+1)l - 1$ and $\kappa(n_2) = (m+1)(l+1) - 1$. Thus, $\frac{\kappa(n_2)}{n_2} < \frac{\kappa(n_1)}{n_1}$. To complete the proof, it suffices to show that $\frac{\kappa(n_1)}{n_1} \leq \frac{\kappa(n)}{n}$ for any $n_1 < n < n_2$.

Any such $n$ satisfies $\kappa(n) = (m+1)l + j$ for some $0 \leq j \leq m$ because $\kappa(n) > \kappa(n_1)$ (this follows from $\frac{\kappa(n_1)}{n_1} \leq \frac{m+1}{2m+1}$) and $\kappa(n) \leq \kappa(n_2)$. Fix $j$ and consider the set of all $n$’s with $\kappa(n) = \kappa(j) = (m+1)l + j$. Since the ratio $\frac{\kappa(n)}{n}$ decreases in $n$, it suffices to verify that $\frac{\kappa(n_1)}{n_1} \leq \frac{\kappa(n)}{n}$ for the largest $n$ in the set. We denote this maximal $n$ as $n(j)$. Then $n(j) = (2m+1)l + 2j - 1$ for $0 \leq j \leq m - 1$ because it satisfies the inequality $\frac{\kappa(j)}{n(j)+1} \leq \frac{m+1}{2m+1} < \frac{\kappa(j)}{n(j)}$, and $n(m) = n_2 - 1$. Verifying that $\frac{\kappa(n_1)}{n_1} \leq \frac{\kappa(n)}{n}$ completes the
Lemma 3. Fix an integer \( m \geq 1 \). For any \( n > 2m - 1 \) and \( k \in k[n] \) there exists \( n_1 \in D \) such that (i) \( n_1 \leq n \), (ii) \( n_1 - \kappa(n_1) \leq n - k \), and (iii) \( \frac{\kappa(n_1)}{n_1} \leq \frac{k}{n} \).

Proof. Fix \( n > 2m - 1 \) and \( k \in k[n] \). Let \( n_1 \leq n \) be a weakly smaller sample size with \( n - k + 1 \) terms in \( P(n, \kappa(n), q) \). (For example, in Figure 6, \( n_1 = 10 \) for \( n = 11 \) and \( k = 8 \).) Such a sample size exists because the number of terms in \( P(n, \kappa(n), q) \) is greater or equal to \( n - k + 1 \), and the number of terms in \( P(n', \kappa(n'), q) \) is either the same or larger by 1 than the number of terms in \( P(n' - 1, \kappa(n' - 1), q) \). By the proof of Lemma 2, the largest \( n \in D \) that is smaller than \( n_1 \) satisfies the desired properties.

Lemma 4. For any integer \( m \geq 1 \), \( \Delta(2) > 0 \).

Proof. Fix integer \( m \geq 1 \). Then \( \Delta(2) \) is equal to:

\[
P \left( 2m - 1, m, \frac{2m + 1}{4m} \right) - P \left( 4m, 2m + 1, \frac{2m + 1}{4m} \right).
\]

We first observe that

\[
P(2m - 1, m, \frac{2m+1}{4m}) = \int_0^{\frac{2m+1}{4m}} P'(2m-1, m, q) dq
\]

\[
= \frac{1}{2} + \int_{\frac{1}{2}}^{\frac{2m+1}{4m}} m\left(\frac{2m-1}{m}\right)q^{m-1}(1-q)^{m-1} dq
\]

\[
\geq (1) \frac{1}{2} + \frac{1}{3}\left(\frac{2m-1}{m}\right)\left(\frac{2m+1}{4m}\right)^{m-1}\left(\frac{2m-1}{4m}\right)^{m-1}
\]

\[
= (2) \frac{1}{2} + \frac{1}{2^{m-1}}\left(\frac{2m}{m}\right)\left(4 - \frac{1}{m^2}\right)^{m-1}
\]

where inequality (1) holds because \( P'(2m-1, m, q) \) decreases in \( q \) over \( \left(\frac{1}{2}, 1\right) \), and we use the identity \( \left(\frac{2m-1}{m}\right) = \frac{1}{2}\left(\frac{2m}{m}\right) \) to obtain equality (2).

We also observe that

\[
P(4m, 2m + 1, \frac{2m+1}{4m}) < (1) P(4m - 2, 2m - 1, \frac{1}{2}) = \frac{1}{2} + \frac{1}{2^{m-1}}\left(\frac{4m-2}{2m-1}\right)
\]

40
where inequality (1) is obtained by applying two times the inequality \( P(n, k, \frac{k}{n}) < P(n - 1, k - 1, \frac{k - 1}{n - 1}) \) (Anderson and Samuels (1967), Theorem 2.2).

To prove the result, it thus suffices to show that

\[
\binom{2m}{m} (4 - \frac{1}{m^2})^{m-1} \geq \binom{4m-2}{2m-1}.
\]

For \( m = 1 \), the inequality holds with equality. For \( m \geq 2 \), we use the identity

\[
\binom{2(n+1)}{n+1} = \frac{4n+2}{2n+1} \binom{2n}{n}
\]

to rewrite the RHS as

\[
\frac{4m-2}{2m-1} = \left( \frac{2m}{m} \right) \left( 4 - \frac{2}{2m-1} \right) \cdot \left( 4 - \frac{2}{2m-2} \right) \cdots \left( 4 - \frac{2}{m+1} \right)
\]

and observe that \( 4 - \frac{1}{m^2} \) is larger than any of the \( m - 1 \) terms on the RHS of the last identity.

\[\square\]

**Lemma 5.** For any integer \( m \geq 1 \), \( \Delta(l) > 0 \) for \( l \geq 3 \).

**Proof.** Fix integers \( m, l \geq 1 \). Then \( \Delta(l) \) is equal to:

\[
P \left( 2m - 1, m \cdot \frac{(m+1)l-1}{(2m+1)l-2} \right) - P \left( (2m+1)l-2, (m+1)l-1, \frac{(m+1)l-1}{(2m+1)l-2} \right).
\]

We observe that

\[
P \left( (2m+1)l-2, (m+1)l-1, \frac{(m+1)l-1}{(2m+1)l-2} \right) \leq (1) \quad P \left( 2ml-2, ml-1, \frac{ml-1}{2ml-2} \right)
\]

\[
\leq (2) \quad \frac{1}{2} + \frac{1}{2} \sqrt{\pi (ml-1+\frac{1}{2})}
\]

where inequality (1) is obtained by applying \( l \) times the inequality \( P(n, k, \frac{k}{n}) < P(n - 1, k - 1, \frac{k - 1}{n - 1}) \), and inequality (2) holds because \( \binom{2n}{n} \leq \frac{4^n}{\sqrt{\pi(n+\frac{1}{2})}} \).

Thus,

\[
\Delta(l) > P \left( 2m - 1, m \cdot \frac{(m+1)l-1}{(2m+1)l-2} \right) - \frac{1}{2} - \frac{1}{2} \sqrt{\pi (ml-1+\frac{1}{2})} \equiv d(l).
\]

41
To complete the proof we treat \( l \) as a continuous variable on \([3, \infty)\) and show that \( d'(l) \geq 0 \) and \( d(3) > 0 \).

The derivative \( d'(l) \) is positive if and only if

\[
\frac{((2m+1)l-2)^2}{(4ml-3)^{3/2} \sqrt{\pi}} \geq \frac{(2m-1)}{2} \left( \frac{(m+1)(l-1)(ml-1)}{(2m+1)(l-2)} \right)^{m-1}.
\] (4)

The RHS is bounded above by \( \frac{2}{\sqrt{\pi (4m+1)}} \). This is because \( \frac{(2m-1)}{m} = \frac{1}{2} \left( \frac{2m}{m} \right) \leq \frac{4^m}{\sqrt{\pi (m+\frac{1}{2})}} \) and \( \frac{ab}{(a+b)^2} \leq \frac{1}{4} \) where \( a = (m+1)l-1 \) and \( b = ml-1 \). The LHS of inequality (4) increases in \( l \). Thus, to verify that \( d'(l) \) is positive, it suffices to verify that \( d'(3) \) is positive, which holds because the inequality \( \frac{((2m+1)3-2)^2}{(12m-3)^{3/2}} \geq \frac{2}{\sqrt{4m+1}} \) holds for \( m \geq 1 \).

It remains to show that \( d(3) > 0 \) for \( m \geq 1 \). For \( m = 1 \) we directly verify the inequality. For \( m \geq 2 \), we start by obtaining a lower bound on

\[
P(2m-1, m, \frac{3m+2}{6m+1}) = \int_0^{\frac{3m+2}{6m+1}} P'(2m-1, m, q) dq
\]

\[
= \frac{1}{2} + \frac{3m}{2(6m+1)} \left( \frac{2m-1}{m} \right) q^{m-1} (1 - q)^{m-1} dq
\]

\[\geq (1) \ \frac{1}{2} + \frac{3m}{2(6m+1)} \left( \frac{2m-1}{m} \right) \left( \frac{3m+2}{6m+1} \right)^{m-1} \left( \frac{3m-1}{6m+1} \right)^{m-1}
\]

\[\geq (2) \ \frac{1}{2} + \sqrt{\frac{3}{\pi (3m+1)}} \frac{3m}{(6m+1)^2} \left( \frac{(6m+4)(6m-2)}{(6m+1)^2} \right)^{m-1}.
\]

Inequality (1) holds because \( P'(2m-1, m, q) \) decreases on \( \left( \frac{1}{2}, \frac{3m+2}{6m+1} \right) \). We use the identity \( \left( \frac{2m-1}{m} \right) = \frac{1}{2} \left( \frac{2m}{m} \right) \) and the inequality \( \left( \frac{2m}{m} \right) \geq \frac{4^m}{\sqrt{\pi (m+\frac{1}{2})}} \) to obtain (2).

Thus, it suffices to prove that

\[
\left( \frac{(6m+4)(6m-2)}{(6m+1)^2} \right)^{m-1} > \frac{6m+1}{3m} \sqrt{\frac{(3m+1)}{5(12m-3)}}.
\]

Because the LHS of the above inequality attains its minimum at \( m = 2 \) and the RHS decreases in \( m \geq 2 \), we complete the proof by verifying it for \( m = 2 \). \( \square \)
References


