stats

MDPI

Article

Saddlepoint Approximation for Data in Simplices: A Review with New Applications

Riccardo Gatto

Institute of Mathematical Statistics and Actuarial Science, University of Bern, 3012 Bern, Switzerland; gatto@stat.unibe.ch; Tel.: +41-31-631-8807

Received: 23 January 2019; Accepted: 14 February 2019; Published: 18 February 2019

Abstract: This article provides a review of the saddlepoint approximation for a M-statistic of a sample of nonnegative random variables with fixed sum. The sample vector follows the multinomial, the multivariate hypergeometric, the multivariate Polya or the Dirichlet distributions. The main objective is to provide a complete presentation in terms of a single and unambiguous notation of the common mathematical framework of these four situations: the simplex sample space and the underlying general urn model. Some important applications are reviewed and special attention is given to recent applications to models of circular data. Some novel applications are developed and studied numerically.

Keywords: bootstrap; circular data; Dirichlet distribution; entropy; likelihood ratio test; multinomial distribution; multivariate hypergeometric distribution; multivariate Polya distribution; spacings; spacing-frequencies; urn model.

MSC: 41A60; 60C05

1. Introduction

The topic of this article is a saddlepoint approximation to the distribution of the M-statistic T_n, precisely $T_n(Y_1, \ldots, Y_n)$, which is the implicit solution with respect to (w.r.t.) t of

$$\sum_{j=1}^{n} \xi_j(Y_j; t) = 0, \tag{1}$$

where the function ξ_j: $\mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$ is continuous (thus measurable), decreasing in its second argument, for $j = 1, \ldots, n$, $\mathbb{R}_+ = [0, \infty)$, and where the random variables $Y_1, \ldots, Y_n$ are nonnegative, dependent and satisfy $\sum_{j=1}^{n} Y_j = k$, for some fixed $k > 0$. Decreasing is meant in the strict sense. The sample vector $(Y_1, \ldots, Y_n)$ takes values in a simplex. It is often referred to as compositional data, by referring to the situation where Y_j represents the number of units of the jth category, for $j = 1, \ldots, n$, given n possible categories (see e.g., [1]). When $(Y_1, \ldots, Y_n)$ follows the multinomial distribution, it is also referred to as categorical data. We consider three discrete and one continuous joint distributions for $(Y_1, \ldots, Y_n)$ and relate these multivariate distributions to three general urn sampling schemes that are given, e.g., in [2].

source: https://doi.org/10.7892/boris.142970 | downloaded: 12.10.2024

The derivation of the saddlepoint approximation to the distribution of T_n relies on the distributional equivalence

$$(Y_1, \ldots, Y_n) \sim \left((X_1, \ldots, X_n) \,\Big|\, \sum_{j=1}^{n} X_j = k \right), \tag{2}$$

which means that $(Y_1, \ldots, Y_n)$ has the conditional distribution of $(X_1, \ldots, X_n)$ given $\sum_{j=1}^{n} X_j = k$. The nonnegative random variables $X_1, \ldots, X_n$ form a conditional triangular array in the sense that, conditionally on their sum, they are independent and their individual distributions may depend on n. We refer to Equation (2) as the conditional representation of $(Y_1, \ldots, Y_n)$ in terms of $(X_1, \ldots, X_n)$. The computation of the distribution of T_n, as function of the dependent random variables $Y_1, \ldots, Y_n$, is generally difficult. It is however simplified by replacing these dependent random variables by the triangular array random variables $X_1, \ldots, X_n$, in the same order, conditional on their sum. Gatto and Jammalamadaka [3] extended the saddlepoint approximation for tail probabilities of Skovgaard [4] to M-statistics and used the conditional representation in Equation (2) to derive saddlepoint approximations for important classes of nonparametric tests, such as tests based on spacings, two-sample tests based on spacing-frequencies and various tests based on ranks. The application of this conditional saddlepoint approximation to the computation of quantiles can be found in [5]. Further applications can be found in [6,7].

This article presents the conditional saddlepoint approximation from the general perspective of the urn sampling model. Four cases of the of conditional representations given in Equation (2) are related to the urn model: the joint multinomial in terms of Poisson random variables conditional on their sum (M-P), the joint multivariate hypergeometric in terms of binomial random variables conditional on their sum (MH-B), the joint multivariate Polya in terms of negative binomial random variables conditional on their sum (MP-NB) and the joint Dirichlet in terms of gamma random variables conditional on their sum (D-G). New applications or examples are given and tested numerically. Various previous applications of the conditional saddlepoint approximation are reviewed. Two other general references on conditional saddlepoint approximations are found in [8] (Chapter 4 and Section 12.5) and [9]. This article completes these references in various ways. It provides a concise and complete presentation of the conditional saddlepoint approximation for M-statistics (that includes an approximation to quantiles). It updates the previous reviews by presenting additional recent important examples. It gives a general reformulation with a consistent and homogeneous notation, that corresponds to a single underlying mathematical model (viz., the urn model and the simplex sample space). It includes new important examples and new numerical comparisons. The numerical illustrations are given for: the distribution of an estimator of the entropy that relates to the urn model, the power of the likelihood ratio test, the distribution of the insurer's total claim amount and the null distribution of a test for symmetry of Dirichlet's distribution.

Mirakhmedov et al. [10] used the three well-known conditional representations M-P, MH-B and MP-NB with the Edgeworth approximation. The Edgeworth is however not a large deviations approximation. Edgeworth approximations to small tail probabilities are usually less accurate than saddlepoint approximations. Butler and Sutton [11] proposed a particular saddlepoint approximation that exploits the conditional representation in Equation (2). It implies that, for all intervals $I_1, \ldots, I_n \subset \mathbb{R}_+$,

$$\mathsf{P}[Y_1 \in I_1, \ldots, Y_n \in I_n] = \mathsf{P}\left[\sum_{j=1}^{n} X_j = k \,\middle|\, X_1 \in I_1, \ldots, X_n \in I_n \right] \frac{\mathsf{P}[X_1 \in I_1, \ldots, X_n \in I_n]}{\mathsf{P}\left[\sum_{j=1}^{n} X_j = k\right]}.$$

Then, the conditional probability above is approximated by a saddlepoint approximation for independent and truncated random variables. This method allows approximating the distribution of $M_n = \max_{j=1,\ldots,n} Y_j$, for example, but does not allow approximating the distribution of the M-statistic in Equation (1). Note that, for the case where $(Y_1, \ldots, Y_n)$ follows the multinomial distribution, given in Equation (3), Good [12] proposed a specific saddlepoint approximation for M_n.

This article has the following structure. Section 2 presents the four conditional representations, in Sections 2.1 and 2.3. They are related to urn sampling schemes in Section 2.2. The three first conditional representations, namely M-P, MH-B and MP-NB, are for counting random variables. The fourth conditional representation is D-G and holds for positive random variables. Section 3 summarizes the conditional saddlepoint approximation for a M-statistics given another one: Sections 3.1 and 3.2 are for tail probabilities and Section 3.3 for quantiles. Then, Section 4 provides new applications and numerical studies for this saddlepoint approximation and briefly reviews other important existing applications. Some final remarks are given in Section 5.

Regarding notation, we define $\mathbb{N} = \{0, 1, \ldots\}$, $\mathbb{N}^* = \mathbb{N} \backslash \{0\}$, $\mathbb{R}_+ = [0, \infty)$ as already defined and $\mathbb{R}_+^* = \mathbb{R}_+ \backslash \{0\}$. The Pochhammer symbol is defined by

$$(x)_k = x \cdot \ldots \cdot (x - k + 1), \quad \forall x \in \mathbb{R}, k \in \mathbb{N}^*.$$

The binomial coefficient is defined by

$$\binom{x}{k} = \begin{cases} 0, & \text{if } k = -1, -2, \ldots, \\ 1, & \text{if } k = 0, \\ \frac{(x)_k}{k!}, & \text{if } k = 1, 2, \ldots, \end{cases} \qquad \forall x \in \mathbb{R}.$$

The indicator function of the statement A is defined by

$$\mathrm{I}\{A\} = \begin{cases} 0, & \text{if } A \text{ is false}, \\ 1, & \text{if } A \text{ is true}. \end{cases}$$

Let $n \in \{2, 3, \ldots\}$. A $(n-1)$-simplex is the $(n-1)$-dimensional polytope determined by the convex hull of its n vertices. We consider only the symmetric simplex. It is obtained by defining the jth vertex $\boldsymbol{v}_j = (v_0, \ldots, v_{n-1})$ by

$$v_i = \begin{cases} x, & \text{if } i = j, \\ 0, & \text{otherwise}, \end{cases} \quad \text{for } i = 0, \ldots, n-1,$$

for any desired size $x \in \mathbb{R}_+^*$ and for $j = 0, \ldots, n-1$. This representation corresponds to the set

$$\Delta_x^{n-1} = \{(x_1, \ldots, x_n) \in \mathbb{R}_+^n \mid x_1 + \ldots + x_n = x\}.$$

We define also by

$$\ddot{\Delta}_k^{n-1} = \Delta_k^{n-1} \cap \mathbb{N}^n = \{(k_1, \ldots, k_n) \in \mathbb{N}^n \mid k_1 + \ldots + k_n = k\}$$

the integer $(n-1)$-simplex of size $k \in \mathbb{N}^*$.

We denote by $X \sim Y$ the fact that the two random elements X and Y have same distribution. The same symbol is used for the asymptotic equivalence.

2. Four Conditional Representations and Their Urn Sampling Interpretations

This section reviews four multivariate distributions for which the conditional representation in Equation (2) holds and relates them to a common urn model. Although these results are classical and can be retrieved perhaps separately in the literature, the contribution of this section must be sought in the single and unambiguous mathematical reformulation: of the multivariate distributions, of their conditional representations and of their urn model. The same notation is used for saddlepoint approximation in Section 3 and for the examples in Section 4. The first three models are presented in Section 2.1 and are related to the three urn sampling schemes in Section 2.2. In these three models, $(Y_1, \ldots, Y_n)$ takes values in $\ddot{\Delta}_k^{n-1}$, for $k \in \mathbb{N}^*$. Section 2.3 presents a fourth multivariate model where $(Y_1, \ldots, Y_n)$ takes values in Δ_k^{n-1}, for $k \in \mathbb{R}_+^*$, and for which an asymptotic relation with one of the urn sampling models holds.

2.1. Three Conditional Representations for Counting Random Variables

The next three multivariate distributions allow for the conditional representation in Equation (2) and relate to the three urn sampling schemes of Section 2.2.

- *Multinomial—conditional Poisson (M-P)*
  Let $X_j \sim \text{Poisson}(qp_j)$, i.e., Poisson distributed with parameter qp_j, for $j = 1, \ldots, n$, be independent, where $(p_1, \ldots, p_n) \in \Delta_1^{n-1}$ and $q \in \mathbb{R}_+^*$. Then, the conditional representation in Equation (2) holds with $(Y_1, \ldots, Y_n) \sim \text{Multinomial}(k; p_1, \ldots, p_n)$, for $k \in \mathbb{N}^*$, that is, with

$$\mathsf{P}[Y_1 = k_1, \ldots, Y_n = k_n] = \binom{k}{k_1 \ldots k_n} p_1^{k_1} \ldots p_n^{k_n}, \tag{3}$$

  $\forall (k_1, \ldots, k_n) \in \ddot{\Delta}_k^{n-1}$, which is the multinomial distribution. Thus, $k = \sum_{j=1}^n k_j$.
- *Multivariate hypergeometric—conditional binomial (MH-B)*
  Let $X_j \sim \text{Binomial}(m_j, q)$, i.e., binomial distributed with m_j trials and elementary probability q, for $j = 1, \ldots, n$, be independent, where $(m_1, \ldots, m_n) \in \ddot{\Delta}_z^{n-1}$, $z = \sum_{j=1}^n m_j$ and $q \in (0, 1)$. Then, the conditional representation in Equation (2) holds with $(Y_1, \ldots, Y_n) \sim \text{Multi-Hypergeometric}(k; m_1, \ldots, m_n)$, for $k \in \mathbb{N}^*$, that is with

$$\mathsf{P}[Y_1 = k_1, \ldots, Y_n = k_n] = \frac{\prod_{j=1}^n \binom{m_j}{k_j}}{\binom{z}{k}}, \tag{4}$$

  for $k_j = 0, \ldots, m_j$, for $j = 1, \ldots, n$, and $k = \sum_{j=1}^n k_j \leq z$, which is the multivariate hypergeometric distribution. Thus, $(k_1, \ldots, k_n) \in \ddot{\Delta}_k^{n-1} \cap ([0, m_1] \times \ldots \times [0, m_n])$.
- *Multivariate Polya—conditional negative binomial (MP-NB)*
  Let $X_j \sim \text{Negative-Binomial}(m_j, q)$, i.e.

$$\mathsf{P}[X_j = l] = \binom{l + m_j - 1}{l} q^{m_j} (1 - q)^l, \quad \text{for } l = 0, 1, \ldots,$$

  for $j = 1, \ldots, n$, be independent, where $(m_1, \ldots, m_n) \in \Delta_u^{n-1}$, for some $u \in \mathbb{R}_+^*$, and $q \in (0, 1)$. Thus, $u = \sum_{j=1}^n m_j$. Then, the conditional representation in Equation (2) holds with $(Y_1, \ldots, Y_n) \sim \text{Multi-Polya}(k; m_1, \ldots, m_n)$, for $k \in \mathbb{N}^*$, that is with

$$\mathsf{P}[Y_1 = k_1, \ldots, Y_n = k_n] = \frac{\prod_{j=1}^n \binom{m_j + k_j - 1}{k_j}}{\binom{u+k-1}{k}}, \tag{5}$$

  $\forall (k_1, \ldots, k_n) \in \ddot{\Delta}_k^{n-1}$, which is the multivariate Polya distribution. Thus, $k = \sum_{j=1}^n k_j$.

We end this section with three remarks of general interest. We first note that in these three situations the conditional representation in Equation (2) holds independently of the choice of q, in $\mathbb{R}_+^*$ for the M-P and in $(0,1)$ for the MH-B and MP-NB representations. This independence can be understood from fact that, in all three cases, $\sum_{j=1}^n X_j$ is a sufficient statistic for q. This is a consequence of the factorization theorem of sufficient statistics.

We can see that each one of the three conditional representations have an interpretation in terms of mixture models. For example, consider the independent random variables $X_j \sim \text{Poisson}(qp_j)$, for $j = 1, \ldots, n$, where $(p_1, \ldots, p_n) \in \Delta_1^{n-1}$ and $q \in \mathbb{R}_+^*$. Then, $\forall k_1, \ldots, k_n \in \mathbb{N}$, for $k = \sum_{j=1}^n k_j$ and $K = \sum_{j=1}^n X_j$,

$$\begin{aligned}\mathsf{P}[X_1 = k_1, \ldots, X_{n-1} = k_{n-1}] &= \sum_{k_n=0}^{\infty} \mathsf{P}[X_1 = k_1, \ldots, X_{n-1} = k_{n-1}, X_n = k_n] \\ &= \sum_{k_n=0}^{\infty} \binom{k}{k_1 \ldots k_{n-1}\, k_n} p_1^{k_1} \ldots p_{n-1}^{k_{n-1}} p_n^{k_n}\, \mathrm{e}^{-q} \frac{q^k}{k!}. \qquad (6)\end{aligned}$$

Thus, $(X_1, \ldots, X_{n-1})$ follows the countable mixture distribution given by multinomial probabilities with Poisson mixing probabilities. Moreover,

$$\begin{aligned}\sum_{k_n=0}^{\infty} \mathsf{P}[X_1 = k_1, \ldots, X_{n-1} = k_{n-1}, X_n = k_n] &= \sum_{k_n=0}^{\infty} \mathsf{P}[X_1 = k_1, \ldots, X_{n-1} = k_{n-1}, X_n = k_n \mid K = k]\mathsf{P}[K = k] \\ &= \sum_{k_n=0}^{\infty} \mathsf{P}[X_1 = k_1, \ldots, X_{n-1} = k_{n-1} \mid K = k]\mathsf{P}[K = k]. \qquad (7)\end{aligned}$$

By equating the multinomial and the Poisson probabilities of Equation (6) to the two probabilities of Equation (7), for any summand, we obtain the M-P conditional representation.

We also note that that the three distributions of $X_1, \ldots, X_n$ (before conditioning) correspond to the three distributions of the $(a, b, 0)$ class. The probability distribution $\{p_n\}_{n \geq 0}$ belongs to the $(a, b, 0)$ class, if it satisfies the recurrence relation $p_n = (a + b/n)p_{n-1}$, for $n = 1, 2, \ldots$ and for some $a, b \in \mathbb{R}$ (see, e.g., Section 6.5 of [13]).

2.2. Three Associated Urn Sampling Schemes

The three multivariate distributions presented in the previous section provide probability models for three sampling schemes: sampling with replacement, sampling without replacement and Polya's sampling. These three sampling schemes are reunited in a single general urn sampling model by Ivchenko and Ivanov [14] (see also [2]). Consider an urn containing balls with the n different colors $\mathcal{C}_1, \ldots, \mathcal{C}_n$. At the beginning, the urn contains: $a_{j,0} \in \mathbb{N}$ balls of color $\mathcal{C}_j$, for $j = 1, \ldots, n$. Each single ball is drawn equiprobably from the urn. Immediately after the lth draw of a ball of color $\mathcal{C}_j$, $a_{j,l-1} \in \mathbb{N}^*$ is updated by $a_{j,l} \in \mathbb{N}$; this holds for $l = 1, 2, \ldots$ and $j = 1, \ldots, n$. Three updating mechanisms are presented in the next paragraph. Thus, immediately after drawing k_j balls of color $\mathcal{C}_j$, for $j = 1, \ldots, n$, and therefore just after a total of $k = \sum_{j=1}^n k_j$ draws, the urn contains a_{j,k_j} balls of color $\mathcal{C}_j$, for $j = 1, \ldots, n$. The updated sampling probability of color $\mathcal{C}_j$ is thus

$$p_j^{(k_1,\ldots,k_n)} = \left(\sum_{j=1}^n a_{j,k_j}\right)^{-1} a_{j,k_j}, \quad \text{for} \quad j = 1, \ldots, n,$$

provided $\sum_{j=1}^n a_{j,k_j} > 0$. The random count Y_j represents the number of randomly drawn balls of color $\mathcal{C}_j$, this for $j = 1, \ldots, n$, after a fixed total number of draws $k = \sum_{j=1}^n Y_j \in \mathbb{N}^*$. Define by $z = \sum_{j=1}^n a_{j,0}$ the initial total number of balls in the urn.

We are interested in the distribution of the M-statistic T_n viz. $T_n(Y_1, \ldots, Y_n)$ defined in Equation (1), under the three following sampling schemes.

- *Sampling with replacement and M-P representation*
  Sampling with replacement from the urn is obtained by setting

$$a_{j,l} = a_{j,0}, \text{ for } l = 1, 2, \ldots \text{ and } j = 1, \ldots, n.$$

  Thus, $p_j^{(k_1,\ldots,k_n)}$, for $j = 1, \ldots, n$, do not depend on $k_1, \ldots, k_n$ and the multinomial distribution in Equation (3) holds with rational $p_j = p_j^{(k_1,\ldots,k_n)} = a_{j,0}/z$, for $j = 1, \ldots, n$. Thus, $(Y_1, \ldots, Y_n)$ takes values in $\ddot{\Delta}_k^{n-1}$ and the M-P representation holds.
- *Sampling without replacement and MH-B representation*
  Sampling without replacement from the urn is obtained by setting

$$a_{j,l} = \begin{cases} a_{j,l-1} - 1 = a_{j,0} - l, & \text{if } l \leq a_{j,0}, \\ 0, & \text{if } l > a_{j,0}, \end{cases} \text{ for } l = 1, 2, \ldots \text{ and } j = 1, \ldots, n.$$

  Assume that $k_j \leq a_{j,0}$ balls of color $\mathcal{C}_j$ have been drawn, for $j = 1, \ldots, n$. The probability of drawing a ball of color $\mathcal{C}_j$ in the next draw is $p_j^{(k_1,\ldots,k_n)} = (a_{j,0} - k_j)/(z - k)$, if $k < z$, and it is undefined, if $k = z$, for $j = 1, \ldots, n$. The multivariate hypergeometric distribution in Equation (4) holds with $m_j = a_{j,0}$, for $j = 1, \ldots, n$, and z equal to the parameter z of the present section. Thus, $(Y_1, \ldots, Y_n)$ takes values in $\ddot{\Delta}_k^{n-1} \cap ([0, a_{1,0}] \times \ldots \times [0, a_{n,0}])$ and the MH-B representation holds.
- *Polya's sampling and MP-NB representation*
  Polya's sampling scheme is obtained by setting

$$a_{j,l} = a_{j,l-1} + r = a_{j,0} + lr, \quad \text{for} \quad l = 1, 2, \ldots \quad \text{and} \quad j = 1, \ldots, n, \tag{8}$$

  where $r \in \mathbb{N}^*$. (Allowing for $r = 0$ would result in sampling with replacement and allowing for $r = -1$ would result in sampling without replacement, which are already presented.) Assume that k_j balls of color $\mathcal{C}_j$ have been drawn, for $j = 1, \ldots, n$. The probability of drawing a ball of color $\mathcal{C}_j$ in the next draw is $p_j^{(k_1,\ldots,k_n)} = (a_{j,0} + k_j r)/(z + kr)$, for $j = 1, \ldots, n$. In this case, the multivariate Polya distribution in Equation (5) holds with rational $m_j = a_{j,0}/r$, for $j = 1, \ldots, n$, and rational $u = z/r$. Thus, $(Y_1, \ldots, Y_n)$ takes values in $\ddot{\Delta}_k^{n-1}$ and the MP-NB representation holds.

2.3. A Conditional Representation for Positive Random Variables and Its Urn Sampling Interpretation

This section presents a fourth model that allows for the conditional representation in Equation (2). It is the Dirichlet distribution and it has a steady state interpretation in terms of Polya's urn. Now, the dependent random variables $Y_1, \ldots, Y_n$ take values in $\mathbb{R}_+$ and cannot yet be considered as counts of the urn model of Section 2.2.

- *Dirichlet—conditional gamma (D-G)*
  Let $X_j \sim \text{Gamma}(a_j, q)$, with density $q^{a_j} e^{-qx} x^{a_j - 1}/\Gamma(a_j)$, $\forall x > 0$, for $j = 1, \ldots, n$, be independent, where $a_1, \ldots, a_n$ and $q \in \mathbb{R}_+^*$. Then, the conditional representation in Equation (2) holds with $(Y_1, \ldots, Y_n) \sim k(\bar{Y}_1, \ldots, \bar{Y}_n)$, where $(\bar{Y}_1, \ldots, \bar{Y}_n)$ is Dirichlet distributed with density

$$\mathsf{P}[\bar{Y}_1 \in (y_1, y_1 + \mathrm{d}y_1), \ldots, \bar{Y}_n \in (y_n, y_n + \mathrm{d}y_n)] = \frac{\Gamma(a_1 + \ldots + a_n)}{\Gamma(a_1) \ldots \Gamma(a_n)} y_1^{a_1 - 1} \ldots y_n^{a_n - 1} \mathrm{d}y_1 \ldots \mathrm{d}y_n, \quad (9)$$

  $\forall (y_1, \ldots, y_n) \in \text{int}\, \Delta_1^{n-1}$ and for $\mathrm{d}y_n = -(\mathrm{d}y_1 + \ldots + \mathrm{d}y_{n-1})$, which is denoted $(\bar{Y}_1, \ldots, \bar{Y}_n) \sim \text{Dirichlet}(a_1, \ldots, a_n)$.

The validity of Equation (2) does not depend on the parameter $q \in \mathbb{R}_+^*$ of the gamma distribution. This independence follows from the factorization theorem of sufficient statistics.

The Dirichlet distribution represents the steady state of Polya's urn sampling scheme, viz. of the multivariate Polya distribution given in Section 2.2.

- *Polya's sampling and D-G representation*
  Precisely, immediately after drawing a ball of color $\mathcal{C}_j$, it is replaced together with $r \in \mathbb{N}^*$ new balls of same color $\mathcal{C}_j$, this for $j = 1, \ldots, n$, cf. Equation (8). If we let the total number of draws k go to infinity, then the vector of the proportions of the n drawn colors follows the $\text{Dirichlet}(a_{1,0}/r, \ldots, a_{n,0}/r)$ distribution, viz.

$$\frac{1}{k}(Y_1, \ldots, Y_n) \xrightarrow{\mathrm{d}} (\bar{Y}_1, \ldots, \bar{Y}_n), \text{ as } k \to \infty, \quad (10)$$

  where $(\bar{Y}_1, \ldots, \bar{Y}_n)$ has the Dirichlet distribution in Equation (9) with $a_j = a_{j,0}/r$, for $j = 1, \ldots, n$. Thus, if $(Y_1, \ldots, Y_n)$ follows the multivariate Polya distribution in Equation (5), taking values in $\ddot{\Delta}_k^{n-1}$, then it is approximatively distributed as $k(\bar{Y}_1, \ldots, \bar{Y}_n)$, taking values in Δ_k^{n-1}.
  To see Equation (10), let $(k_1, \ldots, k_n) \in \ddot{\Delta}_k^{n-1}$. The multivariate Polya probability in Equation (5) can be re-expressed as

$$\mathsf{P}[Y_1 = k_1, \ldots, Y_n = k_n] = \frac{\Gamma(u)}{\prod_{j=1}^n \Gamma(m_j)} \frac{\Gamma(1+k)}{\Gamma(u+k)} \prod_{j=1}^n \frac{\Gamma(m_j + k_j)}{\Gamma(1 + k_j)}.$$

  It follows from Stirling's formula that $\Gamma(x + z_1)/\Gamma(x + z_2) \sim x^{z_1 - z_2}$, as $x \to \infty$, $\forall z_1, z_2 \in \mathbb{R}$. Consequently,

$$\mathsf{P}[Y_1 = k_1, \ldots, Y_n = k_n] \sim c_1(k) \prod_{j=1}^n k_j^{m_j - 1} \sim c_2(k) \prod_{j=1}^n y_j^{m_j - 1} = c_2(k) \prod_{j=1}^n y_j^{\frac{a_{j,0}}{r} - 1}, \text{ as } k \to \infty,$$

  for some positive constants $c_1(k)$ and $c_2(k)$ depending on k and for $y_j = \lim_{k \to \infty} k_j/k$, for $j = 1, \ldots, n$.

3. Conditional Saddlepoint Approximation for M-Statistics

The saddlepoint method, viz. method of steepest descent, allows approximating integrals of the form $\int_\rho f(z) e^{\nu g(z)} \mathrm{d}z$, for large values of $\nu > 0$, where $f\colon \mathbb{C} \to \mathbb{C}$ and $g\colon \mathbb{C} \to \mathbb{C}$ are analytic functions in a domain containing the path ρ and its deformations. Let z_0 be point where the real part of g is the highest. It is a saddlepoint of the surface given by the real part of g. For large values of ν, the value of the integral is accurately approximated as follows. First, restrict ρ to a small neighborhood of z_0. Second, deform ρ such that it crosses z_0 and so that the real part of g decreases fast to $-\infty$, when descending from z_0 down to the endpoints of the deformed ρ. This is the path of steepest descent. The final step is the term-by-term integration, within the neighborhood of z_0, of an asymptotic expansion of the integrand around z_0. Two references are [15,16].

This method yields approximations to densities or tail probabilities of various random variables such as estimators or test statistics. The sample size n takes the role of the asymptotic parameter ν and the relative error of the saddlepoint approximation vanishes at rate n^{-1}, as $n \to \infty$. Unlike normal or Edgeworth approximations, saddlepoint approximations are valid at any fixed point (not depending on n) of the support of the distribution. They are thus large deviations techniques. For these two reasons, they provide accurate approximations to small tail probabilities, in fact even for small values of n. The saddlepoint approximation was introduced into statistics by Daniels [17], for approximating density functions. Lugannani and Rice [18] provided a formula for tail probabilities (see also [19]).

Saddlepoint approximations for conditional distributions were proposed by: Skovgaard [4] for the distribution of a sample mean given another mean; Wang [20] for the distribution of a mean given a nonlinear function of another mean; and Jing and Robinson [21] for the distribution of a nonlinear function of a mean given a nonlinear function of another mean. Kolassa [22] derived higher order terms to the conditional saddlepoint approximation of a sample mean given another mean, by using a different expansion to an integral appearing [4]. DiCiccio [23] provided a different approximation, which is however restricted to the exponential class of distributions.

Some survey articles are [24–27]. General references are [8,28–30].

The saddlepoint approximation to conditional distribution of Skovgaard [4] is re-expressed for the M-statistic defined in Equation (1) by [3]. This is summarized in Section 3.1. A modification for the lattice case is given in Section 3.2. A method for computing quantiles is given in Section 3.3.

3.1. Approximation to the Distribution

Consider n absolutely continuous and independent random variables $X_1, \ldots, X_n$ and the M-statistic $(S_{1,n}, S_{2,n})$ viz. $(S_{1,n}(X_1, \ldots, X_n), S_{2,n}(X_1, \ldots, X_n))$, which is the solution w.r.t. (s_1, s_2) of

$$\sum_{j=1}^{n} \begin{pmatrix} \psi_{1,j}(X_j; s_1, s_2) \\ \psi_{2,j}(X_j; s_2) \end{pmatrix} = 0, \tag{11}$$

where $\psi_{1,j}$: $\mathbb{R}^3 \to \mathbb{R}$ is a continuous function that is decreasing in its second argument and $\psi_{2,j}$: $\mathbb{R}^2 \to \mathbb{R}$ is a continuous function that is decreasing in its second argument, for $j = 1, \ldots, n$. The joint cumulant generating function (c.g.f.) of the summands in Equation (11) is given by

$$K_n(\boldsymbol{v}; \boldsymbol{s}) = \sum_{j=1}^{n} \log \mathsf{E}[\exp\{v_1 \psi_{1,j}(X_j; s_1, s_2) + v_2 \psi_{2,j}(X_j; s_2)\}], \tag{12}$$

where $\boldsymbol{v} = (v_1, v_2) \in \mathbb{R}^2$ and $\boldsymbol{s} = (s_1, s_2) \in \mathbb{R}^2$. Define also $K_{2n}(v_2; s_2) = K_n((0, v_2); (0, s_2))$. The first computational step is to find the saddlepoint $\boldsymbol{\alpha} = (\alpha_1, \alpha_2) \in \mathbb{R}^2$, which is the solution w.r.t. $\boldsymbol{v} = (v_1, v_2)$ of

$$\frac{\partial}{\partial \boldsymbol{v}} K_n(\boldsymbol{v}; \boldsymbol{s}) = 0, \tag{13}$$

and the "marginal saddlepoint" $\beta \in \mathbb{R}$, which is the solution w.r.t. v_2 of

$$\frac{\partial}{\partial v_2} K_{2n}(v_2; s_2) = 0. \tag{14}$$

Next, define

$$K_n''(\boldsymbol{v};\boldsymbol{s}) = \frac{\partial^2}{\partial \boldsymbol{v}\partial \boldsymbol{v}^\top} K_n(\boldsymbol{v};\boldsymbol{s}), \quad K_{2,n}''(v_2;s_2) = \frac{\partial^2}{\partial v_2^2} K_{2n}(v_2;s_2),$$

$$\rho(\boldsymbol{s}) = \operatorname{sgn}(\alpha_1)\{2[K_{2n}(\beta;s_2) - K_n(\boldsymbol{\alpha};\boldsymbol{s})]\}^{\frac{1}{2}} \text{ and } \sigma(\boldsymbol{s}) = \alpha_1 \left(\frac{\det K_n''(\boldsymbol{\alpha};\boldsymbol{s})}{K_{2,n}''(\beta;s_2)}\right)^{\frac{1}{2}}. \tag{15}$$

With these quantities, we obtain the saddlepoint approximation

$$P_n(s_1 \mid s_2) = 1 - \Phi \circ \rho(\boldsymbol{s}) + \phi \circ \rho(\boldsymbol{s}) \left(\frac{1}{\sigma(\boldsymbol{s})} - \frac{1}{\rho(\boldsymbol{s})}\right), \tag{16}$$

where ϕ and Φ are the standard normal density and distribution function. Then,

$$\mathsf{P}[S_{1,n} \geq s_1 \mid S_{2,n} = s_2] = P_n(s_1 \mid s_2)\{1 + \mathrm{O}(n^{-1})\}, \text{ as } n \to \infty. \tag{17}$$

Thus, the saddlepoint approximation in Equation (16) possesses a vanishing relative error and at any value of the argument s_1, that is, over large deviations regions.

By selecting $X_1, \ldots, X_n$ from any one of the four conditional representations, M-P, MH-B, of MP-NB of Section 2.1 or D-G of Section 2.3, and by setting $\psi_{1,j}(x;s_1,s_2) = \xi_j(x;s_1)$ and $\psi_{2,j}(x;s_2) = x - s_2$, for $j = 1, \ldots, n$, we obtain

$$\mathsf{P}[T_n \geq t] = P_n\left(t \middle| \frac{k}{n}\right)\{1 + \mathrm{O}(n^{-1})\}, \text{ as } n \to \infty, \tag{18}$$

for T_n defined in Equation (1).

Precisely, it follows from the conditional representation in Equation (2) that

$$T_n(Y_1, \ldots, Y_n) \sim \left(S_{1,n}(X_1, \ldots, X_n) \middle| S_{2,n}(X_1, \ldots, X_n) = \frac{k}{n}\right).$$

This equivalence and Equation (17) give Equation (18).

The argument s_2 of $\psi_{1,j}(x;s_1,s_2)$ is not considered here, but it is useful in one example in [3].

As mentioned, the justification of this saddlepoint approximation can be found in [4] and it would be too long to reproduce it here. However, we can give a few general ideas. Let us consider $(U_1, V_1), \ldots, (U_n, V_n)$ independent and identically distributed (i.i.d.), absolutely continuous and with joint c.g.f. K. Let $(\bar{U}, \bar{V})$ denote their sample mean. Then, the Fourier inversion and integration of the joint density gives

$$\mathsf{P}[\bar{V} \geq v \mid \bar{U} = u] = \left(\frac{n}{2\pi \mathrm{i}}\right)^2 \int_{c-\mathrm{i}\infty}^{c+\mathrm{i}\infty} \int_{\mathrm{i}\infty}^{\mathrm{i}\infty} \exp\{n[K(s,t) - su - tv]\} \mathrm{d}s \frac{\mathrm{d}t}{nt},$$

for $u, v \in \mathbb{R}$ and $c > 0$. For the integral w.r.t. s, a standard saddlepoint approximation is used. The resulting saddlepoint approximation is an integral w.r.t. t and, due to a singularity, a modified saddlepoint approximation similar to the one in [18] must used to approximate this integral. The generalization from the sample mean to the M-statistic in Equation (11) follows directly from

$$\mathsf{P}[S_{1,n} \geq s_1 \mid S_{2,n} = s_2] = \mathsf{P}\left[\sum_{j=1}^n \psi_{1,j}(X_j, s_1, s_2) \geq 0 \middle| \sum_{j=1}^n \psi_{2,j}(X_j, s_2) = 0\right],$$

for $s_1, s_2 \in \mathbb{R}$, which is due to the fact that $\psi_{1,j}$ and $\psi_{2,j}$ are decreasing in their second argument.

3.2. Modifications for Discrete Statistics

A slight modification of this saddlepoint approximation for the case where T_n takes values in the lattice $\{j\delta/n\}_{j\in\mathbb{Z}}$, for some $\delta > 0$, is obtained by replacing $\sigma(\boldsymbol{s})$ in Equation (16) by

$$\ddot{\sigma}(\boldsymbol{s}) = (1 - \exp\{-\delta\alpha_1\}) \left(\frac{\det K_n''(\boldsymbol{\alpha}; \boldsymbol{s})}{K_{2n}''(\beta; s_2)} \right)^{\frac{1}{2}}. \tag{19}$$

Moreover, the following continuity correction can be considered. For the lattice point s_1, define $\tilde{s}_1 = s_1 - \delta/(2n)$, $\tilde{\boldsymbol{s}} = (\tilde{s}_1, s_2)$ and $\tilde{\boldsymbol{\alpha}} = (\tilde{\alpha}_1, \tilde{\alpha}_2)$ as the solution w.r.t. $\boldsymbol{v}$ of

$$\frac{\partial}{\partial \boldsymbol{v}} K_n(\boldsymbol{v}; \tilde{\boldsymbol{s}}) = 0.$$

Then, replace $\rho(\boldsymbol{s})$ and $\sigma(\boldsymbol{s})$ in Equation (16) by

$$\tilde{\rho}(\tilde{\boldsymbol{s}}) = \operatorname{sgn}(\tilde{\alpha}_1)\{2[K_{2n}(\beta; s_2) - K_n(\tilde{\boldsymbol{\alpha}}; \tilde{\boldsymbol{s}})]\}^{\frac{1}{2}} \text{ and } \tilde{\sigma}(\tilde{\boldsymbol{s}}) = 2\sinh\left(\frac{\delta}{2}\tilde{\alpha}_1\right) \left(\frac{\det K_n''(\tilde{\boldsymbol{\alpha}}; \tilde{\boldsymbol{s}})}{K_{2n}''(\beta; s_2)} \right)^{\frac{1}{2}},$$

respectively. The justifications can be found in [4,19]. The relative error of these modified approximations remains $\mathrm{O}(n^{-1})$.

3.3. Approximation to Quantiles

Define $\zeta(\boldsymbol{s}) = \rho(\boldsymbol{s}) + \log\{\sigma(\boldsymbol{s})/\rho(\boldsymbol{s})\}/\rho(\boldsymbol{s})$, for ρ and σ defined in Equation (15). An asymptotically equivalent version of the saddlepoint approximation in Equation (16) is given by $P_n^*(s_1 \mid s_2) = 1 - \Phi \circ \zeta(\boldsymbol{s})$. This formula leads to a fast algorithm for approximating quantiles, with same asymptotic error as the one entailed by exact inversion of the saddlepoint approximation. The general idea of Wang [31] was adapted to the present situation by Gatto [5].

Let $\varepsilon \in (0, 1)$. One starts with any reasonable approximation to the desired ε-quantile, for example the normal one, given by

$$s_1^{(0)}(\varepsilon) = \frac{\tau(s_2)}{\sqrt{n}} \Phi^{(-1)}(\varepsilon) + \mu(s_2),$$

where $\mu(s_2) \simeq \mathrm{E}[S_{1,n} \mid S_{2,n} = s_2]$ and $\tau^2(s_2) \simeq n\mathrm{var}(S_{1,n} \mid S_{2,n} = s_2)$.

Re-denote by $\boldsymbol{\alpha}(\boldsymbol{s})$ the saddlepoint at $\boldsymbol{s}$, viz. the solution of Equation (13) w.r.t. $\boldsymbol{v}$. Denote $\dot{K}_n(\boldsymbol{v}; \boldsymbol{s}) = \partial/\partial \boldsymbol{s}\, K_n(\boldsymbol{v}; \boldsymbol{s})$. One computes, for $j = 0, 1$,

$$s_1^{(j+1)}(\varepsilon) = s_1^{(j)}(\varepsilon) + \frac{\{\Phi^{(-1)}(\varepsilon)\}^2 - \zeta^2(\boldsymbol{s}^{(j)}(\varepsilon))}{-2\{\dot{K}_n(\boldsymbol{\alpha}(\boldsymbol{s}^{(j)}(\varepsilon)); \boldsymbol{s}^{(j)}(\varepsilon))\}_1}, \tag{20}$$

where $\boldsymbol{s}^{(j)}(\varepsilon) = (s_1^{(j)}(\varepsilon), s_2)$. If $s_1(\varepsilon)$ denotes the exact ε-quantile, then

$$s_1^{(2)}(\varepsilon) = s_1(\varepsilon)\{1 + \mathrm{O}(n^{-\frac{3}{2}})\}, \text{ as } n \to \infty.$$

Moreover, if $\tilde{s}_1(\varepsilon)$ denotes the ε-quantile obtained by exact inversion of the saddlepoint distribution, then $s_1^{(2)}(\varepsilon) = \tilde{s}_1(\varepsilon)\{1 + \mathrm{O}(n^{-3/2})\}$, as $n \to \infty$. Therefore, stopping the iteration of Equation (20) at $j = 1$ is sufficient in terms of asymptotic accuracy.

Consider the simple case $\psi_{1,j}(x; s_1, s_2) = g(x) - s_1$, for some continuous function g: $\mathbb{R} \to \mathbb{R}$. Then, Equation (11) yields $S_{1,n}(X_1, \ldots, X_n) = n^{-1} \sum_{j=1}^{n} g(X_j)$. In this situation, the denominator of the ratio in Equation (20) simplifies to $2\{\boldsymbol{\alpha}(\boldsymbol{s}^{(j)}(\varepsilon))\}_1$.

4. Applications

This section presents various examples that illustrate the relevance and accuracy of the conditional saddlepoint approximation for M-statistics of Section 3 with the M-P, MH-B, MP-NB and D-G representations of Section 2, respectively, in Sections 4.1–4.4. Important applications or examples from previous articles are summarized and novel examples are developed. The common urn sampling model of all examples is always put in the forefront. Many examples are studied numerically. The values obtained by the saddlepoint approximation are always very close to the ones obtained by Monte Carlo simulation. This section is however not a complete list of applications: further examples can be found, e.g., in [8,9] (Chapter 4 and Section 12.5).

As mentioned, the accuracy of the saddlepoint approximation is assessed through comparisons with simple Monte Carlo simulation. The following measures of accuracy for approximating the distribution of the statistic T_n are considered. Let $t > 0$. The probabilities obtained by simulation are considered as exact and denoted $\mathsf{P}_{\mathrm{E}}[T_n < t]$. The probabilities obtained by the saddlepoint approximation are denoted $\mathsf{P}_{\mathrm{S}}[T_n < t]$. Then,

$$\mathrm{ae}(t) = |\mathsf{P}_{\mathrm{S}}[T_n < t] - \mathsf{P}_{\mathrm{E}}[T_n < t]| = |\mathsf{P}_{\mathrm{S}}[T_n \geq t] - \mathsf{P}_{\mathrm{E}}[T_n \geq t]| \tag{21}$$

denotes the absolute error and

$$\mathrm{re}(t) = \frac{|\mathsf{P}_{\mathrm{S}}[T_n < t] - \mathsf{P}_{\mathrm{E}}[T_n < t]|}{\min\{\mathsf{P}_{\mathrm{E}}[T_n < t], 1 - \mathsf{P}_{\mathrm{E}}[T_n < t]\}} = \frac{|\mathsf{P}_{\mathrm{S}}[T_n \geq t] - \mathsf{P}_{\mathrm{E}}[T_n \geq t]|}{\min\{\mathsf{P}_{\mathrm{E}}[T_n \geq t], 1 - \mathsf{P}_{\mathrm{E}}[T_n \geq t]\}} \tag{22}$$

denotes the absolute relative error.

4.1. Sampling with Replacement and M-P Representation

Three new illustrations of the saddlepoint approximation with the M-P representation are presented. Example 1 considers the entropy of the coloration probabilities of the balls of the urn. Numerical evaluations of the saddlepoint approximation to the distribution of the estimator of the entropy are given. Example 2 concerns the likelihood ratio test for the null hypothesis of equality of the coloration probabilities. The power under a particular alternative hypothesis is computed numerically. Example 3 considers the insurer total claim amount when the individual claim settlement is delayed. The saddlepoint approximation to the distribution of the total claim amount is analyzed numerically. Example 4 reviews the application of the saddlepoint approximation to the bootstrap distribution of the M-statistic in Equation (1).

Example 1 (Entropy's estimator under sampling with replacement)**.** *The mathematical study of entropy began with Shannon [32], for the construction of a model for the transmission of information. In sampling with replacement from the urn, the probability of drawing a ball of color $\mathcal{C}_j$ is fixed and given by $p_j = a_{j,0}/z$, for $j = 1, \ldots, n$. Define $\boldsymbol{p} = (p_1, \ldots, p_n) \in \Delta_1^{n-1}$. The entropy of the coloration is given by*

$$\varepsilon_n(\boldsymbol{p}) = -\sum_{j=1}^{n} p_j \log p_j, \tag{23}$$

where $0 \log 0 = 0$ is assumed. The entropy $\varepsilon_n(\boldsymbol{p})$ is an appropriate measure of the uncertainty about the colors of the drawn balls. Indeed, it satisfies the following properties. First, $\varepsilon_n(\boldsymbol{p})$ takes its largest value $\log n$ for $p_1 = \ldots = p_n = n^{-1}$. Second, if we consider the equivalent coloration $\mathcal{C}_1, \ldots, \mathcal{C}_n, \mathcal{C}_{n+1}$ with probabilities $p_1, \ldots, p_n$ and $p_{n+1} = 0$, respectively, then $\varepsilon_n(p_1, \ldots, p_n) = \varepsilon_{n+1}(p_1, \ldots, p_n, p_{n+1})$. Theorem 1 on pp. 9–10 of [33] states that the only continuous function that satisfies these two properties plus another one related to conditional entropy, has the form given in Equation (23) multiplied by a positive constant.

As in Section 2.2, $Y_1, \ldots, Y_n$ denotes the number of drawn balls for each of the n colors $\mathcal{C}_1, \ldots, \mathcal{C}_n$, respectively, after $k \in \mathbb{N}^*$ draws. Define

$$T_n(Y_1, \ldots, Y_n) = \varepsilon_n\left(\frac{Y_1}{k}, \ldots, \frac{Y_n}{k}\right) = -\sum_{j=1}^{n} \frac{Y_j}{k} \log \frac{Y_j}{k} = -\frac{1}{k} \sum_{j=1}^{n} Y_j \log Y_j + \log k \tag{24}$$

and $P_n(Y_1, \ldots, Y_n) = \binom{k}{Y_1 \ldots Y_n} n^{-Y_1} \ldots n^{-Y_n}$, that is, the multinomial probability of the configuration $(Y_1, \ldots, Y_n)$ under uniformity. It is directly shown that $k^{-1} \log P_n = T_n + \mathrm{o}(1)$, as $k \to \infty$ and a.s. Asymptotically, the entropy of the configuration is thus an increasing transform of the probability of the configuration under uniformity. The probability P_n is maximized by the constant configuration and so is the entropy T_n.

Consider now the multinomial model in Equation (3) with unknown vector of probabilities $\boldsymbol{p}$. The frequency Y_j/k is an unbiased estimator of p_j, for $j = 1, \ldots, n$. Thus, T_n is an estimator of the entropy $\varepsilon_n(\boldsymbol{p})$. It takes the form of the M-statistic in Equation (1) with $\xi_j(y; t) = -y \log y + n^{-1} k \log k - n^{-1} kt$. Using the M-P representation and some algebraic manipulations, the c.g.f. in Equation (12) takes the form

$$K_n(\boldsymbol{v}; \boldsymbol{s}) = k(\log k - s_1)v_1 - ns_2 v_2 - q + \sum_{j=1}^{n} \log \left\{ 1 + \sum_{l=1}^{\infty} \frac{1}{l!} \left(q p_j \mathrm{e}^{v_2} l^{-v_1}\right)^l \right\},$$

with $q \in \mathbb{R}_+^*$ arbitrary. We set $s_2 = k/n$ and select q such that $\mathsf{E}[S_{2,n}] = k/n$, i.e., $q = k$. With this choice of q, the marginal saddlepoint equation, cf. Equation (14), has the trivial solution $\beta = 0$. This yields

$$K_n\left(\boldsymbol{v}; \left(s_1, \frac{k}{n}\right)\right) = k\left\{(\log k - s_1)v_1 - v_2 - 1\right\} + \sum_{j=1}^{n} \log \left\{ 1 + \sum_{l=1}^{\infty} \frac{1}{l!} \left(k p_j \mathrm{e}^{v_2} l^{-v_1}\right)^l \right\}. \tag{25}$$

Computing the second order derivatives is long but basic. We only give the simple result $K''_{2,n}(0; (s_1, k/n)) = k$; it can be used for controlling the formula of the second derivative.

We can now apply the saddlepoint approximation of Section 3 to the following case: $p_j = 2j/\{n(n+1)\}$, for $j = 1, \ldots, n$, $n = 6$ and $k = 32$. The saddlepoint approximation is compared with the Monte Carlo distribution of T_6 based on 10^6 simulations. The numerical results are displayed in Figure 1 and Table 1. The probabilities obtained by simulation are denoted $\mathsf{P}_{\mathrm{E}}[T_6 < t]$, the probabilities obtained by the saddlepoint approximation are denoted $\mathsf{P}_{\mathrm{S}}[T_6 < t]$, $\mathrm{ae}(t)$ denotes the absolute error defined in Equation (21) and $\mathrm{re}(t)$ denotes the absolute relative error defined in Equation (22), for $t \in [1.20, 1.77]$. We see that the relative errors are mostly very small. The largest one occurs in the extreme right tail and it is around 31%.

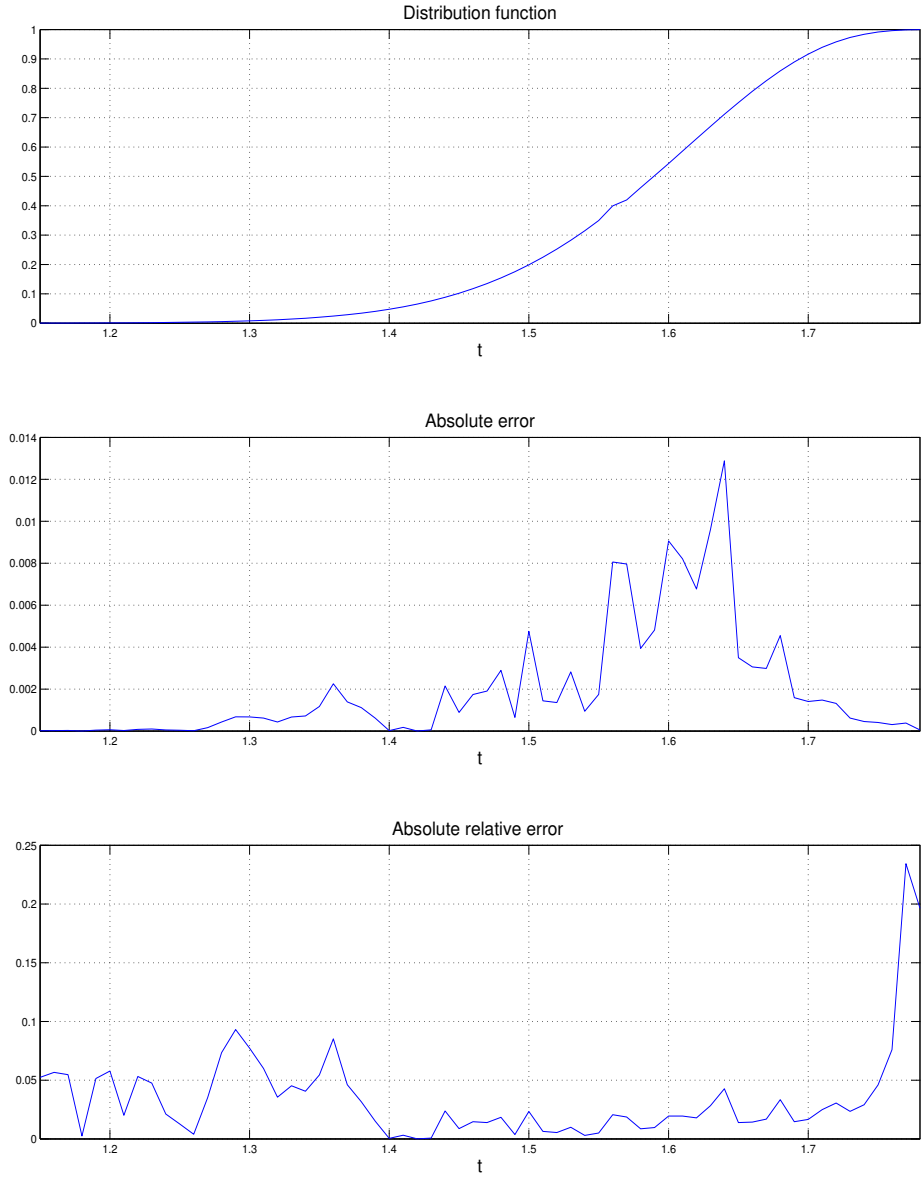

Figure 1. Estimator of coloration's entropy under sampling with replacement (T_6). First graph: saddlepoint approximation to the distribution function, $\mathrm{P_S}[T_6 < t]$. Second graph: absolute error, $\mathrm{ae}(t)$. Third graph: absolute relative error, $\mathrm{re}(t)$.

Table 1. Estimator of coloration's entropy under sampling with replacement (T_6), selected lower and upper tail points; Monte Carlo probability ($\mathrm{P_E}$), saddlepoint probability ($\mathrm{P_S}$), absolute relative error (re).

t	$\mathrm{P_E}[T_6 < t]$	$\mathrm{P_S}[T_6 < t]$	$\mathrm{re}(t)$
1.20	0.00109	0.00103	0.058
1.31	0.01034	0.00972	0.060
1.36	0.02648	0.02422	0.085
1.40	0.04755	0.04753	0.000
1.45	0.10116	0.10205	0.009
1.69	0.89118	0.88959	0.014
1.72	0.95691	0.95823	0.032
1.73	0.97360	0.97303	0.023
1.75	0.99114	0.99155	0.048
1.77	0.99838	0.99876	0.306

Example 2 (Power of likelihood ratio test)**.** *The estimator of entropy in Equation (24) is closely related to the likelihood ratio test. Consider a sample of k i.i.d. random variables and consider any partition of their range that is made by n intervals of positive length. Denote by p_j the probability that any one of the sample values belongs to the jth interval, for $j = 1, \ldots, n$. Denote by Y_j the number of sample values that belong to the jth interval, for $j = 1, \ldots, n$. Then, $(Y_1, \ldots, Y_n)$ takes values in $\ddot{\Delta}_k^{n-1}$ and follows the multinomial distribution in Equation (3). Consider the null hypothesis H_0: $\boldsymbol{p} \in \Pi_0$, where $\Pi_0 \subset \Delta_1^{n-1}$. The likelihood ratio test statistic for H_0 against the general alternative is given by*

$$L_n(Y_1, \ldots, Y_n) = \frac{\sup_{p \in \Pi_0} \left\{ \frac{k!}{Y_1! \ldots Y_n!} p_1^{Y_1} \ldots p_n^{Y_n} \right\}}{\sup_{p \in \Delta_1^{n-1}} \left\{ \frac{k!}{Y_1! \ldots Y_n!} p_1^{Y_1} \ldots p_n^{Y_n} \right\}}.$$

By restricting to $\Pi_0 = \{\boldsymbol{p}_0\}$, for some $\boldsymbol{p}_0 \in \Delta_1^{n-1}$, we obtain

$$T_n^*(Y_1, \ldots, Y_n) = -2 \log L_n(Y_1, \ldots, Y_n) = 2 \sum_{j=1}^{n} Y_j \log Y_j - 2 \sum_{j=1}^{n} Y_j \log p_{0,j} - 2k \log k. \tag{26}$$

In the case $p_{0,1} = \ldots = p_{0,n} = n^{-1}$, which can be obtained without loss of generality by the probability integral transform, $T_n^*(Y_1, \ldots, Y_n)$ is equal to $2 \sum_{j=1}^{n} Y_j \log Y_j$ plus a constant term. Then, $T_n^* \xrightarrow{\mathrm{d}} \chi^2_{n-1}$, as $k \to \infty$. In addition, if $k, n \to \infty$, with $k/n \to l$, for some $l \in (1, \infty)$, then T_n^* is asymptotically normal.

The numerical evaluation of the saddlepoint approximation to the distribution of T_n^*, with $n = 4$, $k = 12$ and under H_0, is given in Table 1 in [5]. We now extend the numerical study to the evaluation of the power function at any point of the alternative, viz. at any $\boldsymbol{p} \in \Delta_1^{n-1} \backslash \{n^{-1}, \ldots, n^{-1}\}$. Because T_n^* is an affine transform of the entropy estimator T_n given in Equation (24), we rather consider T_n as test statistic. Thus, the c.g.f. for the saddlepoint approximation is already given in Equation (25). Consider the power function at the point of alternative hypothesis $p_j = 2j/\{n(n+1)\}$, for $j = 1, \ldots, n$. We select $n = 6$ and $k = 32$. The saddlepoint approximation to the distribution of T_6 under H_0 gives

$$\mathrm{P_S}[T_6 < 1.6060] = 0.0495.$$

The saddlepoint approximation to the distribution of T_6 under the chosen alternative point gives

$$\mathrm{P_S}[T_6 < 1.6060] = 0.5691.$$

This distribution is computed in Example 1. Thus, 0.5691 is the saddlepoint approximation to the power of the test with size 0.0495 at the given alternative.

In situations where Π_0 is the singleton containing the vector of unequal elements $p_{0,1}, \ldots, p_{0,n}$, the saddlepoint approximation can be obtained in a similar way. An important application is with language identification, where these probabilities represent the frequencies of the n letters of the alphabet of a language and $Y_1/k, \ldots, Y_n/k$ are the frequencies of these n letters within a text of k letters. The belonging of the text to the language can be tested with the statistic T_n^*, which is in fact proportional to the Kullback–Leibler information. Precisely, denote

$$\iota_n(\boldsymbol{v}|\boldsymbol{w}) = \sum_{j=1}^{n} v_j \log \frac{v_j}{w_j}$$

the Kullback–Leibler information or discrepancy between the two probability distributions $\boldsymbol{v} = (v_1, \ldots, v_n) \in \Delta_1^{n-1}$ and $\boldsymbol{w} = (w_1, \ldots, w_n) \in \Delta_1^{n-1}$, that satisfy the absolute continuity condition $w_j = 0 \Rightarrow v_j = 0$, for $j = 1, \ldots, n$. Then, $T_n^* = 2k\, \iota_n(Y_1/k, \ldots, Y_n/k \mid p_{0,1}, \ldots, p_{0,n})$.

Example 3 (Total claim amount under delayed settlement)**.** *We are interested in the distribution of the total claim amount of an insurance company over a fixed time horizon. We assume that the delay of claim settlement increases as the individual claim amount increases. This can happen in actuarial practice, partially because large claim amounts require longer controls. Precisely, the individual claim amounts are i.i.d. random variables taking the n values $r_1 < \ldots < r_n$, all in $\mathbb{R}_+^*$, for $n = 2, 3, \ldots$. Let $j \in \{1, \ldots, n\}$. Claims of amount r_j are settled exactly after the jth unit of time (e.g., months). During a given time horizon (e.g., a year), Y_j claims of amount r_j occur. We assume that $k \in \mathbb{N}^*$ claims have occurred during the time horizon under consideration and that $(Y_1, \ldots, Y_n)$, which takes values in $\ddot{\Delta}_k^{n-1}$, follows the multinomial distribution in Equation (3). The total claim amount settled during the time horizon is thus $\sum_{j=1}^{n} r_j Y_j$. We are interested in the distribution of the proportion of total claim amount that is settled exactly after the mth unit of time, viz. of*

$$T_n = T_n(Y_1, \ldots, Y_n) = \frac{\sum_{j=1}^{m} r_j Y_j}{\sum_{j=1}^{n} r_j Y_j}, \tag{27}$$

for some $m \in \{1, \ldots, n-1\}$. It can be re-expressed as the M-statistic in Equation (1) with

$$\xi_j(y; t) = r_j(\mathrm{I}\{j \le m\} - t)y, \;\; \text{for} \;\; j = 1, \ldots, n.$$

The M-P representation tells that the multinomial claim counts have the distribution of independent Poisson occurrences, given a total of k claim occurrences. Thus, with some algebraic manipulations, the c.g.f. in Equation (12) becomes

$$\begin{aligned} K_n(\boldsymbol{v}; \boldsymbol{s}) \;\; = \;\; & -ns_2v_2 - q \\ & + \sum_{j=1}^{n} \log \left(1 + \sum_{l=1}^{\infty} \frac{1}{l!} \exp\{[v_1 r_j(\mathrm{I}\{j \le m\} - s_1) + v_2 + \log(qp_j)]l\} \right), \end{aligned}$$

with arbitrary $q \in \mathbb{R}_+^*$. We set $s_2 = k/n$ and select q such that $\mathsf{E}[S_{2,n}] = k/n$, i.e., $q = k$. Thus, the marginal saddlepoint equation, cf. Equation (14), is solved by $\beta = 0$. This leads to

$$\begin{aligned} K_n\left(\boldsymbol{v}; \left(s_1, \frac{k}{n}\right)\right) \;\; = \;\; & -k(1 + v_2) \\ & + \sum_{j=1}^{n} \log \left(1 + \sum_{l=1}^{\infty} \frac{1}{l!} \exp\{[v_1 r_j(\mathrm{I}\{j \le m\} - s_1) + v_2 + \log(kp_j)]l\} \right). \end{aligned}$$

By computing the second order derivatives, we find $K''_{2,n}(0; (s_1, k/n)) = k$.

For the numerical illustration, consider the multinomial distribution in Equation (3) with probabilities $p_1 = 0.15$, $p_2 = 0.23$, $p_3 = 0.16$, $p_4 = 0.14$, $p_5 = 0.12$, $p_6 = 0.1$, $p_7 = 0.06$, *and* $p_8 = 0.04$ *and the total number of* $k = 30$ *claims. Thus,* $n = 8$ *and the possible claim amounts are* $r_1 = 10$, $r_2 = 15$, $r_3 = 20$, $r_4 = 30$, $r_5 = 50$, $r_6 = 70$, $r_7 = 100$ *and* $r_8 = 120$. *The number of unit of times for the proportion of settled total claim amount, cf. Equation (27), is* $m = 4$. *To assess the accuracy of the saddlepoint approximation, we compute the Monte Carlo distribution of* T_8, *based on* 10^6 *simulations. The numerical results are displayed in Table 2. The probabilities obtained by simulation are denoted* $\mathsf{P}_{\mathrm{E}}[T_8 < t]$, *the probabilities obtained by the saddlepoint approximation are denoted* $\mathsf{P}_{\mathrm{S}}[T_8 < t]$ *and* $\mathrm{re}(t)$ *denotes the relative error, cf. Equation (22), for* $t \in [0.12, 0.72]$. *Most relative errors are below 5%. The largest one occurs in the extreme left tail and is approximatively* 12%.

Table 2. Proportion of total claim amount (T_8): Monte Carlo probability (P_{E}), saddlepoint probability (P_{S}), and absolute relative error (re).

t	$\mathsf{P}_{\mathrm{E}}[T_8 < t]$	$\mathsf{P}_{\mathrm{S}}[T_8 < t]$	$\mathrm{re}(t)$
0.12	0.00035	0.00040	0.124
0.16	0.00523	0.00555	0.059
0.20	0.03119	0.03261	0.044
0.24	0.10347	0.10508	0.016
0.28	0.23178	0.23278	0.004
0.32	0.39024	0.39775	0.019
0.38	0.62988	0.64044	0.029
0.42	0.76566	0.76905	0.015
0.46	0.85601	0.85930	0.023
0.48	0.89281	0.89184	0.009
0.52	0.93960	0.93993	0.006
0.56	0.96733	0.96784	0.016
0.60	0.98294	0.98352	0.035
0.64	0.99137	0.99166	0.035
0.68	0.99577	0.99578	0.003
0.72	0.99799	0.99791	0.037

A practical question would be the following: Which value of t bounds from above the proportion of total claim amount T_8 *with probability 0.99? One computes directly* $\mathsf{P}_{\mathrm{S}}[T_8 < 0.63] = 0.9897$ *and thus* $t = 0.63$, *approximately.*

Example 4 (Bootstrap distribution of M-statistic)**.** *Let* $R_1, \ldots, R_n$ *be a sample of i.i.d. random variables taking values in* $\mathbb{R}$, *for* $n = 2, 3, \ldots$. *Absolute continuity is assumed, in order to avoid repeated values a.s. Consider the M-statistic* U_n *or* $U_n(R_1, \ldots, R_n)$ *defined as the root in t on*

$$\sum_{j=1}^{n} \zeta(R_j; t) = 0,$$

where $\zeta\colon \mathbb{R}^2 \to \mathbb{R}$ *is continuous and decreasing in its second argument. Let* $r_1, \ldots, r_n$ *be a realization of the sample and let* $R_1^*, \ldots, R_n^*$ *be the random variables obtained by sampling with replacement from the values* $r_1, \ldots, r_n$ *with respective probabilities* $p_1, \ldots, p_n$, *for* $(p_1, \ldots, p_n) \in \Delta_1^{n-1}$. *The distribution of* $U_n(R_1^*, \ldots, R_n^*)$, *or simply* U_n^*, *is the bootstrap distribution of* U_n.

This coincides with sampling with replacement from the general urn model of Section 2.2, if the color $\mathcal{C}_j$ *is associated to the value* r_j, *for* $j = 1, \ldots, n$, *and if the number of draws from the urn is* $k = n$. *Define* $\xi_j(y; t) = y\zeta(r_j; t)$, *for* $t \in \mathbb{R}$, $y \in \mathbb{N}$ *and for* $j = 1, \ldots, n$. *Then,* U_n^* *can be represented as the solution w.r.t. t of Equation (1), denoted* T_n, *in which* Y_j *is the number of times that* r_j *has been sampled, for* $j = 1, \ldots, n$. *The conditional saddlepoint approximation of Section 3 yields the distribution of* T_n, *i.e., of* U_n^*, *i.e., of the bootstrap distribution of* U_n. *In most practical cases,* $p_1 = \ldots = p_n = n^{-1}$, *i.e.,* $a_{1,0} = \ldots = a_{n,0}$.

The saddlepoint approximation for bootstrap distributions was introduced by [34–36] and for M-estimators by [37]. A review can be found in [38] (Section 9.5). Thus, the conditional saddlepoint approximation of Section 3 provides an alternative saddlepoint approximation to the bootstrap distribution of M-estimators.

Other applications of this saddlepoint approximation with the M-P representation that can be found the literature are the following. Saddlepoint approximations for likelihood ratio test and for chi-square tests for grouped data, under the null hypotheses, are given in [3]. For the numerical evaluation of the saddlepoint approximation for the likelihood ratio statistic, refer to [5].

4.2. Sampling without Replacement and MH-B Representation

The saddlepoint approximation combined with the MH-B representation can be applied for approximating the distribution of the M-statistic in Equation (1) in finite population sampling, viz. under sampling without replacement. Example 5 analyzes the numerical accuracy of the saddlepoint approximation to the distribution of the coloration entropy when sampling is without replacement.

Example 5 (Entropy's estimator under sampling without replacement)**.** *We consider the entropy estimation of Example 1 in the context of sampling without replacement. We are interested in the coloration entropy $\varepsilon_n(a_{1,0}/z, \ldots, a_{n,0}/z)$, as given by Equation (23), with $a_{1,0}, \ldots, a_{n,0}$ unknown. It is the entropy of the initial state of the urn. In the multivariate hypergeometric model in Equation (4), Y_j/k is an unbiased estimator of $a_{j,0}/z$, for $j = 1, \ldots, n$, where $(Y_1, \ldots, Y_n)$ takes values in $\ddot{\Delta}_k^{n-1} \cap ([0, m_1] \times \ldots \times [0, m_n])$. Thus, an estimator of this entropy is given by Equation (24). The unknown parameters of the multivariate hypergeometric distribution in Equation (4) are $m_j = a_{j,0}$, for $j = 1, \ldots, n$.*

With the MH-B representation and some algebraic manipulations, the c.g.f. in Equation (12) becomes

$$K_n(\boldsymbol{v}; \boldsymbol{s}) = k(\log k - s_1)v_1 - ns_2v_2 + z\log(1-q) + \sum_{j=1}^{n} \log\left\{1 + \sum_{l=1}^{m_j} \frac{(m_j)_l}{l!}\left(\frac{q}{1-q}\mathrm{e}^{v_2}l^{-v_1}\right)^l\right\}, \quad (28)$$

with $q \in (0,1)$ arbitrary. We set $s_2 = k/n$ and select q such that $\mathsf{E}[S_{2,n}] = k/n$, i.e., $q = k/z$. For this purpose, we assume $k < z$. With this choice, the marginal saddlepoint equation, cf. Equation (14), has the trivial solution $\beta = 0$ and the c.g.f. in Equation (28) becomes

$$\begin{aligned} K_n\left(\boldsymbol{v}; \left(s_1, \frac{k}{n}\right)\right) &= k\left\{(\log k - s_1)v_1 - v_2\right\} + z\{\log(z-k) - \log z\} \\ &\quad + \sum_{j=1}^{n} \log\left\{1 + \sum_{l=1}^{m_j} \frac{(m_j)_l}{l!}\left(\frac{k}{z-k}\mathrm{e}^{v_2}l^{-v_1}\right)^l\right\}. \end{aligned}$$

The second order derivatives of K_n can be obtained through long but simple algebraic manipulations. In particular, we find $K''_{2,n}(0; (s_1, k/n)) = k(z-k)/z$.

For the numerical illustration, we consider the multivariate hypergeometric distribution with $n = 7$, $m_1 = 2$, $m_2 = 4$, $m_3 = 6$, $m_4 = 8$, $m_5 = 10$, $m_6 = 12$, $m_7 = 14$ and $k = 25$. We compute the Monte Carlo distribution of T_7 based on 10^6 simulations. The saddlepoint approximation is obtained by following the steps of Section 3. The results are given in Table 3. The saddlepoint probabilities are obtained instantaneously and we see that the relative errors are below 15%, with the exception an extreme left tail point, for which the relative error is 25%.

We now summarize two practical applications of the conditional saddlepoint approximation with the MH-B representation. The first one can be found in [39] and concerns a permutation test of comparison of two groups. The jth individual belongs to the control group, when $Y_j = 0$, and to the treatment group, when $Y_j = 1$, for $j = 1, \ldots, n$. We have $(Y_1, \ldots, Y_n) \sim$ Multi-Hypergeometric$(k; 1, \ldots, 1)$, where k is the number of individuals of the treatment group.

The realizations of $(Y_1, \ldots, Y_n)$ represent the permutations of the individuals and the test statistic T_n is a linear combination of the elements of $(Y_1, \ldots, Y_n)$. The permutation distribution of T_n is obtained from Equation (2), where $X_1, \ldots, X_n$ are i.i.d. Bernoulli random variables.

Table 3. Estimator of coloration's entropy under sampling without replacement (T_7); Monte Carlo probability (P_E), saddlepoint probability (P_S), and absolute relative error (re).

t	$\mathsf{P}_\mathrm{E}[T_n < t]$	$\mathsf{P}_\mathrm{S}[T_n < t]$	$\mathrm{re}(t)$
1.30	0.00010	0.00008	0.247
1.35	0.00034	0.00031	0.075
1.40	0.00124	0.00119	0.042
1.45	0.00377	0.00405	0.074
1.50	0.01271	0.01240	0.024
1.55	0.03340	0.03393	0.015
1.60	0.08396	0.08250	0.017
1.65	0.17648	0.17680	0.002
1.70	0.33407	0.33088	0.010
1.75	0.53940	0.53896	0.001
1.80	0.75566	0.75979	0.017
1.85	0.94083	0.93226	0.145
1.90	0.99594	0.99638	0.109

The second application can be found in [40] and concerns the jackknife distribution of a ratio. Consider the fixed sample $z_1, \ldots, z_n$, sample without replacement $1 \leq d < n$ values and define $Y_j = 0$, if z_j is not sampled, and $Y_j = 1$, if z_j is sampled, for $j = 1, \ldots, n$. This procedure is repeated many times and a statistic of interest is computed $k = n - d$ times, from the k sampled values of each iteration. In the terminology of B. Efron, this is called the delete-d jackknife. We have $(Y_1, \ldots, Y_n) \sim \text{Multi-Hypergeometric}(k; 1, \ldots, 1)$, where k is the sample size of the jackknife samples. The realizations of $(Y_1, \ldots, Y_n)$ represent the permutations of $(z_1, \ldots, z_n)$. The statistic considered in [40] is $T_n = \sum_{j=1}^n v_j Y_j / \sum_{j=1}^n u_j Y_j$, for $u_j, v_j \in \mathbb{R}$, for $j = 1, \ldots, n$. The permutation, viz. delete-d jackknife, distribution of T_n is obtained from Equation (2), where $X_1, \ldots, X_n$ are independent Bernoulli random variables with parameter $1/2$, together with the saddlepoint approximation for M-statistics of Section 3.

4.3. Polya's Sampling and MP-NB Representation

This section provides various applications of the saddlepoint approximation with the MP-NB representation. Example 6 considers the estimator of the entropy of the initial coloration probabilities of the urn, in the setting of Polya's sampling. Example 7 considers the Bayesian analysis if this entropy. The Bayesian Entropy's estimator under multivariate Polya a priori and sampling without replacement is considered. The saddlepoint approximation to this the posterior distribution of the entropy can be obtained by MP-NB representation. Example 8 concerns the saddlepoint approximation with the MP-NB representation for many two-sample tests based on spacing-frequencies.

Example 6 (Entropy's estimator under Polya's sampling)**.** *We consider the entropy estimation problem introduced in Example 1, now in the context of Polya's sampling. We are interested in the entropy of the initial coloration probabilities $\varepsilon_n(a_{1,0}/z, \ldots, a_{n,0}/z)$, given in Equation (23), where $a_{1,0}, \ldots, a_{n,0}$ are unknown. In the multivariate Polya model in Equation (5), Y_j/k is an unbiased estimator of $a_{j,0}/z$, for $j = 1, \ldots, n$, and so an estimator of the entropy is given by Equation (24). The parameters of the multivariate Polya distribution in Equation (5) are k equal to k of the urn model and $m_j = a_{j,0}/r$, for $j = 1, \ldots, n$. Using the MP-NB representation, the c.g.f. in Equation (12) becomes*

$$K_n(\boldsymbol{v};\boldsymbol{s}) = u\log q - k(s_1 - \log k) - ns_2v_2 + \sum_{j=1}^{n}\log\left\{1 + \sum_{l=1}^{\infty}\frac{(l+m_j-1)_l}{l!}\left[(1-q)\mathrm{e}^{v_2}l^{-v_1}\right]^l\right\}, \tag{29}$$

with $q \in (0,1)$ arbitrary. This formula allows for the direct evaluation of the conditional saddlepoint approximation of Section 3.

Example 7 (Bayesian Entropy's estimator under multivariate Polya a priori and sampling without replacement)**.** *The multivariate Polya distribution is often used as a prior distribution in Bayesian statistics, because it constitutes a conjugate class when associated to the multivariate hypergeometric likelihood. Precisely, consider the prior*

$$(M_1,\ldots,M_n) \sim \text{Multi-Polya}(z;\alpha_1,\ldots,\alpha_n) \tag{30}$$

taking value in $\ddot{\Delta}_z^{n-1}$, for $z \in \mathbb{N}^*$, $(\alpha_1,\ldots,\alpha_n) \in \Delta_u^{n-1}$ and $u \in \mathbb{R}_+^*$, and consider the likelihood

$$(Y_1,\ldots,Y_n) \mid \{(M_1,\ldots,M_n) = (m_1,\ldots,m_n)\} \sim \text{Multi-Hypergeometric}(k;m_1,\ldots,m_n),$$

for $(m_1,\ldots,m_n) \in \ddot{\Delta}_z^{n-1}$, $k \in \{0,\ldots,z\}$ and $(Y_1,\ldots,Y_n)$ taking values in $\ddot{\Delta}_k^{n-1} \cap ([0,m_1]\times\ldots\times[0,m_n])$. Then, the posterior is given by

$$\{(M_1,\ldots,M_n)|(Y_1,\ldots,Y_n) = (k_1,\ldots,k_n)\} \sim (k_1,\ldots,k_n) + \text{Multi-Polya}(z-k;\alpha_1+k_1,\ldots,\alpha_n+k_n), \tag{31}$$

for $(k_1,\ldots,k_n) \in \ddot{\Delta}_k^{n-1} \cap ([0,m_1]\times\ldots\times[0,m_n])$. Indeed,

$$\begin{aligned}\mathsf{P}[M_1 = m_1,\ldots,M_n = m_n|Y_1 = k_1,\ldots,Y_n = k_n] &\propto \prod_{j=1}^{n}\binom{m_j}{k_j}\binom{\alpha_j+m_j-1}{m_j}\\ &\propto \prod_{j=1}^{n}\frac{(\alpha_j+m_j-1)!}{(m_j-k_j)!}\\ &\propto \frac{\prod_{j=1}^{n}\binom{(\alpha_j+k_j)+(m_j-k_j)-1}{m_j-k_j}}{\binom{(u+k)+(z-k)-1}{z-k}},\end{aligned}$$

where the last result is in fact equal to the posterior probability. Thus, Equation (31) holds.

The underlying urn model is the sampling without replacement described, in Section 2.2, where the initial number of balls of each one of the colors $\mathcal{C}_1,\ldots,\mathcal{C}_n$, viz. $m_j = a_{j,0}$, for $j = 1,\ldots$, in the same order, is unknown. Only $z = \sum_{j=1}^{n} a_{j,0}$ is known. These initial counts are the elements of the random vector $(M_1,\ldots,M_n)$ with prior distribution in Equation (30). Sampling without replacement has led to the counts $(Y_1,\ldots,Y_n) = (k_1,\ldots,k_n)$, for the colors $\mathcal{C}_1,\ldots,\mathcal{C}_n$, in same order. The updated or posterior distribution of $(M_1,\ldots,M_n)$ is given by Equation (31).

Assume that we are interested in the entropy of the probabilities of the initial coloration. The a priori entropy is thus $T_n(M_1,\ldots,M_n) = \varepsilon_n(M_1/z,\ldots,M_n/z)$, cf. Equation (23). According to Equation (31), the a posteriori entropy is $T_n(k_1+L_1,\ldots,k_n+L_n)$, where $(L_1,\ldots,L_n) \sim \text{Multi-Polya}(z-k;\alpha_1+k_1,\ldots,\alpha_n+k_n)$.

The saddlepoint approximations to the distributions of the a priori and a posteriori entropies can be obtained by the saddlepoint approximation of Section 3 with MP-NB representation, as in Example 6. The a priori and a posteriori c.g.f. can be obtained by minor adaptations of the c.g.f. in Equation (29).

Example 8 (Two-sample tests based on spacing frequencies)**.** *Consider two independent samples: the first consisting of k independent random variables $U_1,\ldots,U_k$ with common absolutely continuous distribution P_U and the second sample consisting of l independent random variables $V_1,\ldots,V_l$ with common absolutely*

continuous distribution P_V. All these random variables have common range given by the real interval $\mathcal{I}$. We wish to test the null hypothesis H_0: $P_U = P_V$. Define $V_{(0)} = \inf \mathcal{I}$, $V_{(l+1)} = \sup \mathcal{I}$ and $V_{(1)} \leq \ldots \leq V_{(l)}$ the ordered $V_1, \ldots, V_l$. Let $n = l + 1$. The random counts

$$Y_j = \sum_{i=1}^{k} \mathrm{I}\{U_i \in [V_{(j-1)}, V_{(j)})\}, \text{ for } j = 1, \ldots, n, \tag{32}$$

are called spacing-frequencies: they provide the number of random variables $U_1, \ldots, U_k$ that lie between gaps made by $V_{(0)}, \ldots, V_{(l+1)}$. Thus, $(Y_1, \ldots, Y_n)$ takes values in $\ddot{\Delta}_k^{n-1}$ and possesses exchangeable components under H_0.

Denote by R_j the rank of the jth largest $V_1, \ldots, V_l$ in the combined sample, for $j = 1, \ldots, l$. It is easily seen that $R_j = \sum_{i=1}^{j}(Y_i + 1)$, or, $Y_j = R_j - R_{j-1} - 1$, for $j = 1, \ldots, l$. Consequently, many two-sample test statistics based on ranks can be re-expressed in terms of spacing-frequencies. Besides this, spacing-frequencies are essential for the analysis of circular data, because they are invariant w.r.t. changes of null direction and sense of rotation (clockwise or anti-clockwise) (for a review, see, e.g., [41]). Circular data are planar directions and can be re-expressed as angles in radians, so that $\mathcal{I} = [0, 2\pi)$, or any other interval of length 2π.

Holst and Rao [42] consider nonparametric test statistics of the form of

$$T_n(Y_1, \ldots, Y_n) = \sum_{j=1}^{n} h_j(Y_j), \tag{33}$$

for some Borel functions $h_1, \ldots, h_n$. If $h_1 = \ldots = h_n = h$, then the test statistic T_n is called symmetric. Under H_0, the multivariate Polya distribution in Equation (5) holds with $m_1 = \ldots = m_n = 1$. Consequently, $u = \sum_{j=1}^{n} m_j = n$ and all Polya's probabilities in Equation (5) are equal to $\binom{n+k-1}{k}^{-1}$. This is in accordance with the result of combinatorics that the number of solutions $(k_1, \ldots, k_n) \in \mathbb{N}^n$ of the equation $k_1 + \ldots + k_n = k$, i.e., card $\ddot{\Delta}_k^{n-1}$, is given by $\binom{n+k-1}{k}$. Thus, the equivalence in Equation (2) holds with the MP-NB representation, where the negative binomial reduces to the geometric distribution. Clearly, Equation (33) takes the form of the M-statistic in Equation (1) and the saddlepoint approximation of Section 3 can be applied.

We now summarize the examples presented in [3,5]. In the classical Wald–Wolfowitz run test, T_n takes the symmetric form of Equation (33) with $h(x) = \mathrm{I}\{x > 0\}$. We define a U-run in the combined sample as a maximal non-empty set of adjacent $U_1, \ldots, U_k$. Since each positive $Y_1, \ldots, Y_n$ is mapped to a different U-run and conversely, T_n yields the number of U-runs and it takes values in $\{1, \ldots, n\}$. Large values of T_n show evidence for equal spread, i.e., for H_0. [5] provides the numerical evaluation of the saddlepoint approximation to the distribution of T_n under H_0. The saddlepoint approximation to the distributions of the Wilcoxon viz. Mann–Whitney, the van der Waerden viz. normal score and the Savage viz. exponential score tests are developed in [3], The numerical study of Savage's test appears in [5]. In the context of directional data, a generalization of Rao's spacings tests (see Section 4.4) to spacing-frequencies together with the saddlepoint approximation is given in [41], which mention its saddlepoint approximation.

The so-called multispacing-frequencies are obtained by gaps of order larger than one made by $V_{(0)}, \ldots, V_{(l+1)}$. Let $g \in \mathbb{N}^*$ denote the differentiation gap order, such that $n = (l+1)/g$ is an integer. Then, the multispacing-frequencies are defined by

$$Y_j = \sum_{i=1}^{k} \mathrm{I}\{U_i \in [V_{(\{j-1\}g)}, V_{(jg)})\}, \text{ for } j = 1, \ldots, n. \tag{34}$$

In the case $g = 1$, Equation (34) coincides with the spacing-frequencies in Equation (32). As before with $g = 1$, $(Y_1, \ldots, Y_n)$ takes values in $\ddot{\Delta}_k^{n-1}$. We reconsider the null hypothesis H_0: $P_U = P_V$ and the general test statistics in Equation (33), however with the multispacing-frequencies in Equation (34). Under H_0, the multivariate Polya distribution in Equation (5) holds with $m_1 = \ldots = m_n = g$, $u = \sum_{j=1}^{n} m_j = ng$ and the MP-NB representation applies.

The saddlepoint approximation with MP-NB representation was analyzed by Gatto and Jammalamadaka [7] in the context of the asymptotically most powerful multispacing-frequencies test against a specific sequence of alternative distributions and also in the context of the test statistic defined by the sum of squared multispacing-frequencies.

It seems difficult to formulate an arbitrary alternative hypothesis in terms of a particular multivariate Polya distribution, for the multispacing-frequencies. In this sense, the conditional saddlepoint approximation with the MP-NB representation may not be easily applied to power computations.

4.4. D-G Representation

Example 9 of this section analyzes the most powerful test of symmetry of the Dirichlet distribution. The saddlepoint approximation based on the D-G representation to the distribution of the test statistic under an asymmetric alternative is developed and its numerical accuracy is studied. The Dirichlet associated to the multinomial distribution is an important conjugate class of distributions in Bayesian statistic. This is illustrated in Example 10, which presents a Bayesian bootstrap test on the entropy. The D-G representation with the conditional saddlepoint approximation allow to compute the Bayes factor of the test, without resampling. Another important application of the saddlepoint approximation with the D-G representation is for the class of one-sample tests based on spacings. This class of nonparametric tests is presented in Example 11 and has some similarities with the two-sample tests based on spacing frequencies of Example 8. Example 11 provides a summary of the applications that can be found in the literature of this saddlepoint approximation to tests based on spacings.

Example 9 (Test for Dirichlet's symmetry)**.** *The symmetric Dirichlet distribution is obtained by setting $a_1 = \ldots = a_n = a$ in Equation (9), for any $a \in \mathbb{R}_+^*$. In Bayesian statistics, symmetric priors are of particular interest in absence of prior knowledge on the individual elements, because they become exchangeable random variables. The single parameter a becomes a concentration parameter: $a = 1$ yields the uniform distribution over Δ_1^{n-1} (thus, the noninformative prior); $a > 1$ yields a concave density over Δ_1^{n-1} (thus, promoting similarity of elements); and $a < 1$ yields a convex density over Δ_1^{n-1} (thus, promoting dissimilarity of elements). For $(\bar{Y}_1, \ldots, \bar{Y}_n) \sim \text{Dirichlet}(a_1, \ldots, a_n)$, consider the testing problem of a particular symmetry against any particular asymmetric alternative. Precisely, given $a, \alpha_1, \ldots, \alpha_n \in \mathbb{R}_+^*$, where at least one the values $\alpha_1, \ldots, \alpha_n$ differs from the other ones, consider H_0: $a_1 = \ldots = a_n = a$, against H_1: $(a_1, \ldots, a_n) = (\alpha_1, \ldots, \alpha_n)$. The test of uniformity is obtained with $a = 1$. Neyman–Pearson's Lemma tells that the most powerful test has the form $T_n > t$, where T_n viz. $T_n(\bar{Y}_1, \ldots, \bar{Y}_n)$ is given by*

$$T_n(\bar{Y}_1, \ldots, \bar{Y}_n) = \sum_{j=1}^{n} (\alpha_j - a) \log \bar{Y}_j. \tag{35}$$

It is the M-statistic in Equation (1) with $\xi_j(y; t) = (\alpha_j - a) \log y - t/n$, for $j = 1, \ldots, n$. From the D-G representation and some algebraic manipulations, the c.g.f. in Equation (12) becomes

$$\begin{aligned} K_n(\boldsymbol{v}; \boldsymbol{s}) &= -s_1 v_1 - n s_2 v_2 + \tilde{\alpha} \log q - \log(q - v_2)\{\tilde{\alpha} + (\tilde{\alpha} - na) v_1\} \\ &\quad + \sum_{j=1}^{n} \{\log \Gamma(\alpha_j + [\alpha_j - a] v_1) - \log \Gamma(\alpha_j)\}, \end{aligned}$$

where $\tilde{\alpha} = \sum_{j=1}^{n} \alpha_j$ and $q \in \mathbb{R}_+^*$ arbitrary. We set $s_2 = 1/n$ and select q such that $\mathsf{E}[S_{2,n}] = 1/n$, i.e. $q = \tilde{\alpha}$. The marginal saddlepoint equation, cf. Equation (14), has then $\beta = 0$ as solution. This leads to

$$\begin{aligned} K_n\left(\boldsymbol{v}; \left(s_1, \frac{k}{n}\right)\right) &= -s_1 v_1 - v_2 + \tilde{\alpha} \log \tilde{\alpha} - \log(\tilde{\alpha} - v_2)\{\tilde{\alpha} + (\tilde{\alpha} - na) v_1\} \\ &\quad + \sum_{j=1}^{n} \{\log \Gamma(\alpha_j + [\alpha_j - a] v_1) - \log \Gamma(\alpha_j)\}. \end{aligned} \tag{36}$$

The second order derivatives of K_n can be expressed in terms of polygamma functions $\psi^{(n)}(z) = (\mathrm{d}/\mathrm{d}z)^{n+1} \log\Gamma(z)$, for $n = 0, 1$. We skip the details but note that $K''_{2,n}(0; (s_1, k/n)) = \tilde{\alpha}^{-1}$.

In the following numerical illustration, $a = 1$ and $\alpha_j = j$, for $j = 1, \ldots, 5$, so $n = 5$. The saddlepoint approximation is computed under H_1, so it gives the power of the test. It is compared with the Monte Carlo distribution of T_5 with 10^6 simulations. The numerical results are displayed in Table 4. The probabilities obtained by simulation are denoted $\mathrm{P_E}[T_5 < t]$, the probabilities obtained by the saddlepoint approximation are denoted $\mathrm{P_S}[T_5 < t]$ and $\mathrm{re}(t)$ denotes the absolute relative error given in Equation (22), for t in the lower and in the upper tails of the distribution. The relative errors of both lower and upper tails do not exceed 7%.

Table 4. Most powerful test statistic for Dirichlet's symmetry (T_5), selected lower and upper tail points: Monte Carlo probability ($\mathrm{P_E}$), saddlepoint probability ($\mathrm{P_S}$), and absolute relative error (re).

t	$\mathbf{P_E}[T_5 < t]$	$\mathbf{P_S}[T_5 < t]$	$\mathbf{re}(t)$
−20.5	0.00105	0.00110	0.044
−18.5	0.00976	0.01017	0.042
−17.6	0.02593	0.02661	0.026
−17.0	0.04814	0.04952	0.029
−16.3	0.09709	0.09951	0.025
−13.4	0.90156	0.90255	0.010
−13.2	0.95632	0.95753	0.029
−13.1	0.97661	0.97727	0.029
−13.0	0.99104	0.99090	0.015
−12.9	0.99833	0.99820	0.070

Example 10 (Bayesian bootstrap and Bayesian entropy test)**.** *In Bayesian statistics, Dirichlet and multinomial distributions are conjugate: Dirichlet prior and multinomial likelihood lead to Dirichlet posterior. Precisely, if*

$$(\bar{Y}_1, \ldots, \bar{Y}_n) \sim \mathrm{Dirichlet}(a_1, \ldots, a_n) \tag{37}$$

and

$$\{(Y_1, \ldots, Y_n) \mid (\bar{Y}_1, \ldots, \bar{Y}_n) = (\bar{y}_1, \ldots, \bar{y}_n)\} \sim \mathrm{Multinomial}(k; \bar{y}_1, \ldots, \bar{y}_n),$$

then

$$\{(\bar{Y}_1, \ldots, \bar{Y}_n) | (Y_1, \ldots, Y_n) = (y_1, \ldots, y_n)\} \sim \mathrm{Dirichlet}(a_1 + y_1, \ldots, a_n + y_n), \tag{38}$$

$\forall a_1, \ldots, a_n \in \mathbb{R}^*_+$ and $(y_1, \ldots, y_n) \in \ddot{\Delta}^{n-1}_k$.

The Bayesian bootstrap was introduced by Rubin [43] as a method for approximating the posterior distribution of a random parameter; precisely the distribution of a function of $\bar{Y}_1, \ldots, \bar{Y}_n$, given the observed data $(Y_1, \ldots, Y_n) = (y_1, \ldots, y_n)$. It consists in sampling of $(\bar{Y}_1, \ldots, \bar{Y}_n)$ from Equation (38). This can be done by generating $Z_j \sim \mathrm{Gamma}(a_j + y_j, q)$, for $j = 1, \ldots, n$, independently, and by setting

$$\bar{Y}_j = \frac{Z_j}{\sum_{i=1}^n Z_i}, \quad \text{for } j = 1, \ldots, n. \tag{39}$$

The value of $q \in \mathbb{R}^*_+$ is irrelevant. Details can be found in Section 10.5 of [38]. Assume that the parameter of interest is $T_n(\bar{Y}_1, \ldots, \bar{Y}_n)$ that admits the M-statistic representation in Equation (1), then the saddlepoint approximation with the D-G representation can be used instead of the described sampling method.

Consider now the urn model of Section 2.2 with sampling with replacement, where the probability of drawing a ball of color $\mathcal{C}_j$ is given by the random variable $\bar{Y}_j$, for $j = 1, \ldots, n$. We are interested in the entropy $\varepsilon_n(\bar{\boldsymbol{Y}})$, viz. Equation (23) as a function of $\bar{\boldsymbol{Y}} = (\bar{Y}_1, \ldots, \bar{Y}_n)$. According to Equation (10), $\varepsilon_n(\bar{\boldsymbol{Y}})$ is the entropy of the sample proportions of colors $\mathcal{C}_1, \ldots, \mathcal{C}_n$ under Polya's sampling at steady state. Thus, $a_j = a_{j,0}/r$, for $j = 1, \ldots, n$; cf. Section 2.3. We consider the Bayesian testing problem H_0: $\{\varepsilon_n(\bar{\boldsymbol{Y}}) \in [\varepsilon_0, \log n]\}$, against H_1: $\{\varepsilon_n(\bar{\boldsymbol{Y}}) \in [0, \varepsilon_0)\}$, for some $\varepsilon_0 \in (0, \log n)$. Then, $\rho_0 = \mathsf{P}[\varepsilon_n(\bar{\boldsymbol{Y}}) \geq \varepsilon_0]$ and $\rho_1 = \mathsf{P}[\varepsilon_n(\bar{\boldsymbol{Y}}) < \varepsilon_0]$ are the prior probabilities of H_0 and H_1, respectively. Their analog posteriors are $r_0(\boldsymbol{y}) = \mathsf{P}[\varepsilon_n(\bar{\boldsymbol{Y}}) \geq \varepsilon_0 | \boldsymbol{Y} = \boldsymbol{y}]$ and $r_1(\boldsymbol{y}) = \mathsf{P}[\varepsilon_n(\bar{\boldsymbol{Y}}) < \varepsilon_0 | \boldsymbol{Y} = \boldsymbol{y}]$, where $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$ and $\boldsymbol{y} = (y_1, \ldots, y_n) \in \ddot{\Delta}_k^{n-1}$. The Bayes factor of H_0 to H_1 is the posterior odds ratio $r_0(\boldsymbol{y})/r_1(\boldsymbol{y})$ over the prior odds ratio ρ_0/ρ_1, namely $\varphi(\boldsymbol{y}) = \rho_1 r_0(\boldsymbol{y})/\{\rho_0 r_1(\boldsymbol{y})\}$. The Monte Carlo solution consists in sampling of $(\bar{Y}_1, \ldots, \bar{Y}_n)$ from the prior in Equation (37) and then from the posterior (38), both levels by means of Equation (39). Thus, $r_0(\boldsymbol{y})$ and $r_1(\boldsymbol{y})$ are Bayesian bootstrap estimators of ρ_0 and ρ_1, respectively, and they allow for the evaluation of $\varphi(\boldsymbol{y})$. Alternatively, these values can be obtained without repeated sampling by using the conditional saddlepoint approximation of Section 3 with the D-G representation.

Example 11 (Tests based on spacings)**.** *The so-called spacings are the first order differences or gaps between successive values of the ordered sample. Let $U_1, \ldots, U_l$ be absolutely continuous and i.i.d. over $[0, 1]$, without loss of generality by the probability integral transform, let $0 \leq U_{(1)} \leq \ldots \leq U_{(l)} \leq 1$ denote the ordered sample and let $U_{(0)} = 0$ and $U_{(l+1)} = 1$. For $n = l + 1$, the spacings are defined by*

$$\bar{Y}_j = U_{(j)} - U_{(j-1)}, \quad \text{for } j = 1, \ldots, n. \tag{40}$$

Thus, $(\bar{Y}_1, \ldots, \bar{Y}_n)$ takes values in Δ_1^{n-1}. Statistics that are defined as functions of spacings are used in various statistical problems, goodness-of-fit testing representing the most important (see, e.g., [44]). Spacings are essential in the analysis of circular data, because they form a maximal invariant w.r.t. changes of null direction and sense of rotation. For Borel functions h_j, for $j = 1, \ldots, n$, important spacings statistics have the form

$$\sum_{j=1}^{n} h_j(n\bar{Y}_j). \tag{41}$$

If $h_1 = \ldots = h_n = h$, then the test statistic is called symmetric. Under the null hypothesis H_0 of uniformity of $U_1, \ldots, U_l$, the D-G representation holds with $a_1 = \ldots = a_n = 1$, so that the n spacings are equivalent in distribution to n i.i.d. exponential random variables conditioned by their sum. As Equation (41) takes the form of the M-statistic in Equation (1), the saddlepoint approximation of Section 3 can be directly applied.

The conditional saddlepoint approximation with the D-G representation under H_0 is analyzed numerically by [3] in the following cases: Rao's spacings test (viz., $h_j(x) = |x - 1|/2$, for $j = 1, \ldots, n$), the logarithm spacings test (viz., $h_j(x) = \log x$, for $j = 1, \ldots, n$), Greenwood's test (viz., $h_j(x) = x^2$, for $j = 1, \ldots, n$) and a locally most powerful spacings test (viz., $h_j(x) = \Phi^{(-1)}(j/(n+1))x$, for $j = 1, \ldots, n$). In the context of reliability, Gatto and Jammalamadaka [6] re-expressed a uniformly most powerful test of exponentially, against alternatives with increasing failure rate, in terms of spacings. They obtained the saddlepoint approximation and show some numerical comparisons.

These spacings can be generalized to higher order differences or gaps. Let $g \in \mathbb{N}^*$ denote the gap order, selected such that $n = (l + 1)/g \in \mathbb{N}^*$. The so-called multispacings are defined as

$$\bar{Y}_j = U_{(jg)} - U_{(\{j-1\}g)}, \text{ for } j = 1, \ldots, n. \tag{42}$$

As previously, $(\bar{Y}_1, \ldots, \bar{Y}_n)$ takes values in Δ_1^{n-1}. When $g = 1$, the random variables in Equation (42) coincide with the spacings in Equation (40). Under H_0, the D-G representation holds with $a_1 = \ldots = a_n = g$.

Gatto and Jammalamadaka [7] provided explicit formulae of the saddlepoint approximations for Rao's multispacings test and for the logarithmic multispacings test, together with a numerical study.

The next problem would be the computation of the distribution of a spacings or multispacings test statistic under a non-uniform alternative distribution. This can be done by saddlepoint approximation with the D-G representation whenever one can find the parameters $a_1, \ldots, a_n \in \mathbb{R}_+^*$ *such that, under the alternative distribution, the spacings or multispacings satisfy* $(\bar{Y}_1, \ldots, \bar{Y}_n) \sim \text{Dirichlet}(a_1, \ldots, a_n)$*. This would give the power of the test. However, re-expressing a non-uniform distribution in terms of a particular Dirichlet distribution does not appear practical, in general.*

5. Final Remarks

This article presents the saddlepoint approximation for M-statistics of dependent random variables taking values in a simplex. Four conditional representations that allow re-expressing these dependent random variables as independent ones are presented. A detailed presentation of the underlying urn sampling model that is common to all four conditional representations is given. Important applications are reviewed. New applications are presented with some numerical comparisons between this saddlepoint approximation and Monte Carlo simulation. The numerical accuracy of the saddlepoint approximation appears very good.

A practical question concerns the relative advantages and disadvantages of using the conditional saddlepoint approximation presented in this article. Indeed, tail probabilities can be computed rapidly and more easily by means of Monte Carlo simulation. However, there is no unique answer to this general question, because several aspects should be considered.

First, when very small tail probabilities, e.g., 10^{-4}, or extreme quantiles are desired, then the simple Monte Carlo used in this article may not always lead to accurate results. The reason is that the saddlepoint approximation is a large deviation technique, with bounded relative error everywhere in the tails, whereas simple Monte Carlo has unbounded relative error in the tails. In fact, simple Monte Carlo is even not logarithmic efficient. This is well explained in [45] (pp. 158–160). To have bounded relative error, importance sampling is required. Then, the mathematical complexity would become close to the one of the saddlepoint approximation. Moreover, computing quantiles by importance sampling may not be straightforward. As shown above, this is quite simple with the saddlepoint approximation.

The computations required for this article were done with *Matlab* (R2017b, The MathWorks, Natick, MA, USA). The minimization program *fminsearch* was used for obtaining the saddlepoint defined in Equation (13). All *Matlab* programs are available at `http://www.stat.unibe.ch`. They should be easily used and modified for new related applications.

One should also mention that, having analytical expression such as a saddlepoint approximation for computing a quantity of interest, may have advantages. Monte Carlo and other purely numerical methods often do not provide such an expression. For example, the saddlepoint approximation can be used for computing the sensitivity of the upper tail probability, viz. the derivative of the tail probability w.r.t. to a parameter of the model. Gatto and Peeters [46] proposed evaluating the sensitivityof the tail probability of the random sum w.r.t. the parameter of the summation index distribution (which is either Poisson or geometric) with the saddlepoint approximation. They showed numerically that the sensitivities obtained by the saddlepoint approximation and by simulation with importance sampling are very close, but this no longer true when simulation is without importance sampling. In the case of computing sensitivity, importance sampling is significantly more computationally intensive than the saddlepoint approximation.

An application of the saddlepoint approximation that exploits a different conditional representation concerns the distribution of the inhomogeneous compound Poisson total claim amount under force of interest, in the context of insurance. It was suggested by [47] and the main idea is the following. The inhomogeneous Poisson process of occurrence times of individual claims is given by $0 \leq T_1 \leq$

$T_2 \leq \ldots$. Let N_t denote the number of occurrences during the time interval $[0, t]$, for some $t > 0$. Then, $\forall n \in \mathbb{N}^*$,

$$\{(T_1, \ldots, T_{N_t}) | N_t = n\} \sim (Y_{(1)}, \ldots, Y_{(n)}), \tag{43}$$

where $Y_{(1)} \leq \ldots \leq Y_{(n)}$ are the ordered values of some random variables $Y_1, \ldots, Y_n$ that are nonnegative, i.i.d. and independent of $\{N_t\}_{t \geq 0}$. The individual claim amounts are represented by the random variables $X_1, X_2, \ldots$ that are nonnegative, i.i.d. and independent of $\{N_t\}_{t \geq 0}$. Let $r \in \mathbb{R}$ denote the force of interest. The discounted total claim amount is $Z_t = \sum_{j=0}^{N_t} \mathrm{e}^{r(t-T_j)} X_j$, for $T_0 = X_0 = 0$, and Equation (43) implies $Z_t \sim \sum_{j=0}^{N_t} \mathrm{e}^{r(t-Y_j)} X_j$, for $Y_0 = 0$. The last random sum has a simple structure and its distribution can be computed by the saddlepoint approximation of [18].

A technique that could exploit the four conditional representations of Section 2 for computing the conditional c.g.f. (and not the conditional saddlepoint approximation) can be found in [48]. It is tentatively applied, with the MP-NB representation, to the symmetric spacing-frequencies test statistic in Equation (33) in [41] (Section 6.3.2). However, this approach seems impractical.

Another extension of the proposed approximation would concern neutrosophic statistics. In standard statistics, observations and parameters are represented by precise values, whereas in neutrosophic statistics they remain indeterminate (see, e.g., [49]).

Funding: This research received no external funding.

Acknowledgments: The author is thankful to three anonymous reviewers, to Sreenivasa Rao Jammalamadaka and to Ilya Molchanov for various discussions, remarks and suggestions that improved the quality of this article.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Aitchison, J. *The Statistical Analysis of Compositional Data*; Chapman & Hall: London, UK, 1986.
2. Kotz, S.; Balakrishnan, N. Advances in urn models during the past two decades. In *Advances in Combinatorial Methods and Applications to Probability and Statistics*; Birkhäuser, Statistics for Industry and Technology: Boston, MA, USA, 1997; pp. 203–257.
3. Gatto, R.; Jammalamadaka, S.R. A conditional saddlepoint approximation for testing problems. *J. Am. Stat. Assoc.* **1999**, *94*, 533–541. [CrossRef]
4. Skovgaard, I.M. Saddlepoint expansions for conditional distributions. *J. Appl. Prob.* **1987**, *24*, 875–887. [CrossRef]
5. Gatto, R. Symbolic computation for approximating the distributions of some families of one and two-sample nonparametric test statistics. *Stat. Comput.* **2000**, *11*, 449–455.
6. Gatto, R.; Jammalamadaka, S.R. A saddlepoint approximation for testing exponentiality against some increasing failure rate alternatives. *Stat. Prob. Lett.* **2002**, *58*, 71–81. [CrossRef]
7. Gatto, R.; Jammalamadaka, S.R. Small sample asymptotics for higher order spacings. In *Advances in Distribution Theory, Order Statistics and Inference Part III: Order Statistics and Applications*; Birkhäuser, Statistics for Industry and Technology: Boston, MA, USA, 2006; pp. 239–252.
8. Butler, R.W. *Saddlepoint Approximations with Applications*; Cambridge University Press: Cambridge, UK, 2007.
9. Reid, N. The roles of conditioning in inference. *Stat. Sci.* **1995**, *10*, 138–157. [CrossRef]
10. Mirakhmedov, S.M.; Jammalamadaka, S. Rao, Ibrahim, B.M. On Edgeworth expansions in generalized urn models. *J. Theor. Prob.* **2014**, *27*, 725–753. [CrossRef]
11. Butler, R.W.; Sutton, R.K. Saddlepoint approximation for multivariate cumulative distribution functions and probability computations in sampling theory and outlier testing. *J. Am. Stat. Assoc.* **1998**, *93*, 596–604. [CrossRef]
12. Good, I.J. Saddlepoint methods for the multinomial distribution. *Ann. Math. Stat.* **1957**, *28*, 861–881. [CrossRef]
13. Klugman, S.A.; Panjer, H.H.; Willmot, G.E. *Loss Models: From Data to Decisions*, 3rd ed.; Wiley & Sons: New York, NY, USA, 2008.

14. Ivchenko, G.I.; Ivanov, A.V. Decomposable statistics in inverse urn problems. *Discr. Math. Appl.* **1995**, *5*, 159–172. [CrossRef]
15. Copson, E.T. *Asymptotic Expansions*; Cambridge University Press: Cambridge, UK, 1965.
16. De Bruijn, N.G. *Asymptotic Methods in Analysis*; Dover Publications: New York, NY, USA, 1981.
17. Daniels, H.E. Saddlepoint approximations in statistics. *Ann. Math. Stat.* **1954**, *25*, 631–650. [CrossRef]
18. Lugannani, R.; Rice, S. Saddle point approximation for the distribution of the sum of independent random variables. *Adv. Appl. Prob.* **1980**, *12*, 475–490. [CrossRef]
19. Daniels, H.E. Tail probability approximations. *Int. Stat. Rev.* **1987**, *55*, 37–48. [CrossRef]
20. Wang, S. Saddlepoint approximations in conditional inference. *J. Appl. Prob.* **1993**, *30*, 397–404. [CrossRef]
21. Jing, B.; Robinson, J. Saddlepoint Approximations for Marginal and Conditional Probabilities of Transformed Variables. *Ann. Stat.* **1994**, *22*, 1115–1132. [CrossRef]
22. Kolassa, J.E. Higher-order approximations to conditional distribution functions. *Ann. Stat.* **1996**, *24*, 353–365. [CrossRef]
23. DiCiccio, T.J.; Martin, M.A.; Young, G.A. Analytical approximations to conditional distribution functions. *Biometrika* **1993**, *80*, 781–790. [CrossRef]
24. Field, C.A.; Tingley, M.A. Small sample asymptotics: Applications in robustness. In *Handbook of Statistics*; North-Holland: Amsterdam, The Netherlands, 1997; Volume 15, pp. 513–536.
25. Gatto, R. Saddlepoint approximations. In *StatsRef: Statistics Reference Online*; Wiley & Sons: New York, NY, USA, 2015; pp. 1–7.
26. Goutis, C.; Casella, G. Explaining the saddlepoint approximation. *Am. Stat.* **1999**, *53*, 216–224.
27. Reid, N. Saddlepoint methods and statistical inference. *Stat. Sci.* **1988**, *3*, 213–238. [CrossRef]
28. Field, C.A.; Ronchetti, E. *Small Sample Asymptotics*; Institute of Mathematical Statistics Lecture Notes-Monograph Series: Hayward, CA, USA, 1990.
29. Jensen, J.L. *Saddlepoint Approximations*; Oxford University Press: Oxford, UK, 1995.
30. Kolassa, J.E. *Series Approximation Methods in Statistics*, 3rd ed.; Springer Lecture Notes in Statistics; Springer: New York, NY, USA, 2006.
31. Wang, S. One-step saddlepoint approximations for quantiles. *Comput. Stat. Data Anal.* **1995**, *20*, 65–74. [CrossRef]
32. Shannon, C.E. The mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423, 623–656. [CrossRef]
33. Khinchin, A.I. *Mathematical Foundations of Information Theory*; English Translation of Two Original Articles in Russian; Dover Publications: New York, NY, USA, 1957.
34. Davison, A.C.; Hinkley, D.V. Saddlepoint approximations in resampling methods. *Biometrika* **1988**, *75*, 417–431. [CrossRef]
35. Feuerverger, A. On the empirical saddlepoint approximation. *Biometrika* **1989**, *76*, 457–464. [CrossRef]
36. Wang, S. Saddlepoint approximations in resampling analysis. *Ann. Inst. Stat. Math.* **1990**, *42*, 115–131. [CrossRef]
37. Ronchetti, E.; Welsh, A.H. Empirical saddlepoint approximations for multivariate M-estimators. *J. R. Stat. Soc. B* **1994**, *56*, 313–326. [CrossRef]
38. Davison, A.C.; Hinkley, D.V. *Bootstrap Methods and Their Application*; Cambridge University Press: Cambridge, UK, 1997.
39. Abd-Elfattah, E.; Butler, R. Saddlepoint approximations for rank-invariant permutation tests and confidence intervals with interval-censoring. *Can. J. Stat.* **2014**, *42*, 308–324. [CrossRef]
40. Booth, J.G.; Butler, R.W. Randomization distributions and saddlepoint approximations in generalized linear models. *Biometrika* **1990**, *77*, 787–796. [CrossRef]
41. Gatto, R.; Jammalamadaka, S.R. On two-sample tests for circular data based on spacing-frequencies. In *Geometry Driven Statistics*; Wiley & Sons: New York, NY, USA, 2015; pp. 129–145.
42. Holst, L.; Rao, J.S. Asymptotic theory for some families of two-sample nonparametric statistics. *Sankhyā Ser. A* **1980**, *42*, 19–52.
43. Rubin, D.B. The Bayesian bootstrap. *Ann. Stat.* **1981**, *9*, 130–134. [CrossRef]
44. Pyke, R. Spacings. *J. R. Stat. Soc. B* **1965**, *27*, 395–449. [CrossRef]
45. Asmussen, S.; Glynn, P.W. *Stochastic Simulation. Algorithms and Analysis*; Springer: New York, NY, USA, 2007.

46. Gatto, R.; Peeters, C. Saddlepoint approximations to sensitivities of tail probabilities of random sums and comparisons with Monte Carlo estimators. *J. Stat. Comput. Simul.* **2015**, *85*, 641–659. [CrossRef]
47. Gatto, R. A saddlepoint approximation to the distribution of inhomogeneous discounted compound Poisson processes. *Methodol. Comput. Appl. Prob.* **2010**, *12*, 533–551. [CrossRef]
48. Bartlett, M.S. The characteristic function of a conditional statistic. *J. Lond. Math. Soc.* **1938**, *13*, 62–67. [CrossRef]
49. Aslam, M. Design of sampling plan for exponential distribution under neutrosophic statistical interval method. *IEEE Access* **2018**, *6*, 64153–64158. [CrossRef]

© 2019 by the author. 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 (http://creativecommons.org/licenses/by/4.0/).