

Article

# Alternative Dirichlet Priors for Estimating Entropy via a Power Sum Functional

Tanita Botha <sup>1,\*</sup>, Johannes Ferreira <sup>1,2,†</sup> and Andriette Bekker <sup>1,2,†</sup>

<sup>1</sup> Department of Statistics, Faculty of Natural and Agricultural Sciences, University of Pretoria, Pretoria 0028, South Africa; johan.ferreira@up.ac.za (J.F.); andriette.bekker@up.ac.za (A.B.)

<sup>2</sup> Centre of Excellence in Mathematical and Statistical Science, University of Witwatersrand, Johannesburg 2050, South Africa

\* Correspondence: Tanita.Botha@up.ac.za

† These authors contributed equally to this work.

**Abstract:** Entropy is a functional of probability and is a measurement of information contained in a system; however, the practical problem of estimating entropy in applied settings remains a challenging and relevant problem. The Dirichlet prior is a popular choice in the Bayesian framework for estimation of entropy when considering a multinomial likelihood. In this work, previously unconsidered Dirichlet type priors are introduced and studied. These priors include a class of Dirichlet generators as well as a noncentral Dirichlet construction, and in both cases includes the usual Dirichlet as a special case. These considerations allow for flexible behaviour and can account for negative and positive correlation. Resultant estimators for a particular functional, the power sum, under these priors and assuming squared error loss, are derived and represented in terms of the product moments of the posterior. This representation facilitates closed-form estimators for the Tsallis entropy, and thus expedite computations of this generalised Shannon form. Select cases of these proposed priors are considered to investigate the impact and effect on the estimation of Tsallis entropy subject to different parameter scenarios.

**Keywords:** generator; multinomial; noncentral; Poisson; power sum; Tsallis



**Citation:** Botha, T.; Ferreira, J.; Bekker, A. Alternative Dirichlet Priors for Estimating Entropy via a Power Sum Functional. *Mathematics* **2021**, *9*, 1493. <https://doi.org/10.3390/math9131493>

Academic Editor: Manuel Alberto M. Ferreira

Received: 29 May 2021  
Accepted: 23 June 2021  
Published: 25 June 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Shannon entropy and related information measures are functionals of probability and a measurement of information contained in a system that arise in information theory, machine learning and text modelling, amongst others. Ref. [1] discussed quantifying the information carried by neural signals to estimating the dependency structure and inferring causal relations, uncertainty and dispersion in statistics being applied in fields such as molecular biology. Other interests include studies measuring complexity of dynamics in physics, to studies measuring diversity in ecology and genetics, fields of coding theory and cryptography [2], financial analysis and data compression [3]. Numerous inferential tasks rely on data-driven procedures to estimate these quantities. In these settings and utilising the estimated quantities, researchers are often confronted with data arising from an unknown discrete distribution, and seek to estimate its entropy. This motivates sustained research interest within entropy, coupled with the current data-driven and computing-rich era, for practitioners.

Entropy estimation remains an openly discussed challenge. Ref. [4] investigated how the maximum likelihood estimator (MLE) performed. This is also referred to as the plug-in principle in functional estimation, where a point estimate of the parameter is used to build an estimate for a functional of the parameter. The classical asymptotic theory of MLEs does not adequately address high-dimensional settings in this current data-driven era [4]. High-dimensional statistics arguably demand theoretical tools to address the needs of these high-dimensional settings. Ref. [5] investigated 18 different estimation measures and

the suitability was determined experimentally based on the bias and the mean squared error. This work takes a Bayesian approach to entropy estimation, building upon work by [1,4,6,7].

Multivariate count data constrained to add up to a certain constant are commonly modelled using the multinomial distribution. This is widely used in modelling categorical data, of which features could be for example, words in the case of textual documents or visual words in the case of images. The Dirichlet distribution, closely related to the probabilistic behaviour of the multinomial distribution, is a conjugate prior for the multinomial distribution when a Bayes perspective is of interest. Ref. [8] highlights how the use of prior distributions in a Bayesian framework makes it possible to work with very limited data sets. Ref. [9] underscores the superior performance by using the hierarchical approach which introduces the construction of the statistical model. Some meaningful studies include [1,4,6,10]. Ref. [11] also showed how using different Dirichlet distributions for the bivariate case gives one the opportunity to include prior information and expert opinion to obtain more realistic results in certain situations. Ref. [4] also experimented with the estimation of entropy and this triggered more exploration with alternative priors. Experimentation on diverse data sets might necessitate parameter-rich priors; therefore, this study proposes these alternative Dirichlet priors to address this potential challenge.

The paper illustrates how a Bayesian approach is applied in a multinomial-Dirichlet family setup, which allows us to obtain a posterior distribution from where explicit expressions for the Tsallis entropy can be derived, by particularly focussing on the Product moment for the power sum functional, and assuming squared error loss. The first of two main contributions of this paper is the addition of flexible priors from a Dirichlet family, utilising them within an information-theoretical world, which also allows for positive correlation in addition to the usual negative correlation characteristic. The second shows that using elegant constructs of the complete product moments of the posteriors, gives one the comparative advantage of obtaining explicit estimators for entropy under these Dirichlet priors. Ref. [8] echos how the computation on moments accelerates the estimation of entropy.

The paper is outlined as follows. In Section 2, the essential components that are used in the paper are outlined. In Section 3, alternative Dirichlet priors will be introduced and studied, as candidates for the Bayesian analysis of entropy. In Section 4, analytical expressions for the entropy expressions under consideration will be derived and studied. Section 5 contains conclusions and final thoughts.

## 2. Essential Components

The countably discrete model under consideration in this paper is given by the well-motivated multinomial distribution. A discrete random variable  $\mathbf{X} = (X_1, \dots, X_K)$  follows the multinomial distribution of order  $K$  (i.e., with  $K$  distinct classes of interest) with parameters  $\mathbf{p} = (p_1, p_2, \dots, p_K)$  and  $n > 0$  if its probability mass function (pmf) is given by

$$f(\mathbf{x}|\mathbf{p}) = \frac{n!}{\prod_{i=1}^K x_i! (n - \sum_{i=1}^K x_i)} \prod_{i=1}^K p_i^{x_i} (1 - \sum_{i=1}^K p_i)^{n - \sum_{i=1}^K x_i}. \quad (1)$$

The Dirichlet distribution (of type 1, see [12]) of order  $K \geq 2$  and parameters  $\mathbf{\Pi} = (\pi_1, \pi_2, \dots, \pi_{K+1})$  for  $\pi_i > 0, i = 1, \dots, k + 1$ , has a probability density function (pdf) with respect to the Lebesgue measure on the Euclidean space  $\mathbb{R}^K$  given by

$$h(\mathbf{p}; \mathbf{\Pi}) = \frac{\prod_{i=1}^{k+1} \Gamma(\pi_i)}{\Gamma(\sum_{i=1}^{k+1} \pi_i)} \prod_{i=1}^{k+1} p_i^{\pi_i - 1} \quad (2)$$

on the  $K$  dimensional simplex, defined by

$$\begin{aligned}
 p_1, p_2, \dots, p_K &> 0 \\
 p_1 + p_2 + \dots + p_K &< 1 \\
 p_{K+1} &= 1 - p_1 - \dots - p_K,
 \end{aligned}$$

and where  $\Gamma(\cdot)$  denotes the usual gamma function (the space and constraints of this  $K$  dimensional simplex is denoted by  $\mathcal{A}$ ).

To derive a Bayesian engine, we need the likelihood function  $f(\mathbf{x}|\mathbf{p})$  in addition to a suitable prior distribution  $h(\mathbf{p})$ . The fundamental relationship between the likelihood function and the prior distribution to form the posterior distribution  $f(\mathbf{p}|\mathbf{x})$  is given by

$$f(\mathbf{p}|\mathbf{x}) = \frac{f(\mathbf{x}|\mathbf{p})h(\mathbf{p})}{\int f(\mathbf{x}|\mathbf{p})h(\mathbf{p})d\mathbf{p}}.$$

The most popular form of entropy is that of Shannon:

$$H(P) = \sum_{i=1}^{K+1} -p_i \ln p_i.$$

Various generalised cases of this entropy exist, which relies on the power sum:

$$F_\alpha(P) = \sum_{i=1}^{K+1} p_i^\alpha \tag{3}$$

where  $\alpha > 0$ . The power sum functional occurs in various operational problems ([4]). Under the assumption of squared error loss within Bayes estimation, the estimates of both these quantities is given by their expected values:

$$E(H(P)) = E\left(\sum_{i=1}^{K+1} -p_i \ln p_i\right)$$

and

$$\hat{F}_\alpha(P) = E(F_\alpha(P)) = E\left(\sum_{i=1}^{K+1} p_i^\alpha\right) = \sum_{i=1}^{K+1} E(p_i^\alpha). \tag{4}$$

Thus, it is of value to consider the expected value of  $p_i^\alpha$  for all values of  $i$ .

Since there are cases, such as the non-extensive system like alignment processing, namely registration, which has complex behaviours associated with the phenomena of radar-imaging systems [13])which cannot be fully explained by Shannon entropy, other generalized forms were designed. The Tsallis entropy considered in this paper, which is a popular generalised entropy, tends to Shannon entropy as  $\alpha$  tends to 1 [14] and is given by

$$T = \frac{\sum_{i=1}^{K+1} p_i^\alpha - 1}{1 - \alpha}; \alpha \geq 0, \alpha \neq 1. \tag{5}$$

The estimate of this generalisation can be written in terms of the estimate of the power sum:

$$E(T) = E\left(\frac{\sum_{j=1}^{K+1} p_j^\alpha - 1}{1 - \alpha}\right) = \frac{\hat{F}_\alpha(p) - 1}{1 - \alpha}.$$

Since the power sum is easier to estimate than the Shannon entropy, the power sum is used in our case. We consider the estimate as the expectation under the posterior distribution, thus under squared-error loss.

### 3. Alternative Dirichlet Priors

In this section, two previously unconsidered Dirichlet priors, namely the Dirichlet generator prior and the noncentral Dirichlet prior will be proposed. Positive correlation

can be observed for special cases of the Dirichlet generator prior, which is a benefit of this generator form. These new contributions add to the field of generative models for count data and have not been previously considered for entropy.

### 3.1. Dirichlet Generator Prior

In this section, Dirichlet generator distributions are proposed as alternative candidates. From this form, numerous flexible candidates can be “generated”.

**Definition 1.** Suppose  $\mathbf{p}$  is Dirichlet-generator distributed. Then, its pdf is given by

$$h(p_1, \dots, p_K; \mathbf{\Pi}) = C p_1^{\pi_1-1} p_2^{\pi_2-1} \dots p_K^{\pi_K-1} \left(1 - \sum_{i=1}^K p_i\right)^{\pi_{K+1}-1} g\left(\theta \sum_{i=1}^K p_i\right) \tag{6}$$

with  $C$  a normalising constant such that

$$C^{-1} = \int_{\mathcal{A}} \dots \int p_1^{\pi_1-1} p_2^{\pi_2-1} \dots p_K^{\pi_K-1} \left(1 - \sum_{i=1}^K p_i\right)^{\pi_{K+1}-1} g\left(\theta \sum_{i=1}^K p_i\right) d\mathbf{p}. \tag{7}$$

The vector  $\mathbf{p} \in \mathcal{A}$  is thus a Dirichlet generator variate with parameters  $\mathbf{\Pi} = (\pi_1, \dots, \pi_{K+1})$ ,  $\theta \in \mathbb{R}$ , and whichever additional parameters  $g(\cdot)$  imposed, which ensures that the pdf  $h(\cdot)$  is non-negative. The following conditions also apply:

1.  $g(\cdot)$  is a Borel-measurable function;
2.  $g(\cdot)$  admits a Taylor series expansion;
3.  $g(0) = 1$ .

The usual Dirichlet distribution with pdf (2) is thus a special case of (6) when  $\theta = 0$ .

For illustration of the implementation of the Dirichlet generator prior, we focus on

$$\begin{aligned} g\left(\theta \sum_{i=1}^K p_i\right) &= {}_rF_q(a_1, \dots, a_r; b_1, \dots, b_q; \theta \sum_{i=1}^K p_i) \\ &= \sum_{n=0}^{\infty} \frac{(a_1)_n \dots (a_r)_n}{(b_1)_n \dots (b_q)_n} \frac{(\theta \sum_{i=1}^K p_i)^n}{n!} \end{aligned}$$

where  ${}_pF_q(\cdot)$  denotes the generalised hypergeometric function (see [15]) and  $(a)_k = \frac{\Gamma(a+k)}{\Gamma(a)}$  is the Pochhammer function.

The prior distribution (6) will then take on the following form with pdf

$$\begin{aligned} h(p_1, \dots, p_k; \mathbf{\Pi}) &= C_*^{-1} \frac{\Gamma\left(\sum_{j=1}^{k+1} \pi_j\right)}{\prod_{j=1}^{k+1} \Gamma(\pi_j)} p_1^{\pi_1-1} \dots p_k^{\pi_k-1} \left(1 - \sum_{j=1}^k p_j\right)^{\pi_{k+1}-1} \\ &\quad \times {}_rF_q\left(a_1, \dots, a_r, b_1, \dots, b_q; \theta \sum_{j=1}^k p_j\right) \end{aligned} \tag{8}$$

where  $C_*$  is equal to

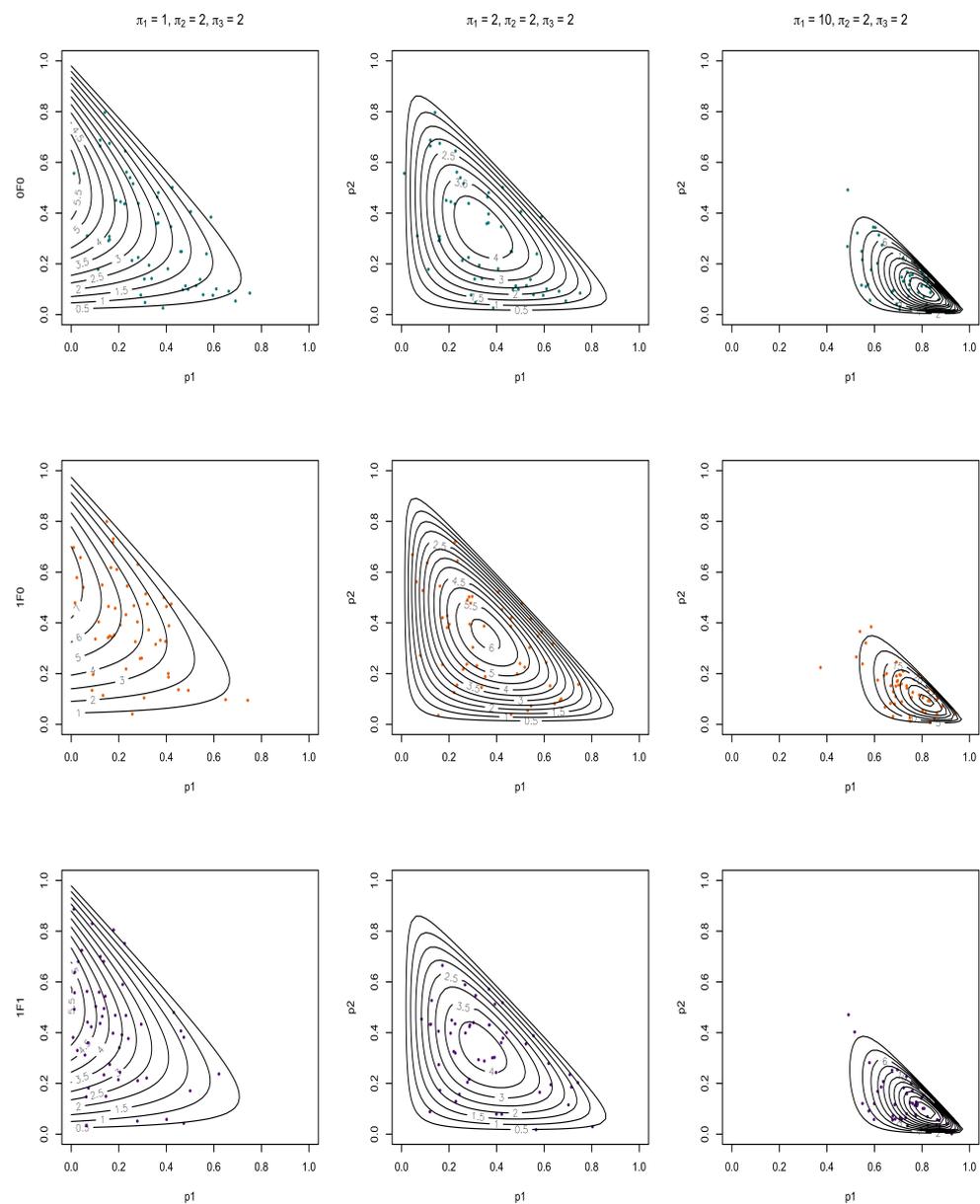
$${}_{r+1}F_{q+1}\left(a_1, \dots, a_r, \sum_{j=1}^k \pi_j; b_1, \dots, b_q, \sum_{j=1}^{k+1} \pi_j; \theta\right). \tag{9}$$

In this paper, three hypergeometric functions are considered ( ${}_0F_0$ ;  ${}_0F_1$  and  ${}_1F_1$ ), since these are commonly considered functions representing exponential, binomial, and the confluent hypergeometric functions. For illustrative investigation, bivariate observations from the corresponding distributions were simulated using Algorithm 1 and the associated

pdfs are overlaid and presented in Figures 1–3. The data were simulated from (8) using the following steps of the Acceptance/Rejection method:

**Algorithm 1** Acceptance/Rejection method

1. Define  $y_i \in (0, 1)$  of size  $n$  for  $i = 1, 2$ ;
2. Calculate the pdf of the Dirichlet (2)  $h(y_1, y_2)$  for  $y_1 + y_2 < 1$ ;
3. Obtain  $m = \max(h(y_1, y_2))$ ;
4. Simulate  $p_i \sim Unif(0, 1)$  of size  $n$  for  $i = 1, 2$ ;
5. Calculate the pdf of the Dirichlet generator (8)  $h(p_1, p_2)$  for  $p_1 + p_2 < 1$ ;
6. Simulate  $z \sim Unif(0, 1)$  of size  $n$ ;
7. If  $\frac{h(p_1, p_2)}{m} > z$ , then keep  $(p_1, p_2)$ , else return to Step 4;
8. Repeat steps 4–7  $k$  times.



**Figure 1.** Dirichlet generator priors (8) for  $\theta = 0.1$  with three different sets of  $\Pi$  (described above Figure 1),  $n = 50$ .

Figures 1 and 2 illustrate the three chosen hypergeometric functions for two choices of  $\theta$  and for three different sets of  $\Pi$ s if  $K = 2$  with  $a_1 = 4$  and  $b_1 = 5$  for  ${}_0F_1$  and  ${}_1F_1$ , respectively. This firstly illustrates the difference between the different hypergeometric candidates as well as the effect a change in  $\pi_1$  has on these three functions (note that a symmetric observation would be made for  $\pi_2$ ). The difference between Figures 1 and 2 shows the effect that  $\theta$  has on these different combinations with Figure 1 having a very small (almost negligible)  $\theta$ , while Figure 2 increases the value of  $\theta$ . An increase in  $\pi_1$  results in a more highly dense concentration of the pdf for corresponding values of  $p_1$  and  $p_2$ . This is observed for all three considered hypergeometric candidates, as seen in Figures 1 and 2, also for an increase in  $\theta$ . For Figure 3, a single set of  $\Pi$ s were selected with  $\theta = 0.1$  to showcase the effect that the parameters  $a$  and  $b$  of the hypergeometric function have on the  ${}_1F_0$  and  ${}_1F_1$  functions. As  $a$  increases, an increased mass is observed closer to the restriction  $p_1 + p_2 < 1$  while an increase in  $b$  results in a lower pdf volume.

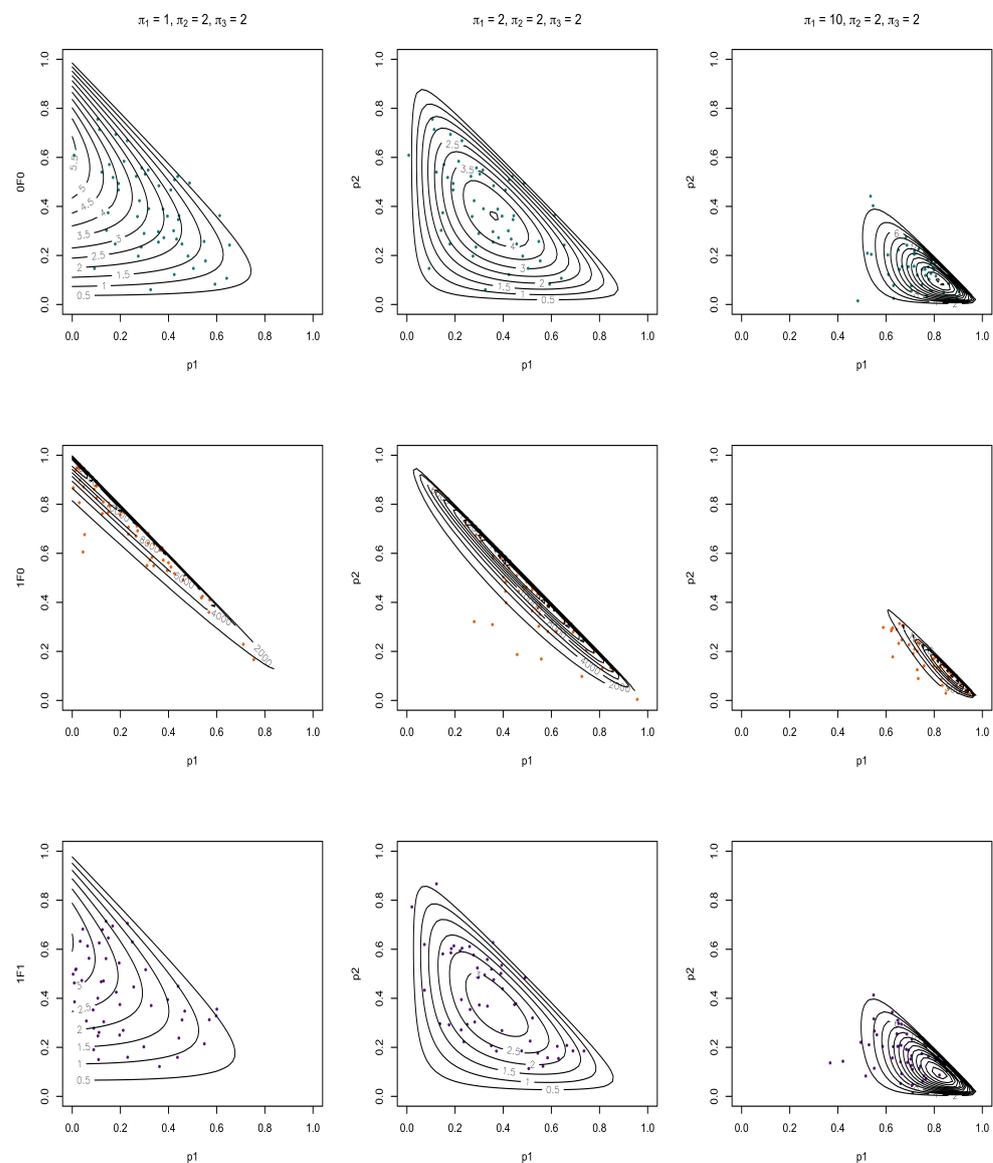


Figure 2. Dirichlet generator priors (8) for  $\theta = 0.9$  with three different sets of  $\Pi$ ,  $n = 50$ .

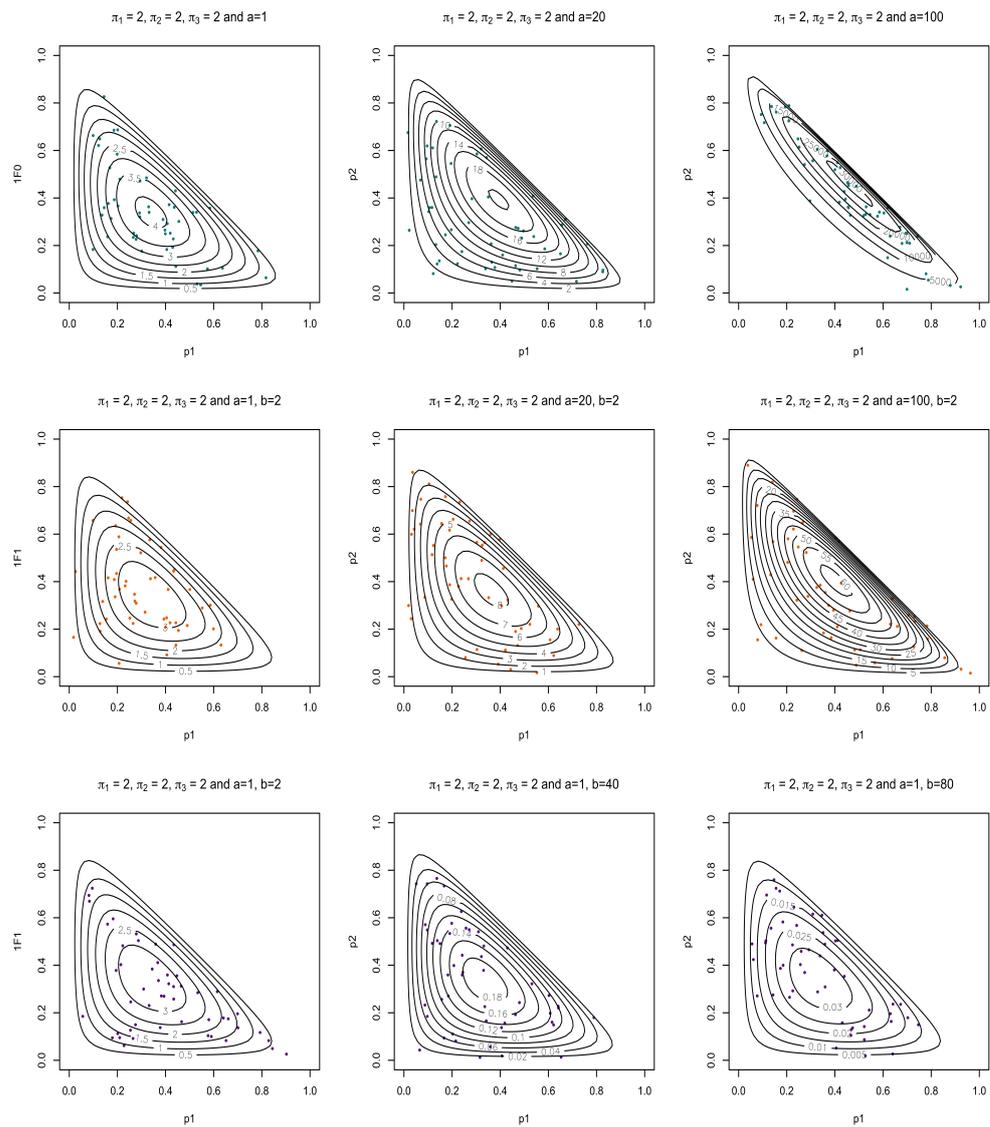


Figure 3. Dirichlet generator priors (8) for  $\theta = 0.1$  with a single set  $\Pi$ ,  $n = 50$ .

Next, the posterior distribution is derived, assuming the Dirichlet generator prior (8) together with a multinomial likelihood (1).

**Theorem 1.** Suppose the likelihood function is given by (1) and the prior distribution for  $\mathbf{p}$  is given by (8). Then, the pdf of the posterior distribution is given by

$$f(\mathbf{p}|\mathbf{x}) \propto p_1^{\pi_1+x_1-1} p_2^{\pi_2+x_2-1} \dots p_K^{\pi_K+x_K-1} (1 - \sum_{i=1}^K p_i)^{\pi_{K+1}+x_{K+1}-1} g(\theta \sum_{i=1}^K p_i) \quad (10)$$

which is identifiable as a Dirichlet generator distribution with parameters  $(\pi_1 + x_1, \dots, \pi_K + x_K, \pi_{K+1} + x_{K+1})$ .

The complete product moment of the Dirichlet generator posterior (10) is of interest for the power sum (4), thus we are interested in  $E(p_1^{k_1} p_2^{k_2} \dots p_k^{k_k} p_{k+1}^{k_{k+1}})$ .

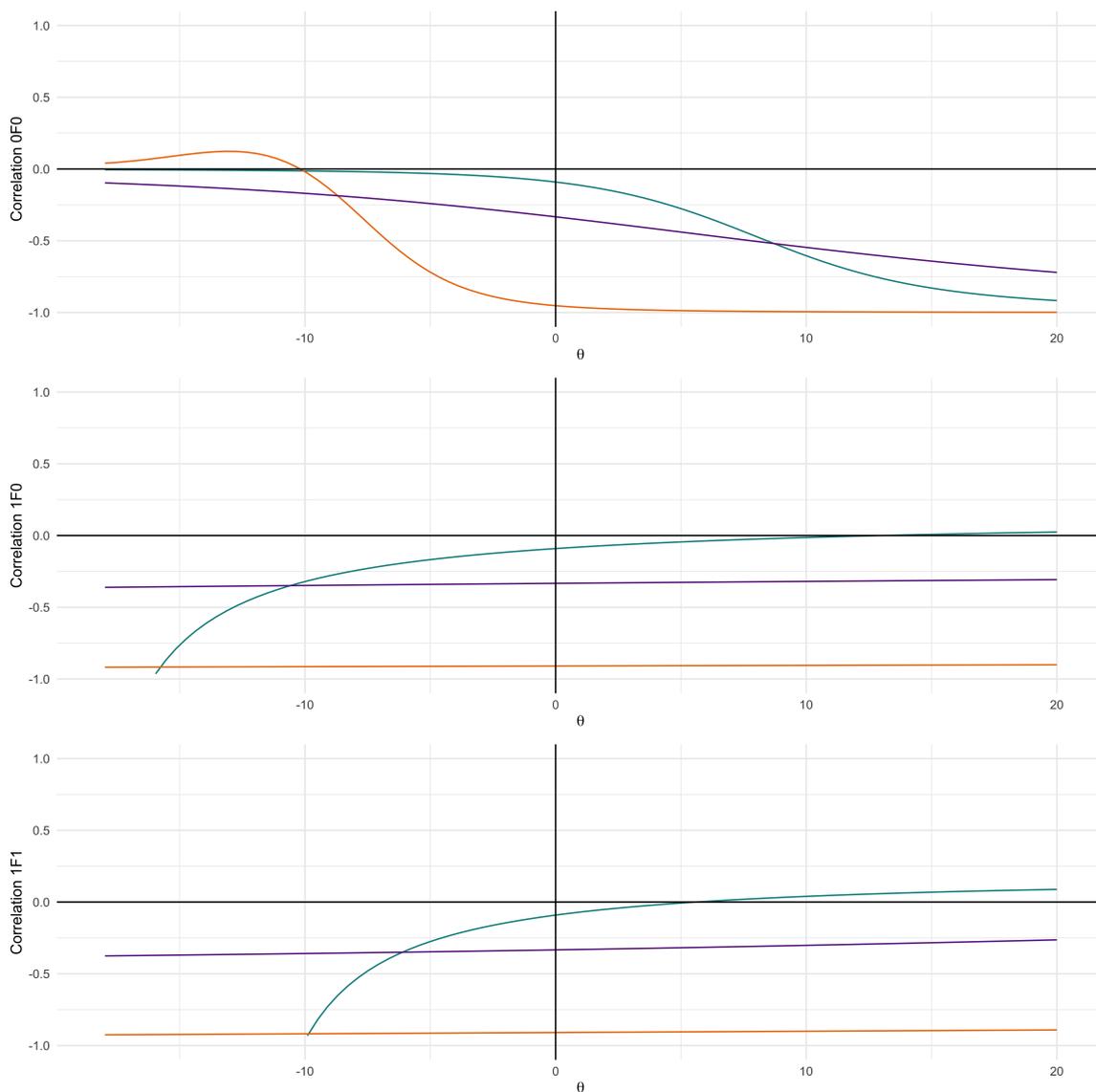
**Theorem 2.** Suppose that  $\mathbf{p}|\mathbf{x}$  follows a Dirichlet generator posterior distribution with pdf given in (10). Then, the complete product moment is given by

$$\begin{aligned}
 & E(p_1^{k_1} p_2^{k_2} \dots p_k^{k_k} p_{k+1}^{k_{k+1}}) \\
 &= \frac{(\pi_1 + x_1)_{k_1} (\pi_2 + x_2)_{k_2} \dots (\pi_{k+1} + x_{k+1})_{k_{k+1}}}{(\pi_1 + x_1 + \dots + \pi_{k+1} + x_{k+1})_{k_1 + \dots + k_{k+1}}} \\
 &\times \frac{{}_{r+1}F_{q+1}(a_1, \dots, a_r, \sum_{i=1}^k (\pi_i + x_i + k_i); b_1, \dots, b_q, \sum_{i=1}^{k+1} (\pi_i + x_i + k_i); \theta)}{{}_{r+1}F_{q+1}(a_1, \dots, a_r, \sum_{i=1}^k (\pi_i + x_i); b_1, \dots, b_q, \sum_{i=1}^{k+1} (\pi_i + x_i); \theta)}. \quad (11)
 \end{aligned}$$

Special cases of the above expression include setting  $k_{k+1} = 0$  to obtain an expression for the usual product moment of the Dirichlet generator distribution under investigation in this paper.

**Proof.** See Appendix A for the proof. □

The product moment can then be used to investigate the correlation for the examples as illustrated in this section. Figure 4 displays the correlation for a range of  $\theta$  values using (8) and the special cases. It is important to note the positive correlation obtained by the introduction of  $g(\cdot) = {}_rF_q(\cdot)$ , which is a major benefit of using these alternative Dirichlet priors.



**Figure 4.** Correlation for different  ${}_pF_q$  candidates. Set 1—blue ( $\pi_1 = 0.5; \pi_2 = 0.5; \pi_3 = 5$ ); Set 2—orange ( $\pi_1 = 2; \pi_2 = 2; \pi_3 = 0.1$ ); Set 3—purple ( $\pi_1 = 5; \pi_2 = 5; \pi_3 = 10$ ) with  $a = 100$  and  $b = 2$ .

### 3.2. Noncentral Dirichlet Prior

In this section, a noncentral Dirichlet distribution will be constructed via the use of Poisson weights. Ref. [16] explored the use of a compounding method as a distributional building tool to obtain bivariate noncentral distributions and showed how this form of the distribution isolated the noncentrality parameter by retaining them in a Poisson probability form and hence introducing mathematical convenience. Ref. [17] extended on this work by introducing new bivariate gamma distributions emanating from a scale mixture of normal class.

**Theorem 3.** Suppose  $\mathbf{p}$  is Dirichlet distributed with pdf given by (2). Then, a noncentral Dirichlet distribution can be constructed in the following manner:

$$\begin{aligned}
 h(\mathbf{p}; \mathbf{\Pi}, \mathbf{\Lambda}) & \tag{12} \\
 &= \sum_{j_1=1}^{\infty} \dots \sum_{j_{K+1}=1}^{\infty} \frac{\exp(\frac{\lambda_1}{2})(\frac{\lambda_1}{2})^{j_1}}{j_1!} \dots \frac{\exp(\frac{\lambda_K}{2})(\frac{\lambda_K}{2})^{j_K}}{j_K!} \frac{\exp(\frac{\lambda_{K+1}}{2})(\frac{\lambda_{K+1}}{2})^{j_{K+1}}}{j_{K+1}!} h(\mathbf{p}; \mathbf{\Pi} | j_1, \dots, j_{K+1}) \\
 &= \sum_{j_1=1}^{\infty} \dots \sum_{j_{K+1}=1}^{\infty} \frac{\exp(\frac{\lambda_1}{2})(\frac{\lambda_1}{2})^{j_1}}{j_1!} \dots \frac{\exp(\frac{\lambda_K}{2})(\frac{\lambda_K}{2})^{j_K}}{j_K!} \frac{\exp(\frac{\lambda_{K+1}}{2})(\frac{\lambda_{K+1}}{2})^{j_{K+1}}}{j_{K+1}!} \\
 &\times \frac{\Gamma(\pi_1 + j_1 + \dots + \pi_K + j_K + \pi_{K+1} + j_{K+1})}{\Gamma(\pi_1 + j_1)\Gamma(\pi_K + j_K)\Gamma(\pi_{K+1} + j_{K+1})} p_1^{\pi_1 + j_1 - 1} \dots p_K^{\pi_K + j_K - 1} (1 - \sum_{i=1}^K p_i)^{\pi_{K+1} + j_{K+1} - 1}
 \end{aligned}$$

where  $h(\mathbf{p} | j_1, \dots, j_{K+1})$  denotes the conditional (central) Dirichlet distribution (see (2)) with parameters  $\mathbf{\Pi}_* = (\pi_1 + j_1, \dots, \pi_{K+1} + j_{K+1})$ , and  $\mathbf{\Lambda}$  denotes the vector of noncentral parameters  $(\lambda_1, \dots, \lambda_K, \lambda_{K+1})$  with  $\lambda_i > 0 \forall i$ . After simplification (12) reflects

$$\begin{aligned}
 h(\mathbf{p}; \mathbf{\Pi}, \mathbf{\Lambda}) & \\
 &= h(\mathbf{p}; \mathbf{\Pi}) \exp(-\sum_{i=1}^{K+1} \frac{\lambda_i}{2}) \\
 &\times \sum_{\phi} \frac{(\pi_1 + \dots + \pi_{K+1})_{j_1 + \dots + j_{K+1}}}{(\pi_1)_{j_1} \dots (\pi_{K+1})_{j_{K+1}} j_1! \dots j_{K+1}!} \left(\frac{\lambda_1}{2} p_1\right)^{j_1} \dots \left(\frac{\lambda_K}{2} p_K\right)^{j_K} \left(\frac{\lambda_{K+1}}{2} (1 - \sum_{i=1}^K p_i)\right)^{j_{K+1}} \tag{13}
 \end{aligned}$$

where  $h(\mathbf{p}; \mathbf{\Pi})$  denotes the (unconditional) Dirichlet distribution (see (2)) with parameter  $\mathbf{\Pi}$  and where  $\sum_{\phi} = \sum_{j_1=1}^{\infty} \dots \sum_{j_{K+1}=1}^{\infty}$ .

**Remark 1.** The pdf in equation (13) reflects a parametrization of the noncentral Dirichlet distribution of [12] and can be represented via the confluent hypergeometric function of several variables:

$$\begin{aligned}
 f(\mathbf{p}; \mathbf{\Lambda}) & \\
 &= h(\mathbf{p}; \mathbf{\Pi}) \exp(-\sum_{i=1}^{K+1} \frac{\lambda_i}{2}) \\
 &\times \Psi_2^{(K+1)}\left(\sum_{i=1}^{K+1} \pi_i; \pi_1, \dots, \pi_{K+1}; \frac{\lambda_1}{2} p_1, \dots, \frac{\lambda_K}{2} p_K, \frac{\lambda_{K+1}}{2} (1 - \sum_{i=1}^K p_i)\right)
 \end{aligned}$$

where

$$\begin{aligned}
 &\Psi_2^{(K+1)}\left(\sum_{i=1}^{K+1} \pi_i; \pi_1, \dots, \pi_{K+1}; \frac{\lambda_1}{2} p_1, \dots, \frac{\lambda_K}{2} p_K, \frac{\lambda_{K+1}}{2} (1 - \sum_{i=1}^K p_i)\right) \\
 &= \sum_{\phi} \frac{(\pi_1 + \dots + \pi_{K+1})_{j_1 + \dots + j_{K+1}}}{(\pi_1)_{j_1} \dots (\pi_{K+1})_{j_{K+1}} j_1! \dots j_{K+1}!} \left(\frac{\lambda_1}{2} p_1\right)^{j_1} \dots \left(\frac{\lambda_K}{2} p_K\right)^{j_K} \left(\frac{\lambda_{K+1}}{2} (1 - \sum_{i=1}^K p_i)\right)^{j_{K+1}}.
 \end{aligned}$$

In particular, when  $\lambda_1 = \lambda_2 = \dots = \lambda_K = \lambda_{K+1} = 0$ , see that

$$h(\mathbf{p}; \mathbf{\Pi}) \exp(0) \Psi_2^{(K+1)} \left( \sum_{i=1}^{K+1} \pi_i; \pi_1, \dots, \pi_{K+1}; 0, \dots, 0, 0 \right) = h(\mathbf{p}; \mathbf{\Pi})$$

which illustrates that the model in (12) reduces to the usual (central) Dirichlet model in (2) when the noncentrality parameters are equal to 0. The model in (12) is thus the multivariate analogue of the doubly noncentral beta distribution (see [18]). In the case when  $\Psi_2^{(K)} \left( \sum_{i=1}^K \pi_i; \pi_1, \dots, \pi_K; \frac{\lambda_1}{2} p_1, \dots, \frac{\lambda_K}{2} p_K \right)$  is considered in (12), this would represent the multivariate analogue of the singly noncentral beta distribution of [18].

Bivariate observations from the corresponding distributions were simulated using Algorithm 1 and the associated pdfs are overlaid and presented in in Figure 5 for different values of  $\lambda_1$  and three combinations of  $\mathbf{\Pi}$ s. These results showcase the effect that  $\lambda_1$  has on these three different functions. Figure 5 clearly demonstrates the movement of the centroid of the contour plot.

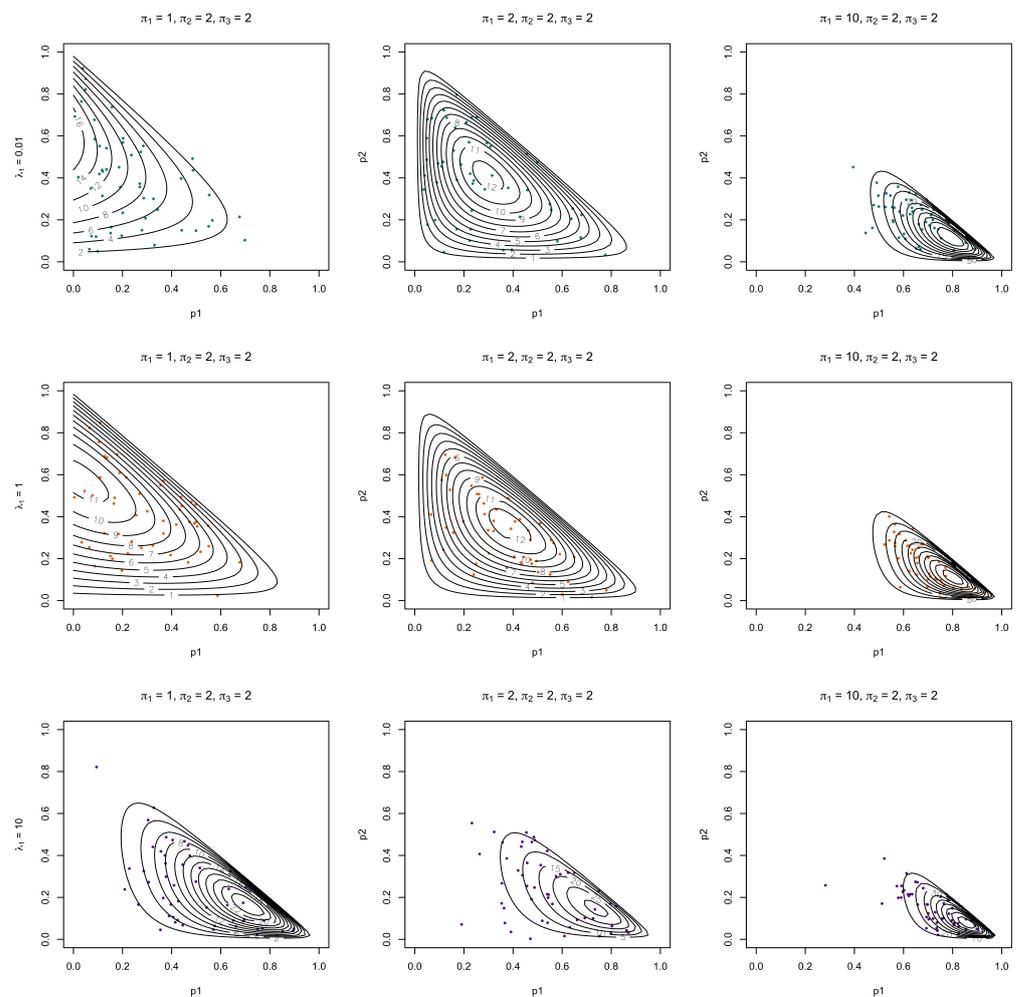


Figure 5. Noncentral Dirichlet Priors (12) for different  $\lambda_1$ s with  $\lambda_2 = 0.8$  and  $\lambda_3 = 0.1, n = 50$ .

Next, the posterior distribution is derived, assuming the noncentral Dirichlet prior (12) together with a multinomial likelihood (1).

**Theorem 4.** Suppose the likelihood function is given by (1) and the prior distribution for  $\mathbf{p}$  is given by (12). Then, the posterior distribution has pdf

$$f(\mathbf{p}|\mathbf{x}) \propto \sum_{\phi} \frac{(\pi_1 + \dots + \pi_{k+1})_{j_1+\dots+j_{k+1}}}{(\pi_1)_{j_1} \dots (\pi_{k+1})_{j_{k+1}} j_1! \dots j_{k+1}!} \left(\frac{\lambda_1}{2} p_1\right)^{j_1} \dots \left(\frac{\lambda_{k+1}}{2} p_{k+1}\right)^{j_{k+1}} \times \left[ \frac{\Gamma(\sum_{j=1}^{k+1} \pi_j + x_j)}{\prod_{j=1}^{k+1} \Gamma(\pi_j + x_j)} \right] p_1^{\pi_1+x_1-1} \dots p_{k+1}^{\pi_{k+1}+x_{k+1}-1} \tag{14}$$

which can be identified as a noncentral Dirichlet distribution with parameters  $(\pi_1 + x_1, \dots, \pi_K + x_K, \pi_{K+1} + x_{K+1})$  and  $\Lambda$ .

**Remark 2.** See that (14) can be represented using the confluent hypergeometric function from Remark 1 as

$$f(\mathbf{p}|\mathbf{x}; \Lambda) = \frac{\Psi_2^{(K+1)}(\sum_{i=1}^{K+1} \pi_i; \pi_1, \dots, \pi_{K+1}; \frac{\lambda_1}{2} p_1, \dots, \frac{\lambda_K}{2} p_K, \frac{\lambda_{K+1}}{2} (1 - \sum_{i=1}^K p_i))}{\Psi_2^{(K+1)}(\sum_{i=1}^{K+1} \pi_i; \pi_1, \dots, \pi_{K+1}; \frac{\lambda_1}{2}, \dots, \frac{\lambda_K}{2}, \frac{\lambda_{K+1}}{2})} \frac{\prod_{j=1}^{K+1} \Gamma(\pi_j + x_j + j_j)}{\Gamma(\sum_{j=1}^{k+1} \pi_j + x_j + j_j)} \times p_1^{\pi_1+x_1-1} \dots p_{k+1}^{\pi_{k+1}+x_{k+1}-1}.$$

The complete product moment of the noncentral Dirichlet posterior is of interest for the power sum, thus we are interested in  $E(p_1^{k_1} p_2^{k_2} \dots p_k^{k_k} p_{k+1}^{k_{k+1}})$ .

**Theorem 5.** Suppose that  $\mathbf{p}|\mathbf{x}$  follows a noncentral Dirichlet distribution with pdf given in (14). Then, the complete product moment is given by

$$E(p_1^{k_1} p_2^{k_2} \dots p_k^{k_k} p_{k+1}^{k_{k+1}}) = \frac{\sum_{\phi} \frac{(\pi_1 + \dots + \pi_{k+1})_{j_1+\dots+j_{k+1}}}{(\pi_1)_{j_1} \dots (\pi_{k+1})_{j_{k+1}} j_1! \dots j_{k+1}!} (\frac{\lambda_1}{2})^{j_1} \dots (\frac{\lambda_{k+1}}{2})^{j_{k+1}} \left[ \frac{\prod_{i=1}^{K+1} \Gamma(\pi_i + x_i + j_i + k_i)}{\Gamma(\sum_{i=1}^{K+1} (\pi_i + x_i + j_i + k_i))} \right]}{\sum_{\phi^*} \frac{(\pi_1 + \dots + \pi_{k+1})_{j_1+\dots+j_{k+1}}}{(\pi_1)_{j_1} \dots (\pi_{k+1})_{j_{k+1}} j_1! \dots j_{k+1}!} (\frac{\lambda_1}{2})^{j_1} \dots (\frac{\lambda_{k+1}}{2})^{j_{k+1}} \left[ \frac{\prod_{i=1}^{K+1} \Gamma(\pi_i + x_i + j_i)}{\Gamma(\sum_{i=1}^{K+1} (\pi_i + x_i + j_i))} \right]}. \tag{15}$$

**Proof.** See Appendix B for the proof. □

### 4. Entropy Estimates

In this section, the Bayesian estimators (16) and (17) based on the posterior distributions (10) and (14) are derived for the power sum (3).

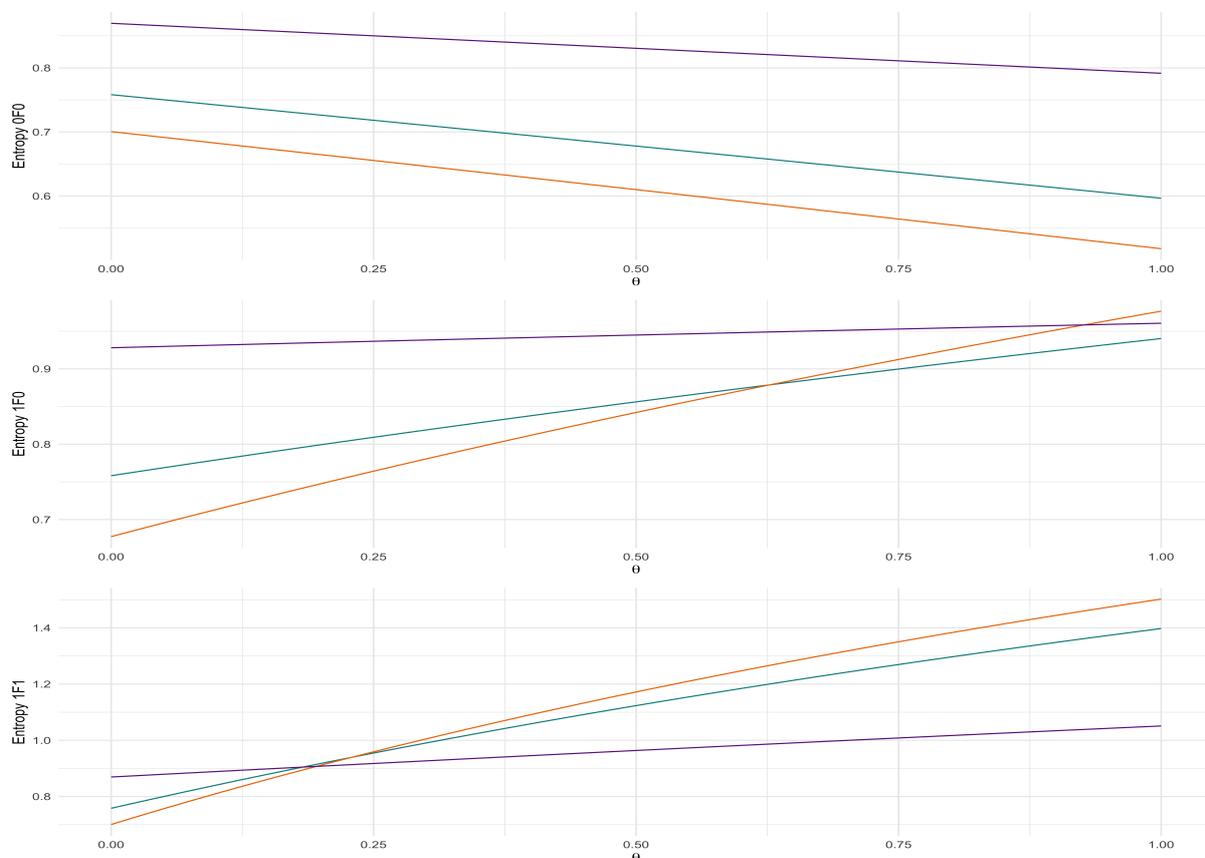
#### 4.1. Dirichlet Generator Prior

Assuming the Dirichlet generator prior, the posterior distribution is given by (10). Using the complete product moments derived in (11), the Bayesian estimator for the power sum (3) can be derived by setting  $k_i = \alpha$  with  $i = 1, \dots, k + 1$  and  $k_{\neq i} = 0$ .

**Theorem 6.** Using (11), the Bayesian estimator for the power sum entropy under the Dirichlet generator posterior (10) is given by:

$$\begin{aligned}
 & \hat{F}_\alpha(p) \\
 &= \sum_{j^*=1}^{K+1} \frac{\Gamma(\pi_{j^*} + x_{j^*} + \alpha) \prod_{j \neq j^*}^{K+1} \Gamma(\pi_j + x_j) \Gamma(\sum_{j=1}^{K+1} \pi_j + x_j)}{\Gamma(\sum_{j=1}^{K+1} \pi_j + x_j + \alpha) \prod_{j=1}^{K+1} \Gamma(\pi_j + x_j)} \\
 &\times \frac{{}_{r+1}F_{q+1}\left(a_1, \dots, a_r, \alpha + \sum_{j=1}^K \pi_j + x_j, b_1, \dots, b_q, \alpha + \sum_{j=1}^{K+1} \pi_j + x_j; \theta\right)}{{}_{r+1}F_{q+1}\left(a_1, \dots, a_r, \sum_{j=1}^K \pi_j + x_j, b_1, \dots, b_q, \sum_{j=1}^{K+1} \pi_j + x_j; \theta\right)} \\
 &= \sum_{j^*=1}^{K+1} \left( \frac{\Gamma\left(\sum_{j=1}^{K+1} \pi_j + x_j\right) \left(\prod_{j=1}^{K+1} \Gamma(\pi_j + x_j)\right)^{-1}}{\left(\Gamma(\pi_{j^*} + x_{j^*} + \alpha) \prod_{j \neq j^*}^{K+1} \Gamma(\pi_j + x_j)\right)^{-1} \Gamma\left(\alpha + \sum_{j=1}^{K+1} \pi_j + x_j\right)} \right)^{-1} \cdot \left( \frac{{}_{r+1}F_{q+1}\left(a_1, \dots, a_r, \alpha + \sum_{j=1}^K \pi_j + x_j, b_1, \dots, b_q, \alpha + \sum_{j=1}^{K+1} \pi_j + x_j; \theta\right)}{{}_{r+1}F_{q+1}\left(a_1, \dots, a_r, \sum_{j=1}^K \pi_j + x_j, b_1, \dots, b_q, \sum_{j=1}^{K+1} \pi_j + x_j; \theta\right)} \right)^{-1}. \tag{16}
 \end{aligned}$$

Using the estimated power sum entropy, we can calculate and investigate the behaviour of Tsallis entropy for various parameter scenarios as illustrated in Figure 6 for the bivariate case  $K = 2$ .



**Figure 6.** Dirichlet generator entropy (16)—Varying  $\theta$ : Set A—blue ( $\pi_1 = 2; \pi_2 = 2; \pi_3 = 2$ ) Set B—orange ( $\pi_1 = 1; \pi_2 = 2; \pi_3 = 2$ ) Set C—purple ( $\pi_1 = 10; \pi_2 = 2; \pi_3 = 2$ ).

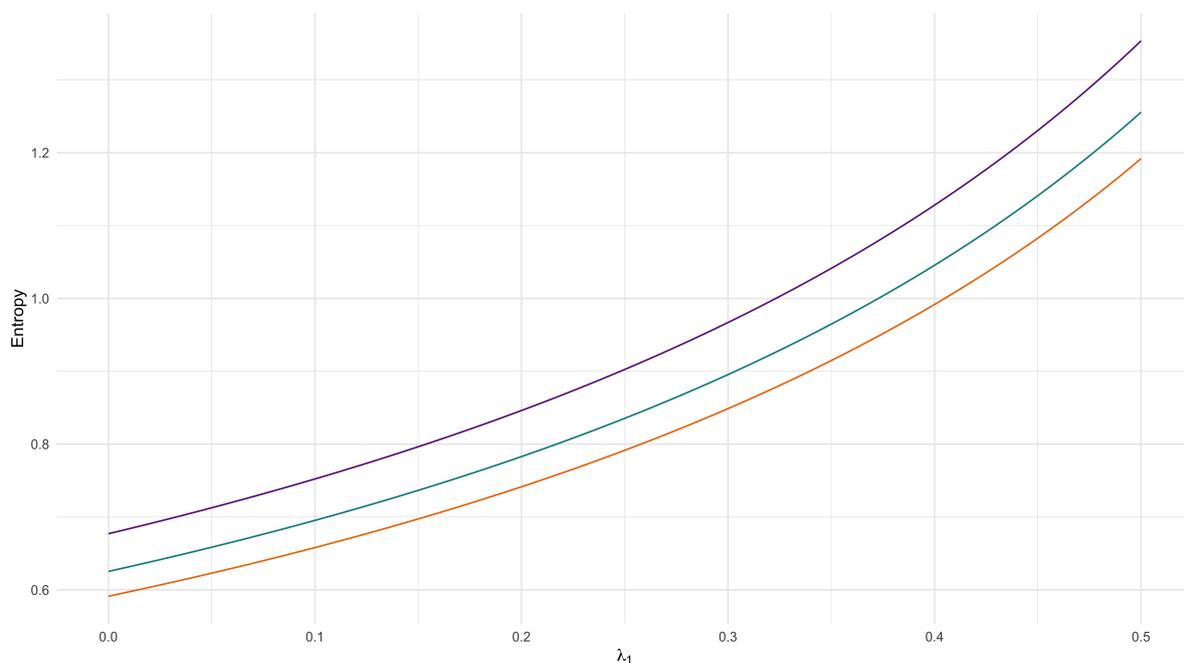
### 4.2. Noncentral Dirichlet Prior

Assuming the noncentral Dirichlet prior and the posterior distribution is given by (14). Using the complete product moments derived in (15), the Bayesian estimator for the power sum entropy (3) can be derived by setting  $k_i = \alpha$  with  $i = 1, \dots, K + 1$  and  $k_{\neq i} = 0$ .

**Theorem 7.** By using (15), the Bayesian estimator for the power sum (4) under the noncentral Dirichlet posterior (14) is given by:

$$\hat{F}_\alpha(p) = \sum_{j^*=1}^{K+1} \left( \frac{\sum_{\phi} \frac{(\pi_1 + \dots + \pi_K + \pi_{K+1})_{j_1 + \dots + j_{K+1}}}{(\pi_1)_{j_1} \dots (\pi_{K+1})_{j_{K+1}}} \left(\frac{\lambda_1}{2}\right)^{j_1} \dots \left(\frac{\lambda_{K+1}}{2}\right)^{j_{K+1}} \frac{\Gamma(\pi_{j^*} + x_{j^*} + j_{j^*} + \alpha) \prod_{j \neq j^*}^{K+1} \Gamma(\pi_j + x_j + j_j)}{\Gamma\left(\alpha + \sum_{j=1}^{K+1} \pi_j + x_j + j_j\right)}}{\sum_{\phi^*} \frac{(\pi_1 + \dots + \pi_K + \pi_{K+1})_{j_1 + \dots + j_{K+1}}}{(\pi_1)_{j_1} \dots (\pi_{K+1})_{j_{K+1}}} \left(\frac{\lambda_1}{2}\right)^{j_1} \dots \left(\frac{\lambda_{K+1}}{2}\right)^{j_{K+1}} \frac{\Gamma(\pi_{j^*} + x_{j^*} + j_{j^*})}{\Gamma\left(\sum_{j=1}^{K+1} \pi_j + x_j + j_j\right)}} \right) \tag{17}$$

Using the estimated power sum entropy, we can calculate the entropies for different parameters of interests as illustrated in Figure 7 for the bivariate case.



**Figure 7.** Noncentral Dirichlet Entropy—Varying  $\lambda_1$ : Set A—blue ( $\pi_1 = 2; \pi_2 = 2; \pi_3 = 2$ ) Set B—orange ( $\pi_1 = 1; \pi_2 = 2; \pi_3 = 2$ ) Set C—purple ( $\pi_1 = 10; \pi_2 = 2; \pi_3 = 2$ ).

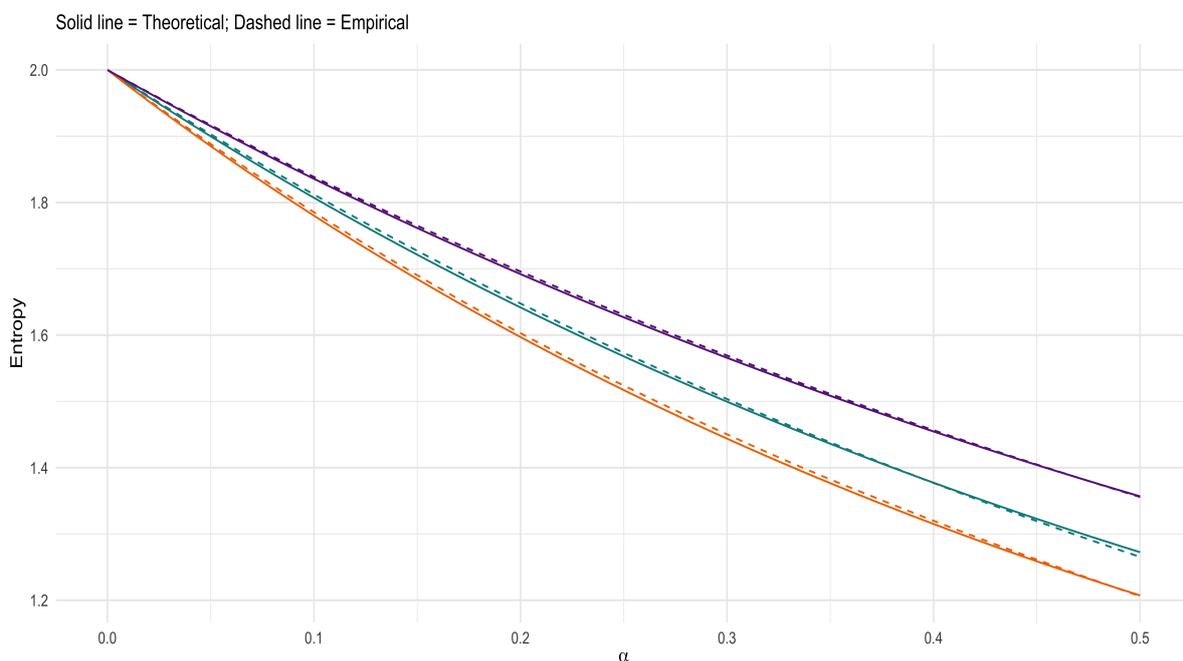
### 4.3. Numerical Experiments of Entropy

The following steps illustrate the empirical behaviour of the Tsallis entropy under consideration for the alternative priors under consideration (Algorithm 2).

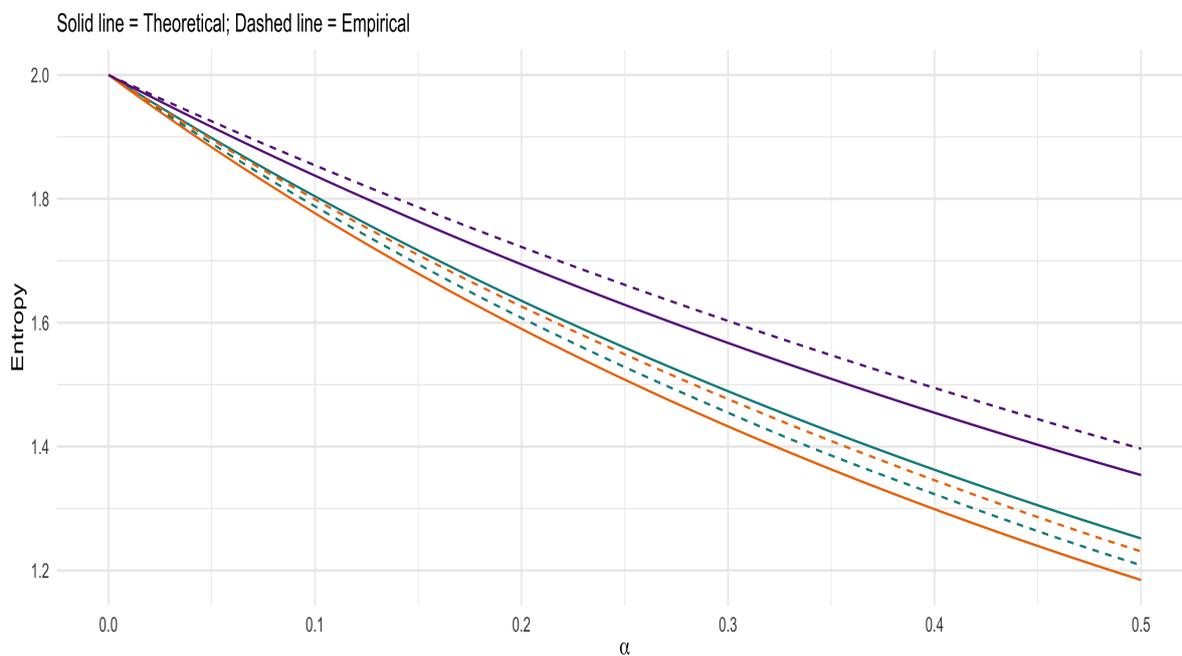
**Algorithm 2** Numerical Experiments of Entropy

1. Simulate  $p_1$  and  $p_2$  from the posterior distribution given by (10) and (14) using the Accept/Rejection method as described earlier for  $n = 50$ ;
2. Calculate  $p_3 = 1 - p_1 - p_2$ ;
3. Calculate the  $p_i^\alpha$  values for all the samples;
4. Determine  $\sum_{i=1}^3 p_i^\alpha$  for each calculation;
5. Calculate the median for the sample of quantities in Step 4 (Note that  $\sum_{i=1}^3 p_i^\alpha$  might not be symmetric, thus the median is used.)
6. The power sum (16) and (17) is used to calculate the Tsallis entropy (5);
7. Steps 1 to 5 are repeated for different parameters and plotted against the analytical entropy in order to illustrate the accuracy of the derived estimates.

Figures 8 and 9 provides validation of the accuracy of the obtained theoretical expression and contribution for the Dirichlet generator  ${}_0F_0$  and noncentral Dirichlet cases with  $x_1 = 1, x_2 = 2$  and  $x_3 = 10$ . From these two figures, it can be seen that the Dirichlet generator prior resulted in empirical results that closely match the theoretical results, while the noncentral Dirichlet shows slight deviations. It is observed that as  $\pi_1$  increases, the Tsallis entropy increases (indicating more uncertainty), while as  $\pi_1$  decreases the Tsallis entropy also decreases (indicating less uncertainty). When considering the location of the density, the changing of  $\pi_1$  leads to densities which tend to the margin of  $p_2$  or towards a specific point along the  $p_1 + p_2 = 1$  line. This shows that as the concentration of the density moves toward a point along the  $p_1 + p_2 = 1$  line, the uncertainty increases and will decrease as the concentration moves towards the small values of  $p_1$ .



**Figure 8.** Dirichlet generator  ${}_0F_0$ —empirical vs calculated Tsallis entropy: for  $\theta = 0.5$  and Set A—blue ( $\pi_1 = 2; \pi_2 = 2; \pi_3 = 2$ ) Set B—orange ( $\pi_1 = 1; \pi_2 = 2; \pi_3 = 2$ ) Set C—purple ( $\pi_1 = 10; \pi_2 = 2; \pi_3 = 2$ ).



**Figure 9.** Noncentral Dirichlet—empirical vs calculated Tsallis entropy: for  $\lambda_1 = 0.1$ ;  $\lambda_2 = 0.8$  and  $\lambda_3 = 0.1$  with Set A—blue ( $\pi_1 = 2$ ;  $\pi_2 = 2$ ;  $\pi_3 = 2$ ) Set B—orange ( $\pi_1 = 1$ ;  $\pi_2 = 2$ ;  $\pi_3 = 2$ ) Set C—purple ( $\pi_1 = 10$ ;  $\pi_2 = 2$ ;  $\pi_3 = 2$ ).

## 5. Conclusions

This study focussed on the power sum functional and its estimation as a key tool to model a generalised entropy form, namely Tsallis entropy, via a Bayesian approach. In particular, previous unconsidered Dirichlet priors have been proposed and studied, offering the practitioner more pliable options given experimental data. Specific choices of the proposed Dirichlet family allow for positive correlation in addition to the usual negative correlation characteristic. An example illustrated theoretical results accurately described empirical entropy. Future work could include further investigations into generalised functionals and their modeling in this information-theoretic environment.

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

**Funding:** This work is based on the research supported in part by the National Research Foundation of South Africa (SARChI Research Chair- UID: 71199; and Grant ref. SRUG190308422768 nr. 120839) as well as the Research Development Programme at the University of Pretoria 296/2021. Opinions expressed and conclusions arrived at are those of the author and are not necessarily to be attributed to the NRF.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** The authors would like to thank the anonymous reviewers for their insightful comments which led to the improvement of this paper. The support of the Department of Statistics at the University of Pretoria is acknowledged.

**Conflicts of Interest:** The authors declare no conflict of interest.

### Abbreviations

The following abbreviations are used in this manuscript:

- MLE Maximum likelihood estimation
- pmf Probability mass function
- pdf Probability density function

### Appendix A. Proof of Complete Product Moments of Dirichlet Generator (11)

**Proof.** The definition of the complete product moment of a  $(K + 1)$  variate variable  $\mathbf{Y}$  with pdf  $f(\mathbf{y})$  is given by

$$E\left(\prod_{i=1}^{k+1} Y_i^{x_i}\right) = \int \cdots \int \prod_{i=1}^{k+1} y_i^{x_i} f(\mathbf{y}) dy_1 \cdots dy_{k+1}$$

and since we know that the posterior distribution are given by Dirichlet generator distributions with parameters  $(\pi_1 + x_1; \dots; \pi_{k+1} + x_{k+1})$  we can show that:

$$\begin{aligned} E(p_1^{k_1} \cdots p_{k+1}^{k_{k+1}}) &= \int \cdots \int_{\mathcal{A}} \prod_{i=1}^{k+1} p_i^{k_i} f(\mathbf{p}) dp_1 \cdots dp_{k+1} \\ &= C^{-1} \frac{\prod_{i=1}^{k+1} \Gamma(\pi_i + x_i + k_i)}{\Gamma(\sum_{i=1}^{k+1} \pi_i + x_i + k_i)} \frac{\Gamma(\pi_1 + x_1 + \cdots + p_{k+1} + x_{k+1})}{\Gamma(\pi_1 + x_1) \cdots \Gamma(\pi_{k+1} + x_{k+1})} C_* \\ &\quad \int \cdots \int C_*^{-1} p_1^{\pi_1 + x_1 + k_1} \cdots p_{k+1}^{\pi_{k+1} + x_{k+1} + k_{k+1}} \frac{\Gamma(\sum_{i=1}^{k+1} \pi_i + x_i + k_i)}{\prod_{i=1}^{k+1} \Gamma(\pi_i + x_i + k_i)} \\ &\quad \times {}_rF_q\left(a_1, \dots, a_r, b_1, \dots, b_q; \theta \left(\sum_{i=1}^k p_i\right)\right) dp_1 \cdots dp_{k+1} \end{aligned}$$

where  $C$  and  $C_*$  corresponds to the normalising constants as in (7) and (9), respectively with parameters with parameters  $(\pi_1 + x_1; \dots; \pi_{k+1} + x_{k+1})$  and  $(\pi_1 + x_1 + k_1; \dots; \pi_{k+1} + x_{k+1} + k_{k+1})$ . Since the integral becomes the Dirichlet generator pdf (6) with parameters  $(\pi_1 + x_1 + k_1, \dots, \pi_{k+1} + x_{k+1} + k_{k+1})$  this will become 1 and the complete product moment will simplify to

$$\begin{aligned} E(p_1^{k_1} \cdots p_{k+1}^{k_{k+1}}) &= \frac{\prod_{i=1}^{k+1} (\pi_i + x_i)_{k_i}}{(\sum_{i=1}^{k+1} \pi_i + x_i)_{k_1 + \dots + k_{k+1}}} \\ &\quad \times \frac{{}_{r+1}F_{q+1}(a_1, \dots, a_r, \sum_{i=1}^k \pi_i + x_i + k_i, b_1, \dots, b_q, \sum_{i=1}^{k+1} \pi_i + x_i + k_i; \theta)}{{}_{r+1}F_{q+1}(a_1, \dots, a_r, \sum_{i=1}^k \pi_i + x_i, b_1, \dots, b_q, \sum_{i=1}^{k+1} \pi_i + x_i; \theta)}. \end{aligned}$$

□

### Appendix B. Proof of Complete Product Moments of Noncentral Dirichlet (15)

**Proof.** The definition of the complete product moment of a  $(K + 1)$  variate variable  $\mathbf{Y}$  with pdf  $f(\mathbf{y})$  is given by

$$E\left(\prod_{i=1}^{k+1} Y_i^{x_i}\right) = \int \cdots \int \prod_{i=1}^{k+1} y_i^{x_i} f(\mathbf{y}) dy_1 \cdots dy_{k+1}$$

and since we know that the posterior distribution are given by a noncentral Dirichlet pdf (12) with parameters  $(\pi_1 + x_1; \dots; \pi_{k+1} + x_{k+1})$  we can show that

$$\begin{aligned}
 E(p_1^{k_1} \dots p_{k+1}^{k_{k+1}}) &= \int \dots \int_{\mathcal{A}} \prod_{i=k_1}^{k_{k+1}} p_i^{k_i} f(\mathbf{p}) dp_1 \dots dp_{k+1} \\
 &= \frac{\sum_{\phi} \frac{(\pi_1 + \dots + \pi_{k+1})_{j_1 + \dots + j_{k+1}}}{(\pi_1)_{j_1} \dots (\pi_{k+1})_{j_{k+1}}} \left(\frac{\lambda_1}{2}\right)^{j_1} \dots \left(\frac{\lambda_{k+1}}{2}\right)^{j_{k+1}} \left[ \frac{\prod_{i=1}^{K+1} \Gamma(\pi_i + x_i + j_i + k_i)}{\Gamma(\sum_{i=1}^{K+1} (\pi_i + x_i + j_i + k_i))} \right]}{\sum_{\phi^*} \frac{(\pi_1 + \dots + \pi_{k+1})_{j_1 + \dots + j_{k+1}}}{(\pi_1)_{j_1} \dots (\pi_{k+1})_{j_{k+1}}} \left(\frac{\lambda_1}{2}\right)^{j_1} \dots \left(\frac{\lambda_{k+1}}{2}\right)^{j_{k+1}} \frac{\prod_{i=1}^{K+1} \Gamma(x_j + x_j + j_j)}{\Gamma(\sum_{j=1}^{K+1} \pi_j + x_j + j_j)}} \\
 &\times \int \dots \int p_1^{\pi_1 + x_1 + j_1 + k_1 - 1} \dots p_{k+1}^{\pi_{k+1} + x_{k+1} + j_{k+1} + k_{k+2} - 1} \\
 &\times \frac{\Gamma(\sum_{i=1}^{k+1} \pi_i + x_i + j_i + k_i)}{\prod_{i=1}^{k+1} \Gamma(\pi_i + x_i + j_i + k_i)} dp_1 \dots dp_{k+1}.
 \end{aligned}$$

Since the integral becomes the noncentral Dirichlet pdf with parameters  $(\pi_1 + x_1 + k_1 + j_1, \dots, \pi_{k+1} + x_{k+1} + k_{k+1} + j_{k+1})$  this will become 1 and the complete product moment will simplify to

$$E(p_1^{k_1} \dots p_{k+1}^{k_{k+1}}) = \frac{\sum_{\phi} \frac{(\pi_1 + \dots + \pi_{K+1})_{j_1 + \dots + j_{K+1}}}{(\pi_1)_{j_1} \dots (\pi_{K+1})_{j_{K+1}}} \left(\frac{\lambda_1}{2}\right)^{j_1} \dots \left(\frac{\lambda_{K+1}}{2}\right)^{j_{K+1}} \left[ \frac{\prod_{i=1}^{K+1} \Gamma(\pi_i + x_i + j_i + k_i)}{\Gamma(\sum_{i=1}^{K+1} (\pi_i + x_i + j_i + k_i))} \right]}{\sum_{\phi^*} \frac{(\pi_1 + \dots + \pi_{K+1})_{j_1 + \dots + j_{K+1}}}{(\pi_1)_{j_1} \dots (\pi_{K+1})_{j_{K+1}}} \left(\frac{\lambda_1}{2}\right)^{j_1} \dots \left(\frac{\lambda_{K+1}}{2}\right)^{j_{K+1}} \left[ \frac{\prod_{i=1}^{K+1} \Gamma(\pi_i + x_i + j_i)}{\Gamma(\sum_{i=1}^{K+1} (\pi_i + x_i + j_i))} \right]}.$$

□

**References**

1. Archer, E.; Park, I.M.; Pillow, J. Bayesian entropy estimation for countable discrete distributions. *J. Mach. Learn. Res.* **2014**, *15*, 2833–2868.
2. Ilić, V.; Korbel, J.; Gupta, S.; Scarfone, A.M. An overview of generalized entropic forms. *arXiv* **2021**, arXiv:2102.10071.
3. Rashad, M.; Iqbal, Z.; Hanif, M. Characterizations and entropy measures of the Libby-Novick generalized beta distribution. *Adv. Appl. Stat.* **2020**, *63*, 235–259. [\[CrossRef\]](#)
4. Jiao, J.; Venkat, K.; Han, Y.; Weissman, T. Maximum likelihood estimation of functionals of discrete distributions. *IEEE Trans. Inf. Theory* **2017**, *63*, 6774–6798. [\[CrossRef\]](#)
5. Contreras Rodríguez, L.; Madarro-Capó, E.J.; Legón-Pérez, C.M.; Rojas, O.; Sosa-Gómez, G. Selecting an Effective Entropy Estimator for Short Sequences of Bits and Bytes with Maximum Entropy. *Entropy* **2021**, *23*, 561. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Wolpert, D.H.; Wolf, D. Estimating functions of probability distributions from a finite set of samples. *Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.* **1995**, *52*, 6841–6854. [\[CrossRef\]](#)
7. Han, Y.; Jiao, J.; Weissman, T. Does Dirichlet prior smoothing solve the Shannon entropy estimation problem? In Proceedings of the IEEE International Symposium on Information Theory, Hong Kong, China, 14–19 June 2015; pp. 1367–1371.
8. Little, D.J.; Toomey, J.P.; Kane, D.M. Efficient Bayesian estimation of permutation entropy with Dirichlet priors. *arXiv* **2021**, arXiv:2104.08991.
9. Zamzami, N.; Bouguila, N. Hybrid generative discriminative approaches based on Multinomial Scaled Dirichlet mixture models. *Appl. Intell.* **2019**, *49*, 3783–3800. [\[CrossRef\]](#)
10. Holste, D.; Grosse, I.; Herzel, H. Bayes’ estimators of generalized entropies. *J. Phys. A Math. Gen.* **1998**, *31*, 2551. [\[CrossRef\]](#)
11. Bodvin, L.J.S.; Bekker, A.; Roux, J.J. Shannon entropy as a measure of certainty in a Bayesian calibration framework with bivariate beta priors: Theory and methods. *S. Afr. Stat. J.* **2011**, *45*, 171–204.
12. Sánchez, L.E.; Nagar, D.; Gupta, A. Properties of noncentral Dirichlet distributions. *Comput. Math. Appl.* **2006**, *52*, 1671–1682. [\[CrossRef\]](#)
13. Kang, M.S.; Kim, K.T. Automatic SAR Image Registration via Tsallis entropy and Iterative Search Process. *IEEE Sens. J.* **2020**, *20*, 7711–7720. [\[CrossRef\]](#)
14. Mathai, A.M.; Haubold, H.J. On generalized entropy measures and pathways. *Phys. A Stat. Mech. Appl.* **2007**, *385*, 493–500. [\[CrossRef\]](#)
15. Gradshteyn, I.S.; Ryzhik, I.M. *Table of Integrals, Series, and Products*; Academic Press: Cambridge, MA, USA, 2014.
16. Ferreira, J.T.; Bekker, A.; Arashi, M. Bivariate noncentral distributions: An approach via the compounding method. *S. Afr. Stat. J.* **2016**, *50*, 103–122.
17. Bekker, A.; Ferreira, J.T. Bivariate gamma type distributions for modeling wireless performance metrics. *Stat. Optim. Inf. Comput.* **2018**, *6*, 335–353. [\[CrossRef\]](#)
18. Ongaro, A.; Orsi, C. Some results on non-central beta distributions. *Statistica* **2015**, *75*, 85–100.