mathematics

MDPI

Article

Context-Dependent Criteria for Dirichlet Process in Sequential Decision-Making Problems

Ksenia Kasianova * and Mark Kelbert

Faculty of Economic Sciences, National Research University Higher School of Economics, Myasnitskaya St. 20, 101000 Moscow, Russia; mkelbert@hse.ru
* Correspondence: kkasyanova@hse.ru

Abstract: In models with insufficient initial information, parameter estimation can be subject to statistical uncertainty, potentially resulting in suboptimal decision-making; however, delaying implementation to gather more information can also incur costs. This paper examines an extension of information-theoretic approaches designed to address this classical dilemma, focusing on balancing the expected profits and the information needed to be obtained about all of the possible outcomes. Initially utilized in binary outcome scenarios, these methods leverage information measures to harmonize competing objectives efficiently. Building upon the foundations laid by existing research, this methodology is expanded to encompass experiments with multiple outcome categories using Dirichlet processes. The core of our approach is centered around weighted entropy measures, particularly in scenarios dictated by Dirichlet distributions, which have not been extensively explored previously. We innovatively adapt the technique initially applied to binary case to Dirichlet distributions/processes. The primary contribution of our work is the formulation of a sequential minimization strategy for the main term of an asymptotic expansion of differential entropy, which scales with sample size, for non-binary outcomes. This paper provides a theoretical grounding, extended empirical applications, and comprehensive proofs, setting a robust framework for further interdisciplinary applications of information-theoretic paradigms in sequential decision-making.

Keywords: experimental design; Bayesian analysis; information gain; Dirichlet process; weighted information

MSC: 62B15

check for updates

Citation: Kasianova, K.; Kelbert, M. Context-Dependent Criteria for Dirichlet Process in Sequential Decision-Making Problems. *Mathematics* **2024**, *12*, 3321. https://doi.org/10.3390/math12213321

Academic Editors: Mohamed T. Madi and Mohammad Z. Raqab

Received: 29 July 2024
Revised: 15 September 2024
Accepted: 21 October 2024
Published: 23 October 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The type of problem discussed, often referred to as the "exploration vs. exploitation" dilemma, is commonly considered in sequential decision-making where a trade-off between statistical power and maximizing some reward function needs to be found [1,2]. A method based on information measures, as suggested in [3], elegantly addresses this problem using the information-theoretic paradigm. When applied to clinical trial designs, the information-theoretic approach has demonstrated several advantages in tackling the balance between power and the expected number of successes, offering flexibility in balancing the two competing goals [4].

The potential scope of this method extends beyond clinical trials, e.g., in economic problems, researchers can balance statistical power and expected profit/loss. Consequently, it becomes highly desirable to extend this analysis from the binary case to scenarios involving an arbitrary and finite number of outcomes. The goal of this paper is to extend the maximal entropy method developed in [3] to the non-binary case using the framework of Dirichlet processes.

However, the weighted entropies for Dirichlet distributions and processes have not yet been thoroughly examined, to the best of our knowledge. In this paper, we extend

the method developed in [3] for the binary case is based on the sequential minimization of the main term of the asymptotic expansion (as the sample size $n \to \infty$) of the difference δ between the weighted and standard differential entropies, enabling it to address non-binary problems. The main result, detailed in Theorem 1, specifies the expansion of δ as $n \to \infty$. The leading term of this expansion has a simple and intuitive form, and, in full analogy with the binary case, serves as the objective function of the optimization problem.

In summary, the main contributions of this paper are as follows:

- The maximal entropy method is extended from binary cases to arbitrary finite outcomes;
- We provide one of the initial explorations into weighted entropies for Dirichlet processes, thereby contributing new insights and frameworks for analyses in this area;
- We outline the theoretical foundations, methodological approaches, and potential applications of weighted entropy measures within the context of Dirichlet processes.

To fully grasp the importance and applications of our findings, the reader is encouraged to refer to the comprehensive treatments of reinforcement learning and sequential decision-making in the context of machine learning provided in [5–7]. These texts offer a rigorous exploration of the mathematical and probabilistic foundations that underpin the methods and theories utilized in this paper, thus providing a broader theoretical framework that enhances the understanding and implications of our work.

The rest of the paper proceeds as follows: Section 2 briefly discusses the application of entropy-based criteria to sequential decision-making, followed by introducing the probabilistic setting in terms of Dirichlet processes for sequential decision-making problems. In Section 3, we define an extension to multiple endpoints and derive more general theoretical criteria, which can be explored experimentally. This section involves examining both standard and weighted versions of the exact Rényi and Tsallis entropy criteria for Dirichlet distributions and determining their asymptotics. Technical details of the proofs are provided in Appendices A and B.

2. Main Results

Sequential decision-making is crucial in scenarios marked by incomplete or asymmetric information, where gathering more data often incurs certain costs. These problems frequently require balancing the expected return against the need for more comprehensive outcome information.

In statistical and economic contexts, traditional information measures are employed to gauge uncertainty, helping to differentiate the impacts of various strategies. However, these measures typically overlook the aspect of exploitation. Fortunately, adaptations can be made to these metrics to incorporate specific target outcomes, thus aligning with the dual objectives of maximizing returns while accruing valuable data using the so-called weighted or context-dependent entropy measures [8].

Consider the simple example of two stocks. The financial analyst bids every day that the closing price will be greater than yesterday. He wants to construct a bidding strategy given that in the beginning he knows nothing about the stocks. Each day he can bid one dollar only for one of the stocks and if he is correct, then he gets his dollar back. Therefore, he needs a criterion which will allow him to sequentially decide which stock to choose, considering that he wants both to loose as little money as possible and learn about the strategies enough to stop his search for a better share at some point.

Following the results provided in [4], we note that for the binary endpoint case the conventional context-free formula for the Shannon differential entropy is symmetric and both strategies would correspond to the same amount of uncertainty, leading to a criterion equivalent to fixed randomization. However, the choice of the criterion based on weighted or context-dependent Shannon differential entropy will allow us to balance between the statistical power and the losses explicitly, accounting for the fact that correct guesses are more preferable for the investor. In particular, based on the entropy gain principle,

the minimization of the following asymptotic Shannon criterion δ_S will allow us to achieve the described result:

$$\delta_S = \frac{(\alpha - \gamma)^2}{2\alpha(1-\alpha)} n^{2\kappa-1}, \kappa \in [0.5, 1) \tag{1}$$

where the numerator of the first factor corresponds to a squared distance between the unknown success probability α and the target probability γ, while the second factor relates to the total number of actual and prior observations n and the penalty parameter κ.

The target probability γ should be set close to 1, indicating an interest in the more profitable stock. Note that a small κ penalizes bad guesses that force the investor to switch between the two stocks more often; however, a higher κ will force the investor to stick to one stock for longer if it led to a successful outcome in the past. To estimate the criteria above, the mean of posterior Beta distribution $\hat{\alpha}$ can be plugged in for α.

After new observations are observed, $\hat{\alpha}$ is recalculated and the strategy can be refined sequentially. Notice that the first multiplication term in (1) relates to the exploitation goal, since the minimization of $\hat{\delta}_S$ will force the design to choose a strategy with values of $\hat{\alpha}$, which are closer the target probability γ, whereas the second term relates to the exploration goal, since the design will tend to choose a strategy with fewer observations.

The main result of this paper is the derivation of a similar criterion that can be applied to sequential decision-making with m outcomes. To derive the Shannon criterion, we set $\nu = 1$ in Equation (16), yielding

$$\delta_S = \frac{1}{2}\left(\sum_{j=1}^{m} \frac{\gamma_j^2}{r_j} - 1\right) n^{2\kappa-1}, \tag{2}$$

where γ_j and r_j represent the desired and estimated probabilities of each outcome, respectively.

Interestingly, for the case where $m = 3$, the obtained formula can be expressed in a similar manner to the criterion for the binary case; see the details in Proposition 1 in Section 3.1. In particular, for the setting described above, the expression will look as follows:

$$\delta_S = \frac{r_1(r_2 - \gamma_2)^2 + r_2(r_1 - \gamma_1)^2 - (r_2\gamma_1 - r_1\gamma_2)^2}{2r_1r_2(1 - r_1 - r_2)} n^{2\kappa-1}, \tag{3}$$

showing that δ_S is always non-negative and has a unique minimum equal to 0, when $r_1 = \gamma_1$ and $r_2 = \gamma_2$. Thus, as the estimated probabilities of each outcome approach the desired probabilities the design retains explicit interest in the exploitation goal.

3. Probabilistic Set-Up

In line with the Bayesian framework outlined by [9], the prior knowledge about a different outcome is incorporated via prior distributions within different strategies. As responses are received in a sequential manner, the prior distributions are dynamically updated to reflect our evolving beliefs. Central to our analysis is the anticipation of the expected total return, which corresponds to the mean of the posterior distribution of the probability parameter analyzed incrementally at each step. Following data acquisition, the value function, optimized according to the chosen action policy, is derived post observation.

Based on the entropy maximization principle, the criteria used to address "exploration vs exploitation" via the use of weighted information measures can be derived. The direct application of information measures can detect the differences between strategies which make these methods inherently suitable for exploration; however, the exploitation will be lacking, since such a policy would just focus on maximizing statistical power. A factor that contributes to this shortcoming is the reliance on the probability density of the outcome, which does not consider the nature of the outcome itself. To mitigate this shortcoming, information measures can be further weighted to combine the desired attributes and

objectives, introducing context dependency. More formally, weighting is accomplished via introducing the weight function $\varphi : \mathbb{R} \longrightarrow \mathbb{R}_+$ and applying it to information measures. Given a random observation with the probability density function (PDF) f, we define the integral as $E(f) = \int_0^1 \varphi(p) f(p) dp$. Then, the formula for the weighted Shannon entropy is given as follows:

$$h^{\varphi}(f) = -\int_0^1 \varphi(p) f(p) \ln(f(p)) \mathrm{d}p \tag{4}$$

where $\varphi \equiv 1$ would represent standard (unweighted) entropy measures.

Following the same approach, other weighted entropy measures can be defined as well, i.e., the weighted Rényi entropy

$$R_{\nu}^{\varphi}(f) = \frac{E(f)}{1-\nu} \ln \frac{\int_0^1 \varphi(p)(f(p))^{\nu} dp}{E(f)}, \nu > 0, \nu \neq 1, \tag{5}$$

and the weighted Tsallis entropy

$$T_q^{\varphi}(f) = \frac{1}{q-1} \left(E(f) - \int_0^1 \varphi(f)(f(p))^q dp \right), q \in \mathbb{R}, q \neq 1, \tag{6}$$

where ν and q are the parameters for Rényi and Tsallis entropy, respectively. In limiting cases where $\nu \to 1$ or $q \to 1$, both of these entropy measures converge to the weighted Shannon entropy. To simplify the formulae, we normalize the weight function so that $E(f) = 1$; see (10).

This method was introduced in [3] by applying weighting to Shannon entropy. The weighting in this case is specifying the desired region of the outcomes. Information gain is defined as the difference between Shannon's differential entropy and its weighted counterpart. This method provides an asymptotically unbiased estimator based on the suggested criterion and its properties and applicability are studied in [10]. It was shown that such a design allows for a careful customization of specific statistical properties and targets [11]. Next, Ref. [4] extends previous research by considering binary endpoint cases, but for several other information measures (Tsallis and Rényi entropy, Fisher Information), and compares it with advanced dynamic programming approaches.

The exact criterion which can be used for the decision rules is the difference the between weighted and non-weighted entropy measure, while the asymptotic criterion is just the first term of its asymptotic expansion. No substantial difference was found between using exact or asymptotic criteria for the design. The explicit formulation of information gain, given the target efficiency and penalty parameters, explains how they contribute to exploitation vs. exploration. Here, we refine the approach developed in [4] and provide theoretical extensions of Rényi and Tsallis entropy-based criteria.

The described criteria will be applied to analyze a case with several strategies and multiple outcomes. Let $(\mathcal{X}, \mathcal{B})$ be a measurable Polish space with a $\sigma-$algebra $\mathcal{B}$. The central object of the Bayesian non-parametric statistics is the so-called Dirichlet process $\mathrm{DP}(\boldsymbol{\alpha}, H)$ with $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_m)$, $\sum_{i=1}^{m} \alpha_i = \alpha_0 = H(\mathcal{X})$, where H is a finite measure of $\mathcal{X}$. This is a random measure or probability distribution on the space of $\mathcal{P}(\mathcal{X})$, i.e., the set of all probability distributions on $\mathcal{X}$. The finite measure H is called the base distribution. We emphasize two main properties of the DP and refer the readers to [12] for their proofs and further details.

(i) For any finite partition $\{B_i, i = 1, \ldots, m\}$, $\bigcup_{i=1}^{m} B_i = \mathcal{X}, B_i \bigcap B_j = \emptyset \; \forall i \neq j$ and random realization $X \sim \mathrm{DP}(\boldsymbol{\alpha}, H)$ the random vector is

$$(X(B_1), \ldots, X(B_m)) \sim \mathrm{Dir}(\alpha_1, \ldots, \alpha_m)$$

where $\mathrm{Dir}(\cdot)$ denotes the Dirichlet distribution on a simplex $\mathbf{S}_{m-1} = \{(z_1, \ldots, z_m) \in \mathbb{R}_+^m : z_i \geqslant 0, \sum_{i=1}^{m} z_i = 1\}$ with $\alpha_j = H(B_j)$, $\sum_{i=1}^{m} \alpha_i = H(\mathcal{X}) = \alpha_0$.

(ii) Let $\mathbf{P}$ be a Dirichlet process on $(\mathcal{X}, \mathcal{B})$ with the base distribution H, and let $X_1, \ldots, X_n$ be a sample of size n from $\mathbf{P}$. Then, the conditional distribution of $\mathbf{P}$, given $X_1, \ldots, X_n$, is a Dirichlet process with the base distribution $\widetilde{H} = H + \sum_{i=1}^{n} \delta_{X_i}$, i.e., the updated parameters of the Dirichlet distribution are $\widetilde{\alpha}_j = \widetilde{H}(B_j) = H(B_j) + \sum_{i=1}^{n} X_i(B_j), j = 1, \ldots, m$ with $\widetilde{\alpha}_0 = \alpha_0 + n$.

Let us fix the multiple outcomes $B_1, \cdots, B_m$, i.e., the partition $\mathcal{X}$ and corresponding random measure $\mathbf{P} \in S_{m-1}$. The corresponding random variable $\mathbf{P}$ is modeled as a Dirichlet distribution:

$$\mathbf{P} \sim \mathrm{Dir}(\boldsymbol{p} \mid \alpha_1, \cdots, \alpha_m) = \frac{\Gamma\left(\sum_{j=1}^{m} \alpha_j\right)}{\prod_{j=1}^{m} \Gamma(\alpha_j)} \prod_{j=1}^{m} p_j^{\alpha_j - 1} \tag{7}$$

where $\alpha_j > 0 \; \forall j \in \{1, \ldots, m\}$ and $\sum_{j=1}^{m} \alpha_j = \alpha_0 = H(\mathcal{X}) < \infty$ are the concentration parameters which denote our prior knowledge about each strategy.

Assume that information is sequentially received and n realizations of random measures $X_i = \delta_{x_i}, i = 1, \ldots, n$ are being observed. Then, the discrete random variable $\boldsymbol{y} = (y_1, \ldots, y_m) \in \mathbb{Z}_+^m$ is defined by $y_j = \sharp\{x_i : x_i \in B_j, j = 1, \ldots, m\}$, where $\sharp$ stands for a number of elements and $\boldsymbol{y}$ follows a multinomial distribution:

$$\boldsymbol{y} \sim \mathrm{Multinomial}(\boldsymbol{y}|n, \boldsymbol{p}) = \binom{n}{y_1, \cdots, y_m} \prod_{j=1}^{m} p_j^{y_j}$$

where $\binom{n}{y_1, \cdots, y_m} = \frac{n!}{y_1! \cdot \cdots \cdot y_m!}$ and $\sum_{j=1}^{m} y_j = n$.

Then, the posterior distribution that is obtained via the Bayesian inference is ensured to be of the same exponential family as before, as it is the conjugate prior to the multinomial distribution. So, it becomes the Dirichlet distribution but with updated parameters $\widetilde{\alpha}_j = \widetilde{H}(B_j)$. Hence, the new base distribution has the form $\widetilde{H} = H + \sum_{i=1}^{n} \delta_{x_j}$ and the posterior distribution

$$f(\boldsymbol{p}|\boldsymbol{y}, \boldsymbol{\alpha}, n) \sim \mathrm{Dir}\big(H(B_j) + y_j, j = 1, \ldots, m\big). \tag{8}$$

This follows from property (ii) of Dirichlet processes, as denoted above.

3.1. Entropy Gain for Context-Dependent Dirichlet Processes

Assume that the limits $r_j = \lim_{n \to \infty} \frac{y_j}{n}$ a.e. exist; therefore, the posterior distribution f in the form (8) is concentrated in an asymptotic neighborhood of $\boldsymbol{r} = (r_1, \ldots, r_m) \in \mathbf{S}_{m-1}$ as the sample grows.

Following [3], we define the weight function $\varphi_\kappa^{(n)}$ in a functional form adopted to the Dirichlet distribution:

$$\varphi_\kappa^{(n)}(\boldsymbol{p}) = \Lambda(\boldsymbol{\gamma}, \boldsymbol{\alpha}, \boldsymbol{y}, n, \kappa) \prod_{j=1}^{m} p_j^{\gamma_j n^\kappa}, 0 < \kappa < 1 \tag{9}$$

where $\boldsymbol{\gamma} = (\gamma_1, \cdots, \gamma_m) \in \mathbf{S}_{m-1}$, $\sum_{j=1}^{m} \gamma_j = 1$ is the vector of prespecified features that distributes the weight to areas of importance with a higher γ_j, κ value, which is a sensitivity parameter that penalizes the sample size, and $\Lambda(\boldsymbol{\gamma}, \boldsymbol{\alpha}, \boldsymbol{y}, n, \kappa)$ is a normalizing constant satisfying the condition

$$\int_{S_{m-1}} \varphi_\kappa^{(n)}(\boldsymbol{p}) f(\boldsymbol{p}|\boldsymbol{y}, \boldsymbol{\alpha}, n) \mathrm{d}\boldsymbol{p} = 1. \tag{10}$$

Asymptotically unbiased estimation is preserved; for $\kappa \in (0,1)$ the mean of the weighted posterior is asymptotically equal to $\boldsymbol{r}$:

$$\lim_{n\to\infty} \mathbb{E}[p_j] = \lim_{n\to\infty} \int_{S_{m-1}} p_j \varphi_\kappa^{(n)}(\boldsymbol{p}) f(\boldsymbol{p}|\boldsymbol{y}, \boldsymbol{\alpha}, n) \mathrm{d}\boldsymbol{p} = \lim_{n\to\infty} \frac{\alpha_j + y_j + \gamma_j n^\kappa}{\alpha_0 + n + n^\kappa} = r_j. \tag{11}$$

Theorem 1 (Asymptotic information gains)**.** *Consider the weight function $\varphi_\kappa^{(n)}$ defined in (9) and a random variable* **P** *with PDF f in the form (8). Then, given $\kappa \in (0,1)$ and $\lim_{n\to\infty} \frac{y_j}{n} = r_j$ a.e., the following limits hold for differences in the weighted and standard differential entropies:*

(i) For Rényi entropy,

$$\lim_{n\to\infty} R_\nu(f) - R_\nu^{\varphi_\kappa^{(n)}}(f) + \frac{1}{1-\nu} \omega_R(\boldsymbol{\gamma}, \boldsymbol{r}, n, \kappa, \nu) = 0 \quad a.e. \tag{12}$$

where

$$\omega_R(\boldsymbol{\gamma}, \boldsymbol{r}, n, \kappa, \nu) = \sum_{i=1}^{\lfloor \frac{\kappa}{1-\kappa} \rfloor} \frac{(-1)^{i-1}}{i(i+1)} \left(\sum_{j=1}^{m} \frac{\gamma_j^{i+1}}{r_j^i} - 1 \right) \left(\frac{1}{\nu^i} - 1 \right) n^{\kappa(i+1)-i} + O(n^{\kappa-1}) \tag{13}$$

and $O(n^{\kappa-1})$ denotes the asymptotic order of the remainder terms, indicating that they are of a smaller order than the leading terms as n grows large.

(ii) For Shannon entropy,

$$\lim_{n\to\infty} h(f) - h^{\varphi_\kappa^{(n)}}(f) + \omega_S(\boldsymbol{\gamma}, \boldsymbol{r}, n, \kappa) = 0 \quad a.e. \tag{14}$$

where

$$\omega_S(\boldsymbol{\gamma}, \boldsymbol{r}, n, \kappa) = \sum_{i=1}^{\lfloor \frac{\kappa}{1-\kappa} \rfloor} \frac{(-1)^{i-1}}{i+1} \left(\sum_{j=1}^{m} \frac{\gamma_j^{i+1}}{r_j^i} - 1 \right) n^{\kappa(i+1)-i} + O(n^{\kappa-1}) \tag{15}$$

and $O(n^{\kappa-1})$ denotes the asymptotic order of the remainder terms.

(iii) For Tsallis entropy:

$$\lim_{n\to\infty} \exp(\omega_T(\boldsymbol{\gamma}, \boldsymbol{r}, n, \kappa, q)) T_q(f) - T_q^{\varphi_\kappa^{(n)}}(f) = 0 \quad a.e.$$

where $\omega_T = \omega_R$ if $q = \nu$.

The results are in agreement with the ones presented in [4] in view of the fact that Tsallis differential entropy is the monotone transformation of Rényi's differential entropy.

Based on the derived asymptotic gains, following [3], two types of selection criteria may be suggested: the exact criteria based on exact forms computed in Equations (12) and (13) and relation (22)

$$\overline{\delta}_R = R_\nu(f) - R_\nu^\varphi(f),$$
$$\overline{\delta}_S = h(f) - h^\varphi(f),$$
$$\overline{\delta}_T = T_q(f) - T_q^\varphi(f)$$

and asymptotic criteria based on a leading term of ω or its monotone transformation (for Tsallis entropy) for $\kappa \in [1/2, 1)$

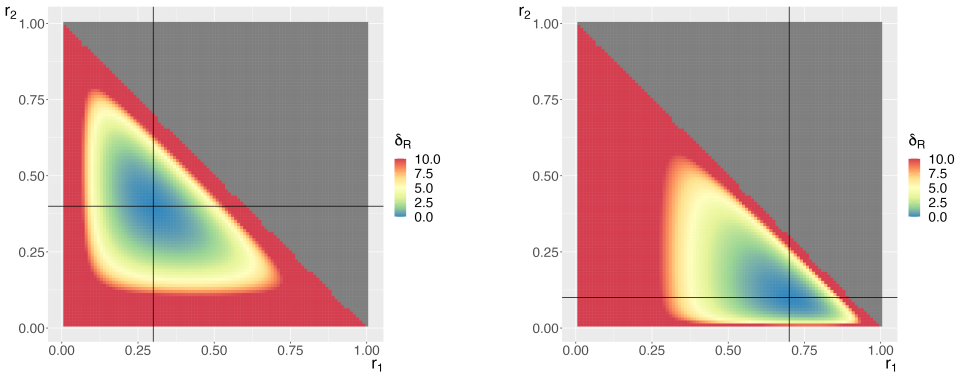
$$\delta_R = \frac{1}{2\nu}\left(\sum_{j=1}^{m} \frac{\gamma_j^2}{r_j} - 1\right) n^{2\kappa-1}. \tag{16}$$

In line with the use of information gain criteria for arm selection, the idea is to choose the arm with the minimum gain due to the non-negativity of information gains, and, for practical purposes, the mean Dirichlet estimator for r_j can be used. The non-negative information gain vanishes at $r_j = \gamma_j$, which makes this whole approach viable.

Proposition 1. *If $m = 3$, Formula (16) takes the form*

$$\delta_R = \frac{1}{2\nu} V n^{2\kappa-1} \tag{17}$$

where $V = \dfrac{r_1(r_2-\gamma_2)^2 + r_2(r_1-\gamma_1)^2 - (r_2\gamma_1 - r_1\gamma_2)^2}{r_1 r_2(1-r_1-r_2)}$. *So, δ_R is always non-negative and has a unique minimum equal to 0 when $r_1 = \gamma_1$ and $r_2 = \gamma_2$; see Figure 1.*



(**a**) (**b**)

Figure 1. The value of asymptotic Rényi criteria δ_R for $\nu = 0.5$, $n = 100$, and (**a**) $\gamma_1 = 0.3, \gamma_2 = 0.4$ or (**b**) $\gamma_1 = 0.7, \gamma_2 = 0.1$. $\delta_R > 8$—red, and $\delta_R < 1$—blue. The minimum of δ_R is at the intercept of two black lines at (γ_1, γ_2).

Proof. In view of the condition that $\sum_{j=1}^{3} \gamma_j = 1$, V is rewritten in terms of the fraction $f_i = \dfrac{\gamma_i}{r_i}$:

$$V = \frac{r_1 r_2^2(1-f_2)^2 + r_1^2 r_2(1-f_1)^2 - r_1^2 r^2(f_1-f_2)^2}{r_1 r_2(1-r_1-r_2)}.$$

Consider the numerator of V. By the triangle inequality,

$$(f_1-1)^2 + (f_2-1)^2 \geqslant (f_1-f_2)^2,$$

$$r_1(f_1-1)^2+r_2(f_2-1)^2 \geqslant r_1r_2(f_1-1)^2+r_1r_2(f_2-1)^2 \geqslant r_1r_2(f_1-f_2)^2,$$
$$r_1^2r_2(f_1-1)^2+r_1r_2^2(f_2-1)^2 \geqslant r_1^2r_2^2(f_1-f_2)^2.$$

This can also be generalized for cases where $m>3$. □

3.2. Proof of Theorem 1

Proof. First, derive the exact expressions for both standard and weighted entropies.

(i) For standard Rényi entropy with $\varphi_\kappa^{(n)}(\boldsymbol{p})\equiv 1$, the derivation is quite straightforward. First, the integral

$$I_1(\boldsymbol{\alpha},\boldsymbol{y},n,\nu)=\int\limits_{S_{m-1}}(f(\boldsymbol{p}|\boldsymbol{y},\boldsymbol{\alpha}))^\nu \mathrm{d}\boldsymbol{p}$$

is computed and evaluated as follows:

$$I_1(\boldsymbol{\alpha},\boldsymbol{y},n,\nu)=\int\limits_{S_{m-1}}(f(\boldsymbol{p}|\boldsymbol{y},\boldsymbol{\alpha}))^\nu \mathrm{d}\boldsymbol{p}=C^\nu\int\limits_{S_{m-1}}\prod_{i=1}^m p_j^{(\alpha_j+y_j)\nu-\nu}\mathrm{d}\boldsymbol{p}$$
$$=C^\nu\frac{\prod\limits_{j=1}^m\Gamma((\alpha_j+y_j)\nu+1-\nu)}{\Gamma((\alpha_0+n)\nu+m-m\nu)}$$

where C is a normalizing constant for (8), defined as

$$C(\boldsymbol{\alpha},\boldsymbol{y})=\frac{\Gamma(\alpha_0+n)}{\prod\limits_{j=1}^m\Gamma(\alpha_j+y_j)}. \tag{18}$$

Then, standard Rényi entropy is expressed as follows:

$$(1-\nu)R_\nu(f(\boldsymbol{p}|\boldsymbol{y},\boldsymbol{\alpha}))=\ln I_1(\boldsymbol{\alpha},\boldsymbol{y},n,\nu)=\nu\left(\ln\Gamma(\alpha_0+n)-\sum_{j=1}^m\ln\Gamma(\alpha_j+y_j)\right)+$$
$$+\sum_{j=1}^m\ln\Gamma(\nu(\alpha_j+y_j)+1-\nu)-\ln\Gamma(\nu(\alpha_0+n)+m-m\nu) \tag{19}$$

To obtain an expression for weighted Rényi entropy with the $\varphi_\kappa^{(n)}(\boldsymbol{p})$ defined in (9), the normalizing constant $\Lambda(\boldsymbol{\gamma},\boldsymbol{\alpha},\boldsymbol{y},n,\kappa)$ is first derived from (11):

$$\Lambda(\boldsymbol{\gamma},\boldsymbol{\alpha},\boldsymbol{y},n,\kappa)=\frac{\prod\limits_{j=1}^m\Gamma(\alpha_j+y_j)\Gamma(\alpha_0+n+n^\kappa)}{\Gamma(\alpha_0+n)\prod\limits_{j=1}^m\Gamma(\alpha_j+y_j+\gamma_jn^\kappa)}.$$

Now, when the weight function is not constant, proceed with computing the following integral:

$$I_2(\boldsymbol{\gamma},\boldsymbol{\alpha},\boldsymbol{y},n,\kappa,\nu)=\int\limits_{S_{m-1}}\varphi_\kappa^{(n)}(p)(f(p|\boldsymbol{y},\boldsymbol{\alpha}))^\nu\mathrm{d}\boldsymbol{p}.$$

Next, retrieve all the constants under the integral sign—one of which is $\Lambda(\boldsymbol{\gamma},\boldsymbol{\alpha},\boldsymbol{y},n,\kappa)$ and the other of which is the normalizing constant from (18) raised to the υ-th power:

$$C^\nu(\boldsymbol{\alpha},\boldsymbol{y})=\frac{\Gamma(\alpha_0+n)^\nu}{\prod\limits_{j=1}^m\Gamma(\alpha_j+y_j)^\nu}.$$

Finally, the integral takes the form of an un-normalized Dirichlet distribution and is easily evaluated:

$$Q(\boldsymbol{\gamma},\boldsymbol{\alpha},\boldsymbol{y},n,\kappa,\nu)=\int\limits_{S_{m-1}}\prod_{j=1}^{m}p_j^{(\alpha_j+y_j)\nu-\nu+\gamma_j n^\kappa}\,\mathrm{d}\boldsymbol{p}=\frac{\prod\limits_{j=1}^{m}\Gamma\big((\alpha_j+y_j)\nu+\gamma_j n^\kappa+1-\nu\big)}{\Gamma\big((\alpha_0+n)\nu+n^\kappa+m-m\nu\big)}.$$

Combining everything together,

$$I_2(\boldsymbol{\gamma},\boldsymbol{\alpha},\boldsymbol{y},n,\kappa,\nu)=\Lambda(\boldsymbol{\gamma},\boldsymbol{\alpha},\boldsymbol{y},n,\kappa)C^\nu(\boldsymbol{\alpha},\boldsymbol{y})Q(\boldsymbol{\gamma},\boldsymbol{\alpha},\boldsymbol{y},n,\kappa,\nu).$$

The weighted Rényi entropy takes the following form:

$$\begin{aligned}(1-\nu)R_\nu^{\varphi(n)}(f(p|\boldsymbol{y},\boldsymbol{\alpha}))&=\ln I_2(\boldsymbol{\alpha},\boldsymbol{y},n,\nu)=\\&=(1-\nu)\left(\sum_{j=1}^{m}\ln\Gamma(\alpha_j+y_j)-\ln\Gamma(\alpha_0+n)\right)\\&-\left(\sum_{i=1}^{m}\ln\Gamma(\alpha_j+y_j+\gamma_j n^\kappa)-\ln\Gamma(\alpha_0+n+n^\kappa)\right)\\&+\sum_{j=1}^{m}\ln\Gamma((\alpha_j+y_j)\nu+\gamma_j n^\kappa+1-\nu)-\ln\Gamma((\alpha_0+n)\nu+n^\kappa+m-m\nu).\end{aligned}\tag{20}$$

To evaluate the limit of the difference δ_R between standard and weighted entropies, recall that $\Gamma(z)=(z-1)!$ and apply the Stirling formula to (19) and (20) when $n\to\infty$:

$$n!=\sqrt{2\pi n}\Big(\frac{n}{e}\Big)^n(1+\theta_n),\text{where } 0<\theta_n<\frac{1}{12n}.$$

After obtaining the asymptotics for standard and weighted entropies, carefully examine each term for the following differences: terms related to 2π; terms related to n and $\ln n$ cancel out. You can see details of the calculations in Appendix A.

What is left are the terms related to n^κ. Hence, one needs to carefully analyze the terms of the type $\sum\limits_{j=1}^{m}(r_j n+\gamma_j n^\kappa)\ln\left(1+\dfrac{\gamma_j n^\kappa}{r_j n\nu}\right)$. First, consider terms of the type $\sum\limits_{j=1}^{m}\gamma_j n^\kappa\ln\left(1+\dfrac{\gamma_j n^\kappa}{r_j n\nu}\right)$. Expanding the logarithmic function, one obtains

$$\begin{aligned}&\sum_{j=1}^{m}\sum_{i=1}^{\infty}\frac{(-1)^{i-1}}{i}\left(\frac{\gamma_j^{i+1}}{(r_j\nu)^i}\right)n^{\kappa(i+1)-i}-\sum_{j=1}^{m}\sum_{i=1}^{\infty}\frac{(-1)^{i-1}}{i}\left(\frac{\gamma_j^{i+1}}{r_j^i}\right)n^{\kappa(i+1)-i}+\\&+\sum_{i=1}^{\infty}\frac{(-1)^{i-1}}{i}n^{\kappa(i+1)-i}-\sum_{i=1}^{\infty}\frac{(-1)^{i-1}}{i}\frac{1}{\nu^i}n^{\kappa(i+1)-i}.\end{aligned}\tag{21}$$

Combining everything together,

$$\sum_{i=1}^{\infty}\frac{(-1)^{i-1}}{i}\left(\sum_{j=1}^{m}\frac{\gamma_j^{i+1}}{r_j^i}-1\right)\left(\frac{1}{\nu^i}-1\right)n^{\kappa(i+1)-i}.$$

Second, by analyzing terms of the type $\sum_{j=1}^{m} r_j n \ln\left(1+\frac{\gamma_j n^{\kappa}}{r_j n \nu}\right)$, one obtains

$$\sum_{j=1}^{m}\sum_{i=1}^{\infty}\frac{(-1)^{i-1}}{i}\frac{\gamma_j^i}{(r_j\nu)^{i-1}}n^{i(\kappa-1)+1}-\sum_{j=1}^{m}\sum_{i=1}^{\infty}\frac{(-1)^{i-1}}{i}\frac{\gamma_j^i}{r_j^{i-1}}n^{i(\kappa-1)+1}+$$

$$\sum_{i=1}^{\infty}\frac{(-1)^{i-1}}{i}n^{i(\kappa-1)+1}-\sum_{i=1}^{\infty}\frac{(-1)^{i-1}}{i}\frac{1}{\nu^{i-1}}n^{i(\kappa-1)+1},$$

which when summed together is equal to

$$\sum_{i=1}^{\infty}\frac{(-1)^{i-1}}{i}\left(\sum_{j=1}^{m}\frac{\gamma_j^i}{r_j^{i-1}}-1\right)\left(\frac{1}{\nu^{i-1}}-1\right)n^{i(\kappa-1)+1}.$$

Notice that for $i=1$, the expression above vanishes and the indexation can be changed, so that the summation starts from $i=2$, which is equivalent to $i \leftarrow i+1$:

$$-\sum_{i=1}^{\infty}\frac{(-1)^{i-1}}{i+1}\left(\sum_{j=1}^{m}\frac{\gamma_j^{i+1}}{r_j^{i}}-1\right)\left(\frac{1}{\nu^{i}}-1\right)n^{\kappa(i+1)-i}.$$

Now, combining everything we get ω_R in (13). The upper bound of summation is equal to $\lfloor\frac{\kappa}{1-\kappa}\rfloor$ due to the fact that after that value terms in ω_R at most become $O(n^{\kappa-1})$.

The obtained results agree with the ones presented in [13] for the Rényi entropy of a random variable coming from a Beta distribution with $m=2$.

(ii) For Shannon entropy:

The limiting value of $R_\nu(f)$ as $\nu \to 1$ is the Shannon entropy, hence, by L'Hospital rule:

$$\omega_S = \lim_{\nu\to 1}\frac{1}{1-\nu}\omega_R(\boldsymbol{\gamma},\boldsymbol{r},n,\kappa,\nu) =$$

$$= \lim_{\nu\to 1}\frac{1}{1-\nu}\sum_{i=1}^{\lfloor\frac{\kappa}{1-\kappa}\rfloor}\left(\frac{1-\nu^i}{\nu^i}\right)\frac{(-1)^{i-1}}{i(i+1)}\left(\sum_{j=1}^{m}\frac{\gamma_j^{i+1}}{r_j^i}-1\right)n^{\kappa(i+1)-i} =$$

$$= \lim_{\nu\to 1}\sum_{i=1}^{\lfloor\frac{\kappa}{1-\kappa}\rfloor} i\nu^{i-1}\frac{(-1)^{i-1}}{i(i+1)}\left(\sum_{j=1}^{m}\frac{\gamma_j^{i+1}}{r_j^i}-1\right)n^{\kappa(i+1)-i} =$$

$$\sum_{i=1}^{\lfloor\frac{\kappa}{1-\kappa}\rfloor}\frac{(-1)^{i-1}}{i+1}\left(\sum_{j=1}^{m}\frac{\gamma_j^{i+1}}{r_j^i}-1\right)n^{\kappa(i+1)-i}.$$

(iii) For Tsallis entropy:

Tsallis entropy is connected to Rényi entropy by the following relation:

$$T_q(f) = \frac{1}{q-1}\left(1-\exp((1-q)R_q(f))\right). \tag{22}$$

Thus, the results from Equations (12) and (13) can be used to define exact the Tsallis entropies both for the standard and weighted counterpart.

Deriving asymptotics for Tsallis entropy is straightforward via Formula (22), which concludes the proof of the limiting value of the difference between the standard and weighted entropies. □

4. The Application of Dirichlet Processes in Economic Problems

Here, we argue that the framework of Dirichlet processes is a valuable mathematical tool for addressing certain problems encountered in economics. To illustrate this, consider a sequential decision model in an economic context.

A factory introduces a new line of chocolate candy and must determine whether to sell the candy in small packages, big boxes, or gift boxes. After a month of sales, the factory evaluates how many of each type of box were sold. The decision-maker must then decide whether it is profitable to continue producing these candies, and, if so, in what quantities. If the decision is made hastily, there is a risk of error; however, if the decision is delayed, the factory may incur losses due to missed opportunities.

To formalize this, consider the set $\mathcal{X}$ of all possible profits. To transition to a Dirichlet process framework with a finite number of outcomes, a researcher needs to choose an initial partition of $\mathcal{X}$ into m possible compartments and associate with this partition the measure $\mathbf{P}$ that describes their prior beliefs. For example, suppose the researcher believes there are three possible outcomes: {"Excellent", "Good", "Bad"}. If they assume that any strategy could equally likely produce any of these outcomes in the absence of prior knowledge, they might partition the outcomes as follows: a "Bad" outcome corresponds to a profit $\Pi \in (-\infty, 0)$, a "Good" outcome to $\Pi \in [0, 100)$, and an "Excellent" outcome to $\Pi \in [100, \infty)$.

Because of the properties of the Dirichlet process, each realization based on sequentially obtained information about $\mathbf{P}$ will be governed by the Dirichlet distribution (7). This allows for a systematic and probabilistically sound approach to updating beliefs and making decisions based on observed data, thus providing a robust framework for tackling this problem within the domain of sequential decision-making in economics. Below are two examples used to illustrate the usefulness of this approach.

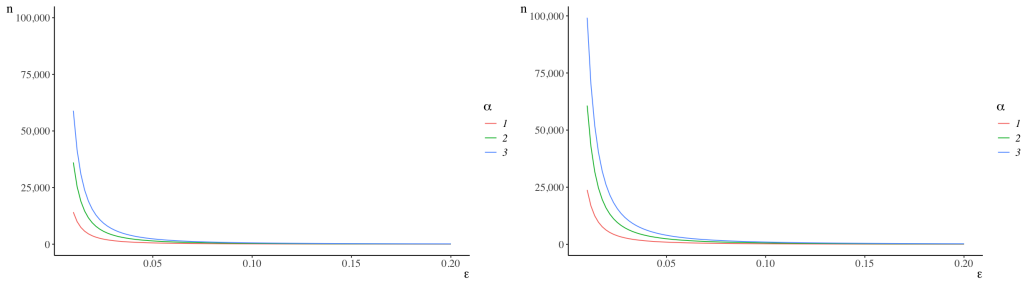
Proposition 2. *Consider two strategies with the true Dirichlet distribution* $\mathbf{P} \sim \mathrm{Dir}(\boldsymbol{\alpha_1})$ *and* $\mathbf{Q} \sim \mathrm{Dir}(\boldsymbol{\alpha_2})$*, where* $\boldsymbol{\alpha_1} = (\alpha_{11}, \alpha_{12}, \alpha_{13}), \boldsymbol{\alpha_2} = (\alpha_{21}, \alpha_{22}, \alpha_{23})$*,* $\alpha_{11} = \alpha_{21}$*,* $\alpha_{22} = \alpha_{12} + \varepsilon$*,* $\alpha_{23} = \alpha_{13} - \varepsilon$*, and* ε *is a constant indicating a small difference between two strategies. Suppose, in this case, there is no Bayesian updating; the observations are independent and the parameters of the Dirichlet distribution represent the true values of the parameters.*

Employing the total variation distance d_{TV} *between* $\mathbf{P}$ *and* $\mathbf{Q}$ *as a measure of confidence, the number of observations conducted for each strategy so that two distributions can be differentiated with confidence* $\Delta = d_{\mathrm{TV}}$ *would be*

$$n \geqslant -\frac{\ln(1 - d_{\mathrm{TV}}^2(\mathbf{P}, \mathbf{Q}))}{\mathrm{KL}[\mathbf{P}||\mathbf{Q}]}. \tag{23}$$

The details of the calculations can be found in Appendix B.1. Suppose that we have one strategy with three equally likely outcomes $\boldsymbol{\alpha_1} = (\alpha, \alpha, \alpha)$. For the second strategy, "Good" outcomes are slightly more probable and "Bad" outcomes are slightly less probable, with $\mathbf{Q} \sim \mathrm{Dir}(\alpha, \alpha + \varepsilon, \alpha - \varepsilon)$.

In Figure 2, observe that the required number of observations to differentiate between the two strategies increases as the confidence level Δ is increased from $d_{\mathrm{TV}} = 0.95$ to $d_{\mathrm{TV}} = 0.99$. The difference in the required number of observations is more pronounced for smaller values of ε, but as the distance between the two distributions becomes larger, the required number tends toward one, given that we need the smallest integer greater than n. Additionally, the higher the values of α, the more equal the probabilities of the outcomes, and the larger the sample size required. This helps decision-makers balance the trade-off between their confidence in distinguishing strategies and the resources spent on data collection.



(**a**) (**b**)

Figure 2. The required number of observations n as a function of ε where $\mathbf{P} \sim \mathrm{Dir}(\alpha, \alpha, \alpha)$ and $\mathbf{Q} \sim \mathrm{Dir}(\alpha, \alpha + \varepsilon, \alpha - \varepsilon)$, with (**a**) $d_{\mathrm{TV}} = 0.95$ and (**b**) $d_{\mathrm{TV}} = 0.99$ for the following values of parameter α: $\alpha = 1$ (red line), $\alpha = 2$ (green line), and $\alpha = 3$ (blue line).

Proposition 3. *Consider two experiments* $\mathbf{P}_n$ *and* $\mathbf{Q}_n$ *producing the data* $\boldsymbol{x} = (x_1, x_2, x_2)$ *and* $\boldsymbol{y} = (y_1, y_2, y_2)$*, respectively. Assuming that the number of prior observations is insignificantly small in comparison to number of posterior observations, the Dirichlet priors can be set to be the same* $\mathrm{Dir}(\alpha_1, \alpha_2, \alpha_3)$.

Suppose that after collecting n observations, one observes that $y_1 = x_1 + z, y_2 = x_2 - z, y_3 = x_3$*. To determine how large the difference z should be to conclude that the variation distance between distributions* $\mathbf{P}_n$ *and* $\mathbf{Q}_n$ *is greater than the* Δ *defined by* d_{TV}*, the following criteria can be used:*

$$\mathrm{KL}[\mathbf{P}||\mathbf{Q}] \geqslant -\ln(1 - d_{\mathrm{TV}}^2(\mathbf{P}, \mathbf{Q})). \tag{24}$$

The details of these calculations can be found in Appendix B.2. Suppose that a company wants to compare two different advertising strategies. In half of their shops, after some time, they received the posterior probability of the {"Excellent", "Good", "Bad"} outcomes that was $\mathrm{Dir}(250 + E, 250 + G, 100)$. In the second half of the shops, where a different strategy was employed, they observed the posterior probability $\mathrm{Dir}(250 + E + z, 250 + G - z, 100)$. Note that we start with 250 for the sake of visualization and so that the values of the parameters are "large enough" that prior observations do not matter.

The values for the corresponding plot with the minimal difference z, showing the difference between the "Excellent" and "Good" outcomes, are given in Figure 3 and were processed by numerical methods. Here, the total number of observations for the first shop $x_1 = 250 + E$ and $x_2 = 250 + G$ was in the range in $[250, 300]$, with $E + G \leq 100$. Notice that the smaller the difference between the first and second outcomes (i.e., $x_1 \approx x_2$), the larger z should be. In some cases, if $x_1 + z < 0$ or $x_2 - z < 0$, it is not be possible to observe a difference.

Both propositions illustrate the effectiveness of the Dirichlet process framework in integrating prior beliefs and updating them as new data are received. They provide quantitative criteria for determining the number of observations required to distinguish between competing strategies or experiments. This is particularly crucial in sequential decision-making, where decisions are made based on the accumulation of data over time. Notably, these criteria can be reformulated using the weighted Kullback–Leibler divergence and the total weighted variation (see [8]).

To elucidate the practical application of sequential experiments using the weighted asymptotic Shannon criterion derived earlier (see Formula (2)), consider the following example. A company may employ three different strategies over the course of a year, each with three potential outcomes: {"Excellent", "Good", "Bad"}. For simplicity, assume the outcomes yield profits of −100, +50, and +100, respectively.

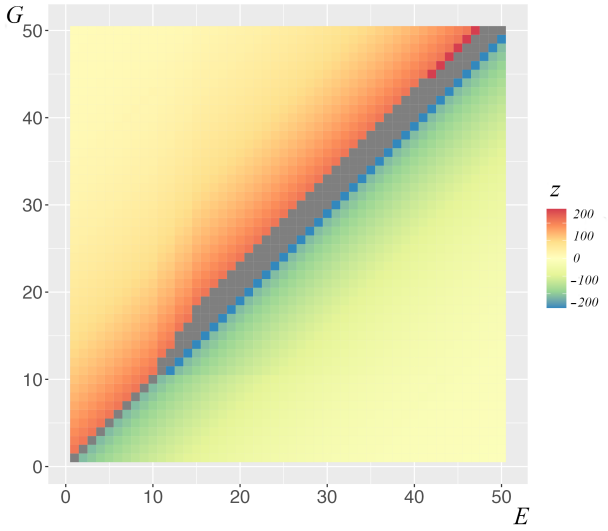


Figure 3. Minimal difference $z = y_1 - x_1$ in number of observations so that $\mathbf{P} \sim \text{Dir}(x_1, x_2, 100)$ and $\mathbf{Q} \sim \text{Dir}(x_1 + z, x_2 - z, 100)$ can be distinguished with 0.99 confidence given that $d_{\text{TV}} > 0.99$. Here, $x_1 = 250 + E$ and $x_2 = 250 + G$, with $E, G \in [0, 50]$; values of z around zero—yellow squares, $z > 200$—red squares, $z < 200$—blue squares, and $x_1 + z < 0$ or $x_2 - z < 0$—gray squares.

The first strategy follows the Dirichlet distribution $\mathbf{S1} \sim \text{Dir}(\mathbf{0.1, 0.8, 0.1})$, the second strategy follows $\mathbf{S2} \sim \text{Dir}(\mathbf{0.2, 0.7, 0.1})$, and the third strategy follows $\mathbf{S3} \sim \text{Dir}(\mathbf{0.3, 0.6, 0.1})$. Outcomes are observed at the end of each day, and switching between strategies incurs no costs, with previous choices not affecting current profits.

In weighted entropy designs, the experiment begins with the strategy that minimizes the criterion (2). To calculate the value of δ_S, the stakeholder must choose the following parameters: (1) the target probability γ, in this case $(0.65, 0.34, 0.01)$, indicating no interest in "Bad" outcomes, and a higher interest in "Excellent" outcomes than in "Good" outcomes; (2) a penalty parameter κ, e.g., $\kappa = 0.5$, indicating an interest in exploitation, or $\kappa = 0.999$, an indicating interest in exploration; and (3) a prior distribution, say $\text{Dir}(\mathbf{1, 1, 1})$, indicating one prior observation of each outcome for each arm.

Upon observing the outcome of the chosen strategy, the total number of responses is updated, and the plug-in estimator r_j and information gains are recalculated. The strategy for the following day is determined by the minimum value of the criterion. This procedure is repeated until the total number of observations is attained, which in this case is 365; one for each day of the year.

Figures 4 and 5 present the results of a simulated experiment with 10,000 iterations using the aforementioned parameters. For comparison, a fixed randomization reference is included, wherein each strategy has an equal probability (1/3) of being chosen daily.

Figure 4 shows that for both selected values of the penalty parameter κ, the weighted entropy design predominantly favors Strategy 3, as it has the highest probability of yielding an "Excellent" outcome. Strategies 1 and 2 are chosen less frequently compared to fixed randomization. Specifically, when $\kappa = 0.5$, the design ceases to explore Strategy 1, identifying it as the furthest from the desired distribution. Figure 5 shows that the total average profit from all three strategies accumulates over time to be higher than with fixed randomization. By the end of the year, the total profits are 2.91 times higher for $\kappa = 0.5$ and 2.85 times higher for $\kappa = 0.999$ than for fixed randomization.

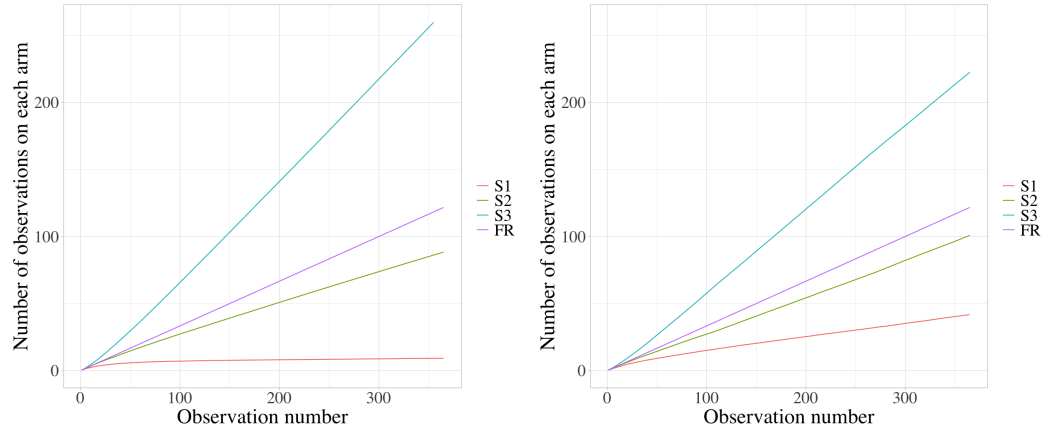


Figure 4. Average number of observation days for each strategy: red line—Strategy 1, with **S1** $\sim$ Dir$(\mathbf{0.1, 0.8, 0.1})$; green line—Strategy 2, with **S2** $\sim$ Dir$(\mathbf{0.2, 0.7, 0.1})$; blue line—Strategy 3, with **S3** $\sim$ Dir$(\mathbf{0.3, 0.6, 0.1})$; violet line—fixed randomization. Target probability $\gamma = (0.65, 0.34, 0.01)$. Value of penalty parameter: (**a**) $\kappa = 0.5$, (**b**) $\kappa = 0.999$.

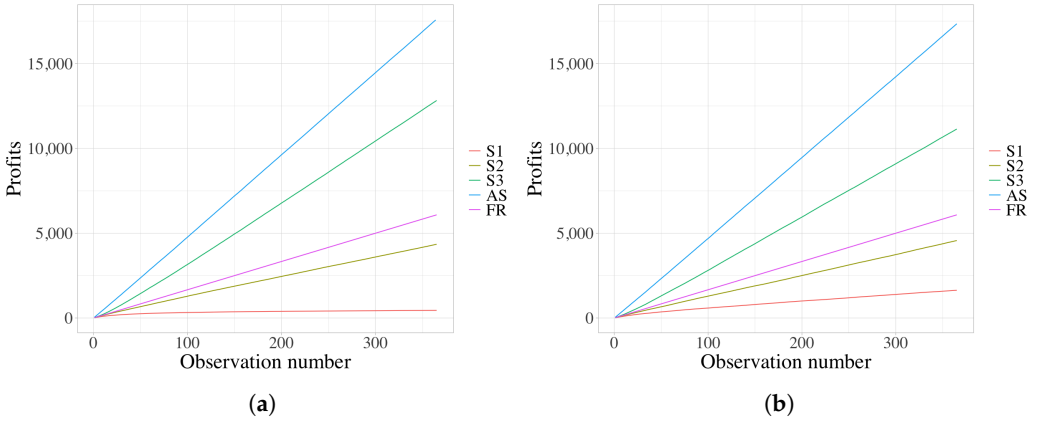


Figure 5. Average accumulated profits for each strategy: red line—Strategy 1, with **S1** $\sim$ Dir$(\mathbf{0.1, 0.8, 0.1})$; olive line—Strategy 2, with **S2** $\sim$ Dir$(\mathbf{0.2, 0.7, 0.1})$; green line—Strategy 3, with **S3** $\sim$ Dir$(\mathbf{0.3, 0.6, 0.1})$; blue line—total profits for asymptotic Shannon design, i.e., sum of profits for Strategy 1, 2, and 3; violet line—fixed randomization. Target probability $\gamma = (0.65, 0.34, 0.01)$. Value of penalty parameter: (**a**) $\kappa = 0.5$, (**b**) $\kappa = 0.999$. Profit outcomes $\{-100,50,100\}$.

In conclusion, this example demonstrates the potential of using the Dirichlet process framework and weighted entropy measures in sequential decision-making. The results show a significant improvement in accumulated profits when strategies are chosen based on the weighted asymptotic Shannon criterion compared to fixed randomization.

Possible applications of this methodology extend across various fields and include financial markets, where investors often face the challenge of deciding between multiple investment strategies. By using this methodology, they can sequentially update their strategies based on accumulating data, thus potentially improving their investment returns and minimizing risks. This approach can also be applied to data-driven business decisions in continuous A/B testing, where it is crucial to quickly achieve optimal results, especially in case of scarce data when launching a new product.

However, it is important to note that the results may vary significantly for different parameter values. This simple example is intended to illustrate the potential application of the proposed methodology. Conducting a more sophisticated computational experiment, while beyond the scope of this paper, would provide further insights.

Furthermore, it is essential to compare this methodology with other established methods to fully understand its strengths and weaknesses. Future work should include comparisons with alternative techniques such as the Multi-arm Bandit models, Upper Confidence Bound method, and the Bayesian Optimization approach. These comparisons would help in validating the effectiveness and reliability of the proposed methodology.

5. Discussion

This research adapts information-theoretic methods to address the "exploration versus exploitation" dilemma through the innovative application of weighted entropy measures. We extend existing methodologies to accommodate multiple outcome categories using the Dirichlet process framework. This approach allows for the modeling of multinomial outcomes with Rényi and Tsallis information measures, demonstrating their theoretical feasibility and practical potential.

While the core of this study is theoretical, it provides foundational insights and establishes a comprehensive framework for further interdisciplinary applications of information-theoretic principles in sequential decision-making. From a practical perspective, generalizing the results to accommodate an arbitrary number of outcomes, for both exact and asymptotic criteria, results in only a linear increment in computational demands. This observation, predicated on the functional forms of Rényi and Tsallis entropies, underscores the practical feasibility of the extended methodologies.

Our findings show promise for applications in financial markets, clinical research, and in business scenarios such as continuous A/B testing, where the rapid achievement of optimal results is critical, especially with limited data available.

Future research is essential to fully explore the potential of this methodology. Comparative studies using established methods like Multi-arm Bandit models and Bayesian Optimization approaches are necessary to validate its effectiveness and thoroughly investigate the behavior of various parameters within simulated environments. Empirical validation in multi-arm settings will help investigate the method's parameter behavior in simulated environments. Additionally, deriving the Fisher Information for multinomial outcomes and exploring its asymptotics will further solidify the theoretical basis and practical applicability of our approach.

Author Contributions: Conceptualization, M.K.; methodology, M.K.; software, K.K.; validation, K.K. and M.K.; formal analysis, K.K. and M.K.; investigation, K.K.; writing—original draft preparation, K.K.; writing—review and editing, K.K. and M.K.; visualization, K.K.; supervision, M.K.; project administration, M.K.; funding acquisition, M.K. All authors have read and agreed to the published version of the manuscript.

Funding: K.K. and M.K. were supported by the Russian Science Foundation, grant number N 23-21-00052.

Data Availability Statement: The dataset used is available on request from the authors.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

DP	Dirichlet process
PDF	Probability density function

Appendix A. Details of the Proof of Theorem 1

Recall the formula for standard Rényi entropy (19) and expand it using the Stirling approximation as $n \to \infty$ for the Gamma function:

$$(1-\nu)R_\nu(f(\boldsymbol{p}|\boldsymbol{y},\boldsymbol{\alpha})) = \ln I_1(\boldsymbol{\alpha},\boldsymbol{y},n,\nu) = \nu\left(\ln\Gamma(\alpha_0+n) - \sum_{j=1}^{m}\ln\Gamma(\alpha_j+y_j)\right) +$$

$$+\sum_{j=1}^{m}\ln\Gamma(\nu(\alpha_j+y_j)+1-\nu)-\ln\Gamma(\nu(\alpha_0+n)+m-m\nu). \tag{A1}$$

As a result, one obtains

$$\begin{gathered}\ln I_1=\tfrac{\nu}{2}\ln 2\pi+\tfrac{\nu}{2}\ln(\alpha_0+n-1)+\nu(\alpha_0+n-1)\cdot\ln(\alpha_0+n-1)-\nu(\alpha_0+n-1)-\\ -\nu\textstyle\sum_{j=1}^{m}\Big[\frac{1}{2}\ln 2\pi+\frac{1}{2}\ln\big(\alpha_j+nr_j-1\big)+\big(\alpha_j+nr_j-1\big)\cdot\ln\big(\alpha_j+nr_j-1\big)-\\ -\big(\alpha_j+nr_j-1\big)\Big]+\textstyle\sum_{j=1}^{m}\Big[\frac{1}{2}\ln 2\pi+\frac{1}{2}\ln\big(\alpha_j\nu+nr_j\nu-\nu\big)+\big(\alpha_j\nu+nr_j\nu-\nu\big)\cdot\\ \cdot\ln\big(\alpha_j\nu+nr_j\nu-\nu\big)-\big(\alpha_j\nu+nr_j\nu-\nu\big)\Big]-\Big[\frac{1}{2}\ln 2\pi+\frac{1}{2}\ln(\alpha_0\nu+n\nu+m-m\nu-1)+\\ +(\alpha_0\nu+n\nu+m-m\nu-1)\cdot\ln(\alpha_0\nu+n\nu+m-m\nu-1)-\\ -(\alpha_0\nu+n\nu+m-m\nu-1)\Big]+O(n^{\kappa-1})\simeq A+B+C+D\end{gathered}$$

where A stands for the terms with 2π, B indicates linear terms, C indicates terms with $\ln n$, and D stands for terms with $n\ln n$.

Consider the terms with 2π :

$$\begin{gathered}A=\frac{\nu}{2}\ln 2\pi-\frac{\nu m}{2}\ln 2\pi+\frac{m}{2}\ln 2\pi-\frac{1}{2}\ln 2\pi=\\ =\frac{1}{2}\ln 2\pi(\nu-\nu m+m-1)=\frac{1}{2}(m-1)(1-\nu)\ln 2\pi.\end{gathered}$$

Consider the linear terms:

$$B=-\nu(\alpha_0+n-1)+\nu\sum_{j=1}^{m}\big(\alpha_j+nr_j-1\big)-\sum_{j=1}^{m}\big(\alpha_j\nu+nr_j\nu-\nu\big)+$$

$$+(\alpha_0\nu+n\nu+m-m\nu-1)=(-\nu\alpha_0-\nu n+\nu)+(\nu+\nu n-\nu m)-(\nu+\nu n-\nu m)+$$

$$(\alpha_0\nu+n\nu+m-m\nu)=(m-1)(1-\nu).$$

Consider the terms with $\ln n$:

$$C=\frac{\nu}{2}\ln(\alpha_0+n-1)-\frac{\nu}{2}\sum_{j=1}^{m}\ln\big(\alpha_j+nr_j-1\big)+\frac{1}{2}\sum_{j=1}^{m}\ln\big(\alpha_j\nu+nr_j\nu-\nu\big)-$$

$$-\frac{1}{2}\ln(\alpha_0\nu+n\nu+m-m\nu-1)\simeq\frac{\nu}{n}\ln n-\frac{\nu m}{2}\ln n-\frac{\nu}{2}\sum_{j=1}^{m}\ln r_j+\frac{m}{2}\ln n+\frac{1}{2}\sum_{j=1}^{m}\ln r_j+$$

$$+\frac{m}{2}\ln\nu-\frac{1}{2}\ln n-\frac{1}{2}\ln\nu=\frac{1}{2}(1-\nu)\sum_{j=1}^{m}\ln r_j+\frac{1}{2}(1-\nu)(m-1)\ln n+\frac{1}{2}(m-1)\ln\nu$$

Consider the terms with $n\ln n$:

$$D=\nu(\alpha_0+n-1)\ln n-\nu\sum_{j=1}^{m}\big(\alpha_j+nr_j-1\big)\ln nr_j+\sum_{j=1}^{m}\big(\alpha_j\nu+nr_j\nu-\nu\big)\ln nr_j\nu-$$

$$-(\alpha_0\nu+n\nu+m-m\nu-1)\ln n\nu=\nu\alpha_0\ln n+\nu n\ln n-\nu\ln n\nu\ln n-\nu n\ln n+$$

$$+\nu m\ln n-\nu\sum_{j=1}^{m}\alpha_j\ln r_j-\nu n\sum_{j=1}^{m}r_j\ln r_j+\nu\sum_{j=1}^{m}\ln r_j+\nu\ln n+\nu n\ln n-$$

$$-\nu m\ln n+\nu\sum_{j=1}^{m}\alpha_j\ln r_j+n\nu\sum_{j=1}^{m}r_j\ln r_j-\nu\sum_{j=1}^{m}\ln r_j+\nu\ln\nu+\nu n\ln\nu-\nu m\ln\nu-$$

$$-\alpha_0\nu\ln n-\alpha_0\nu\ln\nu-n\nu\ln n-n\nu\ln\nu-m(1-\nu)\ln n-m(1-\nu)\ln\nu+\ln n+\ln\nu=$$

$$=(\nu-\alpha_0\nu-m+1)\ln\nu+(\nu-1)(m-1)\ln n.$$

Recall the formula for weighted Rényi entropy (20) and expand it using the Stirling approximation as $n \to \infty$ for the Gamma function:

$$(1-\nu)R_\nu^{\varphi^{(n)}}(f(p|y,\alpha)) = \ln I_2(\alpha, y, n, \nu) = (1-\nu)\left(\sum_{j=1}^{m} \ln\Gamma(\alpha_j + y_j) - \ln\Gamma(\alpha_0 + n)\right) -$$

$$-\left(\sum_{i=1}^{m} \ln\Gamma(\alpha_j + y_j + \gamma_j n^\kappa) - \ln\Gamma(\alpha_0 + n + n^\kappa)\right) +$$

$$+\sum_{j=1}^{m} \ln\Gamma((\alpha_j + y_j)\nu + \gamma_j n^\kappa + 1 - \nu) - \ln\Gamma((\alpha_0 + n)\nu + n^\kappa + m - m\nu)$$

As a result, one obtains

$$\begin{aligned}
\ln I_2 = (1-\nu)\sum_{j=1}^{m}\Big[&\tfrac{1}{2}\ln 2\pi + \tfrac{1}{2}\ln\big(\alpha_j + nr_j - 1\big) + \big(\alpha_j + nr_j - 1\big)\cdot\ln\big(\alpha_j + nr_j - 1\big) - \\
&-\big(\alpha_j + nr_j - 1\big)\Big] - (1-\nu)\Big[\tfrac{1}{2}\ln 2\pi + \tfrac{1}{2}\ln(\alpha_0 + n - 1) + (\alpha_0 + n - 1)\ln(\alpha_0 + n - 1) - \\
&-(\alpha_0 + n - 1)\Big] - \sum_{j=1}^{m}\Big[\tfrac{1}{2}\ln 2\pi + \tfrac{1}{2}\ln\big(\alpha_j + nr_j + \gamma_j n^\kappa - 1\big) + \big(\alpha_j + nr_j + \gamma_j n^\kappa - 1\big)\cdot \\
&\cdot\ln\big(\alpha_j + nr_j + \gamma_j n^\kappa - 1\big) - \big(\alpha_j + nr_j + \gamma_j n^\kappa - 1\big)\Big] + \Big[\tfrac{1}{2}\ln 2\pi + \\
&+\tfrac{1}{2}\ln(\alpha_0 + n + n^\kappa - 1) + (\alpha_0 + n + n^\kappa - 1)\ln(\alpha_0 + n + n^\kappa - 1) - \\
&-(\alpha_0 + n + n^\kappa - 1)\Big] + \sum_{j=1}^{m}\Big[\tfrac{1}{2}\ln 2\pi + \tfrac{1}{2}\ln\big(\alpha_j\nu + nr_j\nu + \gamma_j n^\kappa - \nu\big) + \\
&+\big(\alpha_j\nu + nr_j\nu + \gamma_j n^\kappa - \nu\big)\ln\big(\alpha_j\nu + nr_j\nu + \gamma_j n^\kappa - \nu\big) - \\
&-\big(\alpha_j\nu + nr_j\nu + \gamma_j n^\kappa - \nu\big)\Big] - \Big[\tfrac{1}{2}\ln 2\pi + \tfrac{1}{2}\ln(\alpha_0\nu + n\nu + n^\kappa + m - m\nu - 1) + \\
&+(\alpha_0\nu + n\nu + n^\kappa + m - m\nu - 1)\ln(\alpha_0\nu + n\nu + n^\kappa + m - m\nu - 1) + \\
&-(\alpha_0\nu + n\nu + n^\kappa + m - m\nu - 1)\Big] + O(n^{\kappa-1}) \simeq E + F + G + H + K
\end{aligned}$$

where E stands for the terms with 2π, F for linear terms, G for terms with $\ln n$, H for terms with $n\ln n$, and K for terms related to n^κ.

Consider the terms with 2π

$$E = (1-\nu)\frac{m}{2}\ln 2\pi - (1-\nu)\frac{1}{2}\ln 2\pi - \frac{m}{2}\ln 2\pi + \frac{1}{2}\ln 2\pi + \frac{m}{2}\ln 2\pi - \frac{1}{2}\ln 2\pi =$$

$$= \frac{1}{2}(m-1)(1-\nu)\ln 2\pi$$

and notice that $A = E$.

Consider the linear terms

$$F = -(1-\nu)\sum_{j=1}^{m}\big(\alpha_j + nr_j - 1\big) + (1-\nu)(\alpha_0 + n - 1) + \sum_{j=1}^{m}\big(\alpha_j + nr_j + \gamma_j n^\kappa - 1\big) -$$

$$-(\alpha_0 + n + n^\kappa - 1) - \sum_{j=1}^{m}\big(\alpha_j\nu + nr_j\nu + \gamma_j n^\kappa - \nu\big) + (\alpha_0\nu + n\nu + n^\kappa + m - m\nu - 1) =$$

$$= (m-1)(1-\nu)$$

and notice that $B = F$.

Consider the terms with $\ln n$

$$G = (1-\nu)\frac{1}{2}\sum_{j=1}^{m}\ln\big(\alpha_j + nr_j - 1\big) - (1-\nu)\frac{1}{2}\ln(\alpha_0 + n - 1) -$$

$$-\frac{1}{2}\sum_{j=1}^{m}\ln\left(\alpha_j+nr_j+\gamma n^{\kappa}-1\right)+\frac{1}{2}\ln(\alpha_0+n+n^{\kappa}-1)+$$

$$+\frac{1}{2}\sum_{j=1}^{m}\ln\left(\alpha_j\nu+nr_j\nu+\gamma_j n^{\kappa}-\nu\right)-\frac{1}{2}\ln(\alpha_0\nu+n\nu+n^{\kappa}+m-m\nu-1)\simeq$$

$$\simeq\frac{m}{2}(1-\nu)\ln n+\frac{1}{2}(1-\nu)\sum_{j=1}^{m}\ln r_j-\frac{1}{2}(1-\nu)\ln n$$

$$-\frac{m}{2}\ln n-\frac{1}{2}\sum_{j=1}^{m}\ln r_j+\frac{1}{2}\ln n+\frac{m}{2}\ln n+\frac{1}{2}\sum_{j=1}^{m}\ln r_j+\frac{m}{2}\ln\nu-\frac{1}{2}\ln n-\frac{1}{2}\ln\nu=$$

$$=\frac{1}{2}(m-1)(1-\nu)\ln n+\frac{1}{2}(1-\nu)\sum_{j=1}^{m}\ln r_j+\frac{1}{2}(m-1)\ln\nu$$

and notice that $C=G$.

Recall that $\sum\gamma_j=1$. Consider the terms with $n\ln n$

$$H=(1-\nu)\sum_{j=1}^{m}\left(\alpha_j+nr_j-1\right)\ln\left(\alpha_j+nr_j-1\right)-(1-\nu)(\alpha_0+n-1)\ln(\alpha_0+n-1)-$$

$$-\sum_{j=1}^{m}\left(\alpha_j+nr_j+\gamma_j n^{\kappa}-1\right)\cdot\ln\left(\alpha_j+nr_j+\gamma_j n^{\kappa}-1\right)+(\alpha_0+n+n^{\kappa}-1)\cdot$$

$$\cdot\ln(\alpha_0+n+n^{\kappa}-1)+\sum_{j=1}^{m}\left(\alpha_j\nu+nr_j\nu+\gamma_j n^{\kappa}-\nu\right)\ln\left(\alpha_j\nu+nr_j\nu+\gamma n^{\kappa}-\nu\right)-$$

$$-(\alpha_0\nu+n\nu+n^{\kappa}+m-m\nu-1)\cdot\ln(\alpha_0\nu+n\nu+n^{\kappa}+m-m\nu-1)\simeq$$

$$\simeq(1-\nu)\ln n+(1-\nu)n\ln n-(1-\nu)m\ln n+(1-\nu)\sum_{j=1}^{m}\alpha_j\ln r_j+(1-\nu)n\sum_{j=1}^{m}r_j\ln r_j-$$

$$-(1-\nu)\sum_{j=1}^{m}\ln r_j-(1-\nu)\alpha_0\ln n-(1-\nu)n\ln n++(1-\nu)\ln n-\ln n-\sum_{j=1}^{m}\alpha_j\ln r_j-n\ln n-$$

$$-n\sum_{j=1}^{m}r_j\ln r_j-n^{\kappa}\ln n-\sum_{j=1}^{m}\gamma_j n^{\kappa}\ln r_j+m\ln n+\sum_{j=1}^{m}\ln r_j+\alpha_0\ln n+n\ln n+n^{\kappa}\ln n-\ln n+$$

$$+\nu\ln n+\nu\sum_{j=1}^{m}\alpha_j\ln r_j+\nu\ln\nu+\nu n\ln n+\nu n\sum_{j=1}^{m}r_j\ln r_j+\nu n\ln\nu+n^{\kappa}\ln n+\sum_{j=1}^{m}\gamma_j n^{\kappa}\ln r_j+$$

$$+n^{\kappa}\ln\nu-\nu m\ln n-\nu\sum_{j=1}^{m}\ln r_j-\nu m\ln\nu-\alpha_0\nu\ln n-\alpha_0\nu\ln\nu-n\nu\ln n-n\nu\ln\nu-n^{\kappa}\ln n-$$

$$-n^{\kappa}\ln\nu-m\ln n-m\ln\nu+m\nu\ln n+m\nu\ln\nu+\ln n+\ln\nu=$$

$$=(\nu-\alpha_0\nu-m+1)\ln\nu+(\nu-1)(m-1)\ln n$$

and notice that $D=H$.

As a result, $A+B+C+D=E+F+G+H$. Therefore, what is left is K, with all non-decreasing the terms related to n^{κ}. Collecting all non-decreasing terms with n^{κ} from the expression for weighted Rényi entropy (20) leads to Formula (14) for ω_R.

Appendix B

Appendix B.1. Details of the Calculations in Proposition 2

Let $\mathbf{P} \sim \mathrm{Dir}(\boldsymbol{\alpha_1}), \mathbf{Q} \sim \mathrm{Dir}(\boldsymbol{\alpha_2})$. Compute the Kullback–Leibler divergence between $\mathbf{P}$ and $\mathbf{Q}$:

$$\mathrm{KL}[\mathbf{P}||\mathbf{Q}] = \int\limits_{S_{m-1}} \mathrm{Dir}(\boldsymbol{\alpha_1}) \ln \frac{\mathrm{Dir}(\boldsymbol{\alpha_1})}{\mathrm{Dir}(\boldsymbol{\alpha_2})} \mathrm{d}\boldsymbol{p} = \mathbb{E}_{\mathbf{P}}\left[\ln \frac{\mathrm{Dir}(\boldsymbol{\alpha_1})}{\mathrm{Dir}(\boldsymbol{\alpha_2})}\right] \tag{A2}$$

where $S_{m-1} = \{p \in \mathbb{R}^m : \sum\limits_{i=1}^{m} p_i = 1, 0 \leqslant p_i \leqslant 1, i = 1, \ldots, m\}$.

$$\begin{gathered}
\mathrm{KL}[\mathbf{P}||\mathbf{Q}] = \mathbb{E}_{\mathbf{P}}\left[\ln \frac{\Gamma\left(\sum\limits_{i=1}^{m} \alpha_{1i}\right)}{\prod\limits_{i=1}^{m} \Gamma(\alpha_{1i})} \prod_{i=1}^{m} p_i^{\alpha_{1i}-1} - \ln \frac{\Gamma\left(\sum\limits_{i=1}^{m} \alpha_{2i}\right)}{\prod\limits_{i=1}^{m} \Gamma(\alpha_{2i})} \prod_{i=1}^{m} p_i^{\alpha_{2i}-1}\right] = \\
\mathbb{E}_{\mathbf{P}}\left[\ln\left(\frac{\Gamma\left(\sum\limits_{i=1}^{m} \alpha_{1i}\right)}{\Gamma\left(\sum\limits_{i=1}^{m} \alpha_{2i}\right)} \cdot \frac{\prod\limits_{i=1}^{m} \Gamma(\alpha_{2i})}{\prod\limits_{i=1}^{m} \Gamma(\alpha_{1i})} \cdot \prod_{i=1}^{m} p_i^{\alpha_{1i}-\alpha_{2i}}\right)\right] = \\
\mathbb{E}_{\mathbf{P}}\left[\ln \frac{\Gamma\left(\sum\limits_{i=1}^{m} \alpha_{1i}\right)}{\Gamma\left(\sum\limits_{i=1}^{m} \alpha_{2i}\right)} + \sum_{i=1}^{m} \ln \frac{\Gamma(\alpha_{2i})}{\Gamma(\alpha_{1i})} + \sum_{i=1}^{m} (\alpha_{1i} - \alpha_{2i}) \cdot \ln(p_i)\right] = \\
\ln \frac{\Gamma\left(\sum\limits_{i=1}^{m} \alpha_{1i}\right)}{\Gamma\left(\sum\limits_{i=1}^{m} \alpha_{2i}\right)} + \sum_{i=1}^{m} \ln \frac{\Gamma(\alpha_{2i})}{\Gamma(\alpha_{1i})} + \sum_{i=1}^{m} (\alpha_{1i} - \alpha_{2i}) \cdot \mathbb{E}_{\mathbf{P}}[\ln(p_i)]
\end{gathered} \tag{A3}$$

where

$$\mathbb{E}_{\mathbf{P}}[\ln(p_i)] = \psi(\alpha_{1i}) - \psi\left(\sum_{j=1}^{m} \alpha_{1j}\right), \psi(t) = \frac{\Gamma'(t)}{\Gamma(t)}. \tag{A4}$$

To justify the identity in (A4), make $\alpha_0 = \sum\limits_{i=1}^{m} \alpha_i$ and write

$$\mathbb{E}_{\mathbf{P}}[\ln(p_i)] = \frac{\Gamma\left(\sum\limits_{i=1}^{m} \alpha_i\right)}{\prod\limits_{i=1}^{m} \Gamma(\alpha_i)} \int (\ln p_i) \prod_{i=1}^{m} p_i^{\alpha_i - 1} \mathrm{d}\boldsymbol{p},$$

$$p_i^{\alpha_i - 1} = e^{(\alpha_i - 1)\ln p_i} \qquad (\ln p_i) p_i^{\alpha_i - 1} = \frac{\partial}{\partial \alpha_i} p_i^{\alpha_i - 1},$$

$$\frac{\partial}{\partial \alpha_i} \frac{\Gamma\left(\sum\limits_{i=1}^{m} \alpha_i\right)}{\prod\limits_{i=1}^{m} \Gamma(\alpha_i)} \int \prod_{i=1}^{m} p_i^{\alpha_i - 1} \mathrm{d}\boldsymbol{p} = 0,$$

therefore

$$\mathbb{E}_{\mathbf{P}}[\ln(p_i)] = \frac{\Gamma(\alpha_0)}{\prod \Gamma(\alpha_i)} \frac{\partial}{\partial \alpha_j} \frac{\prod \Gamma(\alpha_i)}{\Gamma(\alpha_0)} = \psi(\alpha_j) - \psi(\alpha_0).$$

The next step is based on the Bretagnolle–Huber inequality [14]. Recall that

$$d_{\mathrm{TV}}(\mathbf{P}, \mathbf{Q}) \leqslant \sqrt{1 - e^{-n\,\mathrm{KL}[\mathbf{P}||\mathbf{Q}]}} \tag{A5}$$

where $d_{\mathrm{TV}}(\mathbf{P},\mathbf{Q}) = \frac{1}{2}\int \mid p(x) - q(x) \mid \mathrm{d}x$ is the total variation distance. This inequality implies that

$$e^{-n\,\mathrm{KL}[\mathbf{P}||\mathbf{Q}]} \leqslant 1 - d^2_{\mathrm{TV}}(\mathbf{P},\mathbf{Q}). \tag{A6}$$

Thus,

$$n \geqslant -\frac{\ln(1 - d^2_{\mathrm{TV}}(\mathbf{P},\mathbf{Q}))}{\mathrm{KL}[\mathbf{P}||\mathbf{Q}]}. \tag{A7}$$

The corresponding plots are given in Figure 2.

Appendix B.2. Details of the Calculations in Proposition 3

Let $\mathbf{P} \sim \mathrm{Dir}(\boldsymbol{\alpha_1} + \boldsymbol{x}), \mathbf{Q} \sim \mathrm{Dir}(\boldsymbol{\alpha_2} + \boldsymbol{y})$. Compute the Kullback–Leibler divergence between $\mathbf{P}$ and $\mathbf{Q}$, where $S_{m-1} = \{p \in \mathbb{R}^m : \sum\limits_{i=1}^{m} p_i = 1, 0 \leqslant p_i \leqslant 1, i = 1,\ldots,m\}$:

$$\begin{aligned}
&\mathrm{KL}[\mathbf{P}||\mathbf{Q}] = \int\limits_{S_{m-1}} \mathrm{Dir}(\boldsymbol{\alpha_1} + \boldsymbol{x}) \ln \frac{\mathrm{Dir}(\boldsymbol{\alpha_1} + \boldsymbol{x})}{\mathrm{Dir}(\boldsymbol{\alpha_2} + \boldsymbol{y})} \mathrm{d}\boldsymbol{p} = \mathbb{E}_{\mathbf{P}}\left[\ln \frac{\mathrm{Dir}(\boldsymbol{\alpha_1} + \boldsymbol{x})}{\mathrm{Dir}(\boldsymbol{\alpha_2} + \boldsymbol{y})}\right] = \\
&= \mathbb{E}_{\mathbf{P}}\left[\ln \frac{\Gamma\left(\sum\limits_{i=1}^{m} \alpha_{1i} + n\right)}{\prod\limits_{i=1}^{m} \Gamma(\alpha_{1i} + x_i)} \prod_{i=1}^{m} p_i^{\alpha_{1i}+x_i-1} - \ln \frac{\Gamma\left(\sum\limits_{i=1}^{m} \alpha_{2i} + n\right)}{\prod\limits_{i=1}^{m} \Gamma(\alpha_{2i} + y_i)} \prod_{i=1}^{m} p_i^{\alpha_{2i}+y_i-1}\right] = \\
&= \mathbb{E}_{\mathbf{P}}\left[\ln\left(\frac{\Gamma\left(\sum\limits_{i=1}^{m} \alpha_{1i} + n\right)}{\Gamma\left(\sum\limits_{i=1}^{m} \alpha_{2i} + n\right)} \cdot \frac{\prod\limits_{i=1}^{m} \Gamma(\alpha_{2i} + y_i)}{\prod\limits_{i=1}^{m} \Gamma(\alpha_{1i} + x_i)} \cdot \prod_{i=1}^{m} p_i^{\alpha_{1i}+x_i-\alpha_{2i}-y_i}\right)\right] = \\
&= \mathbb{E}_{\mathbf{P}}\left[\ln \frac{\Gamma\left(\sum\limits_{i=1}^{m} \alpha_{1i} + n\right)}{\Gamma\left(\sum\limits_{i=1}^{m} \alpha_{2i} + n\right)} + \ln \frac{\prod\limits_{i=1}^{m} \Gamma(\alpha_{2i} + y_i)}{\prod\limits_{i=1}^{m} \Gamma(\alpha_{1i} + x_i)} + \ln \prod_{i=1}^{m} p_i^{\alpha_{1i}+x_i-\alpha_{2i}-y_i}\right] = \\
&= \ln \frac{\Gamma\left(\sum\limits_{i=1}^{m} \alpha_{1i} + n\right)}{\Gamma\left(\sum\limits_{i=1}^{m} \alpha_{2i} + n\right)} + \ln \frac{\prod\limits_{i=1}^{m} \Gamma(\alpha_{2i} + y_i)}{\prod\limits_{i=1}^{m} \Gamma(\alpha_{1i} + x_i)} + \sum_{i=1}^{m} \mathbb{E}_{\mathbf{P}}\left[\ln p_i^{\alpha_{1i}+x_i-\alpha_{2i}-y_i}\right].
\end{aligned} \tag{A8}$$

Assume that the number of prior observations is insignificantly small in comparison to the number of posterior observations and that the initial priors are set to be the same. Hence, $\mathbf{P} \sim \mathrm{Dir}(x_1, x_2, x_3), \mathbf{Q} \sim \mathrm{Dir}(y_1, y_2, y_3)$. Also, let us specify that $y_1 = x_1 + z, y_2 = x_2 - z, y_3 = x_3$.

Using the asymptotics of the Gamma function,

$$\Gamma(x+1) \sim (x/e)^x \sqrt{2\pi x},$$

the following formula is obtained:

$$\mathrm{KL}[\mathbf{P}||\mathbf{Q}] = \sum (y_i \ln y_i - x_i \ln x_i) + \sum (x_i - y_i)\mathbb{E}_{\mathbf{P}}[\ln p_i]. \tag{A9}$$

The minimal distance z between x_1 and y_1 so that the strategies are distinguished with confidence, $d_{\mathrm{TV}}(\mathbf{P},\mathbf{Q})$, can be determined as follows:

$$\mathbb{E}_{\mathbf{P}}[\ln(p_i)] = \psi(x_i) - \psi\left(\sum_{j=1}^{m} x_j\right), \text{ where } \psi(t) = \frac{\Gamma'(t)}{\Gamma(t)} \sim \ln(t) - \tfrac{1}{2t}, \tag{A10}$$

and in view of Bretagnolle–Huber inequality [14],

$$d_{\mathrm{TV}}(\mathbf{P},\mathbf{Q}) \leqslant \sqrt{1 - e^{-\mathrm{KL}[\mathbf{P}||\mathbf{Q}]}}, \tag{A11}$$

$$e^{-\mathrm{KL}[\mathbf{P}||\mathbf{Q}]} \leqslant 1 - d_{\mathrm{TV}}^2(\mathbf{P},\mathbf{Q}), \tag{A12}$$

$$\mathrm{KL}[\mathbf{P}||\mathbf{Q}] \geqslant -\ln(1 - d_{\mathrm{TV}}^2(\mathbf{P},\mathbf{Q})), \tag{A13}$$

and (A9) is processed by numerical methods.

References

1. Azriel, D.; Mandel, M.; Rinott, Y. The treatment versus experimentation dilemma in dose finding studies. *J. Stat. Plan. Inference* **2011**, *141*, 2759–2768. [CrossRef]
2. Sutton, R.S.; Barto, A.G. *Reinforcement Learning: An Introduction*; MIT Press: Cambridge, MA, USA, 2018.
3. Mozgunov, P.; Jaki, T. An information theoretic approach for selecting arms in clinical trials. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2020**, *82*, 1223–1247. [CrossRef]
4. Kasianova, K.; Kelbert, M.; Mozgunov, P. Response adaptive designs for Phase II trials with binary endpoint based on context-dependent information measures. *Comput. Stat. Data Anal.* **2021**, *158*, 107187. [CrossRef] [PubMed]
5. Deisenroth, M.P.; Faisal, A.A.; Ong, C.S. *Mathematics for Machine Learning;* Cambridge University Press: Cambridge, UK, 2020.
6. Murphy, K.P. *Machine Learning: A Probabilistic Perspective (Adaptive Computation and Machine Learning Series)*; MIT Press: Cambridge, MA, USA, 2018.
7. Murphy, K.P. *Probabilistic Machine Learning: Advanced Topics*; MIT Press: Cambridge, MA, USA, 2023.
8. Stuhl, I.; Kelbert, M.; Suhov, Y.; Yasaei Sekeh, S. Weighted gaussian entropy and determinant inequalities. *Aequationes Math.* **2022**, *96*, 85–114. [CrossRef]
9. Williamson, S.F.; Jacko, P.; Villar, S.S.; Jaki, T. A Bayesian adaptive design for clinical trials in rare diseases. *Comput. Stat. Data Anal.* **2017**, *113*, 136–153. [CrossRef] [PubMed]
10. Suhov, Y.; Stuhl, I.; Yasaei Sekeh, S.; Kelbert, M. Basic inequalities for weighted entropies. *Aequationes Math.* **2016**, *90*, 817–848. [CrossRef]
11. Kelbert, M.; Stuhl, I.; Suhov, Y. Weighted entropy and its use in computer science and beyond. In *Analytical and Computational Methods in Probability Theory*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 293–308.
12. Ferguson, T.S. A Bayesian analysis of some nonparametric problems. *Ann. Stat.* **1973**, *1*, 209–230. [CrossRef]
13. Kelbert, M.; Mozgunov, P. Asymptotic behaviour of the weighted Rényi, Tsallis and Fisher entropies in a Bayesian problem. *Eurasian Math. J.* **2015**, *6*, 6–17.
14. Bretagnolle, J.; Huber, C. Estimation des densités: Risque minimax. *Séminaire De Probab. De Strasbg.* **1978**, *12*, 342–363.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.