



Heriot-Watt University  
Research Gateway

# Poisson and Gaussian approximations of the power divergence family of statistics

## Citation for published version:

Daly, FA 2024, 'Poisson and Gaussian approximations of the power divergence family of statistics', *ALEA: Latin American Journal of Probability and Mathematical Statistics*, vol. 21, pp. 25-37.  
<https://doi.org/10.30757/ALEA.v21-02>

## Digital Object Identifier (DOI):

[10.30757/ALEA.v21-02](https://doi.org/10.30757/ALEA.v21-02)

## Link:

[Link to publication record in Heriot-Watt Research Portal](#)

## Document Version:

Peer reviewed version

## Published In:

ALEA: Latin American Journal of Probability and Mathematical Statistics

## General rights

Copyright for the publications made accessible via Heriot-Watt Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

## Take down policy

Heriot-Watt University has made every reasonable effort to ensure that the content in Heriot-Watt Research Portal complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [open.access@hw.ac.uk](mailto:open.access@hw.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.

# Poisson and Gaussian approximations of the power divergence family of statistics

Fraser Daly\*

November 1, 2023

**Abstract** Consider the family of power divergence statistics based on  $n$  trials, each leading to one of  $r$  possible outcomes. This includes the log-likelihood ratio and Pearson's statistic as important special cases. It is known that in certain regimes (e.g., when  $r$  is of order  $n^2$  and the allocation is asymptotically uniform as  $n \rightarrow \infty$ ) the power divergence statistic converges in distribution to a linear transformation of a Poisson random variable. We establish explicit error bounds in the Kolmogorov (or uniform) metric to complement this convergence result, which may be applied for any values of  $n$ ,  $r$  and the index parameter  $\lambda$  for which such a finite-sample bound is meaningful. We further use this Poisson approximation result to derive error bounds in Gaussian approximation of the power divergence statistics.

**Key words and phrases:** Cressie–Read statistics; log-likelihood ratio; Pearson's statistic; Poisson approximation; normal approximation; uniform metric

**MSC 2020 subject classification:** 62E17; 60E15; 60F05

## 1 Introduction and main result

In the setting of a multinomial goodness-of-fit test, suppose we have  $n$  independent trials, each of which is assigned to one of  $r = r(n)$  classes. We denote by  $\mathbf{p} = (p_1, \dots, p_r)$  the corresponding classification probabilities, where  $p_1 + \dots + p_r = 1$ , and by  $N_1, \dots, N_r$  the observed frequencies in each class following these  $n$  trials, so that  $N_1 + \dots + N_r = n$ .

Cressie and Read [5] introduced the power divergence family of statistics for testing the hypothesis  $H_0 : \mathbf{p} = \mathbf{p}_0$  against alternatives of the form  $H_1 : \mathbf{p} = \mathbf{p}_1$ . For a given index parameter  $\lambda \in \mathbb{R}$ , these statistics are defined by

$$T_\lambda = n \sum_{j=1}^r p_j g_\lambda \left( \frac{N_j}{np_j} \right),$$

---

\*Department of Actuarial Mathematics and Statistics and the Maxwell Institute for Mathematical Sciences, Heriot–Watt University, Edinburgh EH14 4AS, UK. E-mail: f.daly@hw.ac.uk

where

$$g_\lambda(x) = \begin{cases} 2x \log(x), & \text{if } \lambda = 0, \\ -2 \log(x), & \text{if } \lambda = -1, \\ \frac{2x(x^\lambda - 1)}{\lambda(\lambda + 1)}, & \text{otherwise.} \end{cases} \quad (1.1)$$

As special cases, these statistics include Pearson's  $\chi^2$  ( $\lambda = 1$ ), the log-likelihood ratio ( $\lambda = 0$ ), the Freeman–Tukey statistic ( $\lambda = -1/2$ ) and the modified log-likelihood ratio ( $\lambda = -1$ ). Cressie and Read [5] also suggest  $T_{2/3}$  as a good alternative to Pearson's statistic and the log-likelihood ratio. The choice of the index parameter  $\lambda$  for applications will depend on the type of departure from  $H_0$  that the practitioner wishes to detect. For example, Cressie and Read [5] note that the effect of a single ratio  $N_j/p_j$  on the test statistic  $T_\lambda$  increases as  $|\lambda|$  increases; see their Section 6.2. See also Sections 4.5, 5.5 and 6.7 of [12] for discussions of appropriate choices of  $\lambda$  for testing different hypotheses and in different settings.

It is well known that in many cases  $T_\lambda$  has an asymptotic chi-square distribution as  $n \rightarrow \infty$ . Rates of convergence and error bounds in the chi-square approximation of  $T_\lambda$  for  $\lambda > -1$  have been obtained to complement and quantify this limit theorem by several authors: see, for example, [1, 2, 7, 11, 17] and references therein. A Gaussian limit for a suitably scaled version of  $T_\lambda$  in some regimes is also well known [5, 10]. Though we are not aware of any error bounds that accompany this central limit theorem, there are rates of convergence known in particular examples, such as for Pearson's statistic (see Section 4.2 of [6]). Other limiting distributions are also possible. Recently, Rempała and Wesolowski [13] have shown that in some regimes, a suitably normalised version of  $T_\lambda$  has an asymptotic Poisson distribution as  $n \rightarrow \infty$ ; we discuss their result in more detail below. Although this limit theorem is a recent observation, Poisson limits in this setting are not unexpected: Steck [16] observed many years ago that in the special case of equally likely classes, Pearson's statistic has an asymptotic Poisson distribution.

Our main aim in this paper is to complement the limit theorem of [13] for power divergence statistics by providing an explicit error bound in the Poisson approximation of  $T_\lambda$  in Theorem 1 below. That is, we give an upper bound on the Kolmogorov (or uniform) distance between  $\tilde{T}_\lambda$ , a suitably normalised version of  $T_\lambda$  to be defined precisely in (1.3) below, and a suitable Poisson random variable  $Z$ , given by

$$d_K(\tilde{T}_\lambda, Z) = \sup_{y \in \mathbb{R}} |\mathbb{P}(\tilde{T}_\lambda \geq y) - \mathbb{P}(Z \geq y)|. \quad (1.2)$$

We will also use this Poisson approximation result to derive error bounds in Gaussian approximation of the power divergence family of statistics, again in the Kolmogorov distance.

This choice of metric is governed by practical considerations in the use of the statistics  $T_\lambda$ . Theorem 1 can be used to give upper bounds on the probability of type-I and type-II errors in the test of the hypotheses  $H_0$  versus  $H_1$  described above. For example, to bound the probability of a type-I error, set  $\mathbf{p} = \mathbf{p}_0$ . One is then interested in bounding the probability that the observed value of  $T_\lambda$  falls into the rejection region. This can be done by approximating  $T_\lambda$  by a suitably scaled Poisson random variable, for which the corresponding error probability can be calculated or bounded, and bounding the error in this approximation using Theorem 1. These estimates will be useful in settings where this error estimate is small, and the Poisson approximation is known to hold. Similar comments apply to the estimation of type-II error probabilities.

In a Poisson approximation setting it is relatively straightforward to directly derive error bounds in the Kolmogorov distance, but this is typically not the case for chi-square and Gaussian

approximations. In the setting of chi-square approximation for  $T_\lambda$ , [7] was able to directly obtain bounds only in a metric based on smooth test functions. Similarly, the bounds given by [6] in the Gaussian approximation of Pearson's  $\chi^2$  are for smooth test functions. Bounds in Kolmogorov distance can be obtained from such results for smooth test functions, but typically such bounds give sub-optimal rates of convergence. Rates of convergence (but not explicit error bounds) in Kolmogorov distance for the chi-square approximation of  $T_\lambda$  have also been established directly by several authors: see [1, 2, 17], among others. The recent paper [11] also derives bounds in the Kolmogorov distance, employing some auxiliary randomisation and showing that the upper bounds hold with high probability with respect to this randomisation.

We will focus here on the case  $\lambda > -1$ , since this is the largest class of power divergence statistics we can reasonably consider. For  $\lambda \leq -1$ , if one or more of  $N_1, \dots, N_r$  is zero then we have  $T_\lambda = \infty$ . Since this will happen with high probability in our regimes of interest (in which a typical application will have  $r$  of order  $n^2$ ), bounds on  $d_K(\tilde{T}_\lambda, Z)$  become meaningless in this case.

We use the remainder of this section to discuss the Poisson limit theorem of Rempała and Wesolowski [13], and to state our main Poisson result in Theorem 1 below. Some examples and illustrations of this result are then given in Section 2. In Section 3 we derive our Gaussian approximation result. The proof of Theorem 1 is given in Section 4, following which some concluding remarks make up Section 5.

Let the random variable  $P$  be defined by  $\mathbb{P}(P = p_j) = p_j$ , for  $j = 1, \dots, r$ . With this notation, Theorem 2.1 of Rempała and Wesolowski [13] establishes a Poisson limit theorem for  $T_\lambda$  for any  $\lambda \geq 0$  under the following three conditions:

- (i)  $n \max_j p_j \rightarrow 0$  as  $n \rightarrow \infty$ ;
- (ii)  $\frac{n \text{Var}(P^{-\lambda})}{\mathbb{E}[P^{-\lambda}]^2} \rightarrow 0$  as  $n \rightarrow \infty$  when  $\lambda > 0$ , or  $n \text{Var}(\log(P)) \rightarrow 0$  as  $n \rightarrow \infty$  when  $\lambda = 0$ ; and
- (iii) there exists a constant  $0 < \eta < \infty$  such that  $\frac{n^2 \mathbb{E}[P^{1-k\lambda}]}{\mathbb{E}[P^{-\lambda}]^k} \rightarrow \eta$  as  $n \rightarrow \infty$ , for each  $k = 1, 2, 3$ .

More precisely, under the conditions (i)–(iii) above, Rempała and Wesolowski show that the normalised version  $\tilde{T}_\lambda$  of  $T_\lambda$ , defined by

$$\tilde{T}_\lambda = \frac{n^\lambda}{g_\lambda(2)\mathbb{E}[P^{-\lambda}]} (T_\lambda - n^2 \mathbb{E}[Pg_\lambda([nP]^{-1})]) , \quad (1.3)$$

converges in distribution to a Poisson distribution with mean  $\eta/2$  as  $n \rightarrow \infty$ . Our Theorem 1 below gives a quantitative version of this result. Rempała and Wesolowski [13] are motivated by [4, 8, 9] to consider the doubly asymptotic setting where  $r = r(n) \rightarrow \infty$  as  $n \rightarrow \infty$ . In considering a rate of convergence based on our Theorem 1 we may also have such a setting in mind, but we emphasise that our result gives an explicit error bound which may be evaluated for any given values of  $n$  and  $r$ .

**Theorem 1.** *Let  $\lambda > -1$  and  $n \geq 4$ . Assume that  $p_j \leq 0.13$  and  $(n+1)p_j < 4$  for each  $j = 1, \dots, r$ . Let  $\pi_j = \binom{n}{2} p_j^2 (1-p_j)^{n-2}$  and  $\mu = \pi_1 + \dots + \pi_r$ , and let  $Z$  have a Poisson*

distribution with mean  $\mu$ . Then

$$d_K(\tilde{T}_\lambda, Z) \leq 45 \min\{1, \mu^{-1}\} \left( 6n \left[ \sum_{j=1}^r p_j \pi_j \right]^2 + \sum_{j=1}^r \pi_j^2 + \frac{24\mu^2}{n} \right) \\ + 18 (5.55\mu c_\lambda)^{0.49/c_\lambda} + \frac{8n\mu \max_j p_j}{4 - (n+1) \max_j p_j} + \min \left\{ 8.1d_\lambda, 2.15 \left( \frac{d_\lambda}{\mu} \right)^{1/3} \right\}, \quad (1.4)$$

where

$$c_\lambda = \max_j \left| \frac{1}{p_j^\lambda \mathbb{E}[P^{-\lambda}]} - 1 \right|, \quad \text{and} \quad d_\lambda = \begin{cases} \frac{n \text{Var}(\log(P))}{4 \log(2)^2}, & \text{if } \lambda = 0, \\ \frac{n \text{Var}(P^{-\lambda})}{4(2^\lambda - 1)^2 \mathbb{E}[P^{-\lambda}]^2}, & \text{otherwise.} \end{cases}$$

If  $c_\lambda = 0$ , the second term of the upper bound (1.4) should be read as zero. The proof of Theorem 1 is given in Section 4 below, before which we consider some applications of that result in Section 2 and a Gaussian approximation result which may be derived from it in Section 3. We conclude this section with some brief discussion of our upper bound, beginning by noting that while the Poisson limit theorem of [13] applies only for  $\lambda \geq 0$ , our result provides bounds for all  $\lambda > -1$ . See Section 2.3 below for an example where we obtain a Poisson limit for this full range of values of  $\lambda$ .

**Remark 2.** *The approximating Poisson random variable  $Z$  in Theorem 1 has a mean  $\mu$  which depends on  $n$ . In some applications it may be more convenient to replace this with a different Poisson distribution whose mean is equal to  $\lim_{n \rightarrow \infty} \mu$ . Note that under the conditions of Theorem 2.1 of [13], this limit is equal to  $\eta/2$ . Letting  $Z'$  have a Poisson distribution with mean  $\eta/2$ , we obtain an upper bound for  $d_K(\tilde{T}_\lambda, Z')$  by taking the upper bound (1.4) of Theorem 1 and adding to it the additional term*

$$45 \min \left\{ \sqrt{\frac{2}{e}} \left| \sqrt{\mu} - \sqrt{\frac{\eta}{2}} \right|, \left| \mu - \frac{\eta}{2} \right| \right\}.$$

*This is apparent from the proof of Theorem 1 below, by using the triangle inequality at an appropriate point and applying Theorem 1 of [14] to bound  $d_K(Z, Z')$ , noting that the total variation distance bounded therein serves as an upper bound for the Kolmogorov distance we use here.*

**Remark 3.** *The assumptions that  $p_j \leq 0.13$  and  $(n+1)p_j < 4$  for each  $j$  in Theorem 1 may be relaxed at the cost of a somewhat worse upper bound. However, since we have in mind the setting where  $n \max_j p_j \rightarrow 0$  as  $n \rightarrow \infty$ , we lose very little by making these assumptions here.*

**Remark 4.** *It is clear that we need  $d_\lambda \rightarrow 0$  if the final term of (1.4) is to go to zero as  $n$  grows large, with the rate of convergence then governed by the first term of the minimum therein. However, in particular applications, for a given values of the parameters of the problem, it may be the case that the second term of the minimum is smaller than the first. We therefore leave both terms in the statement of our bound.*

We emphasise that Theorem 1 applies for any given  $\lambda > -1$ ,  $n \geq 4$ ,  $r$  and the  $p_j$ , and can be used to estimate the Poisson approximation error for any set of values of these parameters. Nevertheless, thinking of the limit theorem implied by our upper bound, it is natural to compare it with the conditions (i)–(iii) above for Poisson convergence derived by Rempala and Wesolowski [13]. We have that our upper bound (1.4) converges to zero as  $n \rightarrow \infty$  if conditions (i), (ii), (iv) and (v) hold, where these latter two conditions are given as follows:

(iv) there exists a constant  $0 < \eta < \infty$  such that  $n^2\mathbb{E}[P] \rightarrow \eta$  as  $n \rightarrow \infty$ ; and

(v)  $c_\lambda \rightarrow 0$  as  $n \rightarrow \infty$ .

To see this, we note that convergence of the final term of our upper bound follows immediately from (ii). Noting that  $\mu \leq \frac{n^2}{2} \sum_{j=1}^r p_j^2 = \frac{n^2}{2} \mathbb{E}[P] \rightarrow \eta/2$  by (iv), convergence of the third term to zero is a consequence of (i). Condition (v) gives convergence of the second term. For the first term of the upper bound, (iv) has already given a finite limit for  $\mu$ , so that  $\mu^2/n \rightarrow 0$ . Similarly,  $\sum_{j=1}^r \pi_j^2 \leq \frac{n^4}{4} (\max_j p_j)^2 \mathbb{E}[P] \rightarrow 0$  by (i) and (iv). Finally,

$$n \left[ \sum_{j=1}^r p_j \pi_j \right]^2 \leq n \left[ \sum_{j=1}^r \frac{n^2}{2} p_j^3 \right]^2 \leq \frac{n^5}{4} (\max_j p_j)^2 \mathbb{E}[P]^2 \rightarrow 0,$$

again using (i) and (iv).

Some comparison of the conditions (i)–(iii) used in the Poisson limit theorem of [13] and our Theorem 1 will be made in the examples considered in Section 2. We remark here that we have  $c_\lambda = 0$  when either  $\lambda = 0$  or the allocation is uniform (i.e.,  $p_j = 1/r$  for each  $j = 1, \dots, r$ ). So (v) suggests that the allocation should be asymptotically uniform. In either of these cases where  $c_\lambda = 0$ , we have that (iii) and (iv) are equivalent, since (iii) reduces to the statement that  $n^2\mathbb{E}[P] \rightarrow \eta$ . The proof of Theorem 1 makes it clear that our upper bound is geared towards the natural setting where the allocation is asymptotically uniform and  $c_\lambda \rightarrow 0$ , in which case it appears that conditions (iii) and (iv) play similar roles.

## 2 Examples and illustrations

In this section we illustrate the bound (1.4) of Theorem 1 with some examples and applications. As noted above, if either  $\lambda = 0$  or the allocation is uniform then  $c_\lambda = 0$ , and so the second term of the upper bound (1.4) vanishes. We will detail both these special cases here, beginning in Section 2.1 with the log-likelihood ratio,  $\lambda = 0$ , as an initial illustration. For the case of uniform allocations, we will first study the more general example of the discrete power distribution in Section 2.2, which we will then specialise to the case of uniform allocation in Section 2.3.

### 2.1 The log-likelihood ratio

In the case  $\lambda = 0$  the upper bound (1.4) simplifies, giving the following Poisson approximation bound for the log-likelihood ratio  $T_0 = 2 \sum_{j=1}^r N_j \log \left( \frac{N_j}{np_j} \right)$ .

**Corollary 5.** *Let  $n \geq 4$ , and assume that  $p_j \leq 0.13$  and  $(n+1)p_j < 4$  for each  $j = 1, \dots, r$ . Let  $\pi_j$  and  $\mu$  be as in Theorem 1, and let  $Z$  have a Poisson distribution with mean  $\mu$ . Then*

$$d_K(\tilde{T}_0, Z) \leq 45 \min\{1, \mu^{-1}\} \left( 6n \left[ \sum_{j=1}^r p_j \pi_j \right]^2 + \sum_{j=1}^r \pi_j^2 + \frac{24\mu^2}{n} \right) \\ + \frac{8n\mu \max_j p_j}{4 - (n+1) \max_j p_j} + \min \left\{ 8.1d_0, 2.15 \left( \frac{d_0}{\mu} \right)^{1/3} \right\},$$

where  $d_0 = \frac{n \text{Var}(\log(P))}{4 \log(2)^2}$ .

As noted above, since  $c_0 = 0$ , the condition (iii) from the limit theorem of Rempala and Wesolowski [13] is equivalent to (iv) above in this example.

## 2.2 The discrete power distribution

We now consider the example in which  $p_j = \frac{j^{-a}}{z_r(a)}$  for  $j = 1, \dots, r$  and for some parameter  $a \in [0, 1]$ , where the normalising constant is the generalised harmonic number defined by  $z_r(a) = \sum_{k=1}^r k^{-a}$ . This is the discrete power distribution with parameters  $a$  and  $r$ . This example was also considered in Example 3.1 of [13], where it is shown that the restriction  $a < 1/2$  is necessary to ensure that  $n \max_j p_j \rightarrow 0$  as  $n \rightarrow \infty$  in the case where  $r$  is of order  $n^2$ . In doing so they used the asymptotics  $z_r(a) \approx \frac{r^{1-a}}{1-a}$  for  $a \neq 1$ , which we can also employ in understanding the behaviour of the terms that appear in our bound in this setting. For future use we also note that  $z_r(1) = O(\log r)$ .

In the special case  $a = 0$  we have the uniform distribution  $p_j = 1/r$  for each  $j$ ; we will look in more detail at this special case in Section 2.3 below, but begin here with the more general example. For concreteness we will focus on the case  $\lambda > 0$  throughout this section.

In this setting we have that  $\max_j p_j = p_1 = z_r(a)^{-1}$ ,

$$\pi_j = \binom{n}{2} \frac{j^{-2a}}{z_r(a)^2} \left( 1 - \frac{j^{-a}}{z_r(a)} \right)^{n-2} \leq \frac{n^2 j^{-2a}}{2z_r(a)^2}, \quad \text{and} \quad \mu = \sum_{j=1}^r \pi_j \leq \sum_{j=1}^r \frac{n^2 j^{-2a}}{2z_r(a)^2} = \frac{n^2 z_r(2a)}{2z_r(a)^2}.$$

We further note that, for  $x \in \mathbb{R}$ ,

$$\mathbb{E}[P^x] = \sum_{j=1}^r p_j^{1+x} = \frac{z_r(a(1+x))}{z_r(a)^{1+x}},$$

and hence

$$d_\lambda = \frac{n}{4(2^\lambda - 1)^2} \left( \frac{\mathbb{E}[P^{-2\lambda}]}{\mathbb{E}[P^{-\lambda}]^2} - 1 \right) = \frac{n}{4(2^\lambda - 1)^2} \left( \frac{z_r(a(1-2\lambda))z_r(a)}{z_r(a(1-\lambda))^2} - 1 \right).$$

Finally, we have that

$$c_\lambda = \max_j \left| \frac{z_r(a)j^{a\lambda}}{z_r(a(1-\lambda))} - 1 \right| = \frac{z_r(a)r^{a\lambda}}{z_r(a(1-\lambda))} - 1. \quad (2.1)$$

Combining all these ingredients, we have that (1.4) gives us

$$d_K(\tilde{T}_\lambda, Z) \leq \frac{45n^3}{z_r(a)^4} \left( \frac{3n^2 z_r(3a)^2}{2z_r(a)^2} + \frac{nz_r(4a)}{4} + 6z_r(2a)^2 \right) + 18 \left( \frac{2.78n^2 c_\lambda z_r(2a)}{z_r(a)^2} \right)^{0.49/c_\lambda} + \frac{4n^3 z_r(2a)}{z_r(a)^2(4z_r(a) - (n+1))} + \frac{2.025n}{(2^\lambda - 1)^2} \left( \frac{z_r(a(1-2\lambda))z_r(a)}{z_r(a(1-\lambda))^2} - 1 \right), \quad (2.2)$$

where  $c_\lambda$  is given by (2.1) and for simplicity we have taken the first term in each of the minima in (1.4). As noted above, we do not lose much by doing this.

For illustration, consider the case where  $r$  is of order  $n^2$ , as in Example 3.1 of [13]. Of the four terms in the upper bound (2.2), the second and fourth rely on  $a$  being close to zero in order to be small, whereas the first and third will go to zero as  $n \rightarrow \infty$  for any fixed  $a < 1/2$ . These terms are of order  $O(n^{-1})$  and  $O(n^{2a-1})$ , respectively, up to logarithmic terms. The bound (2.2) continues to hold for fixed  $a \geq 1/2$ , but these terms will no longer go to zero as  $n$  grows and cannot be expected to be small. In particular, the third term will grow logarithmically in  $n$  at  $a = 1/2$  and as a power of  $n$  for  $a > 1/2$ .

### 2.3 Uniform allocations

We now specialise the example of Section 2.2 to the case where the allocation is uniform, i.e.,  $a = 0$  and so  $p_j = 1/r$  for each  $j = 1, \dots, r$ . Noting that  $z_r(0) = r$  and that  $c_\lambda = d_\lambda = 0$  in this case, we obtain the following corollary to Theorem 1:

**Corollary 6.** *Let  $n \geq 4$ ,  $r \geq 8$  and  $(n+1)/r < 4$ . Let  $T_\lambda$  be the power divergence statistic in the uniform case  $p_j = 1/r$  for  $j = 1, \dots, r$ , and  $Z$  have a Poisson distribution with mean  $\frac{1}{r} \binom{n}{2} (1 - 1/r)^{n-2}$ . Then, for all  $\lambda > -1$ ,*

$$d_K(\tilde{T}_\lambda, Z) \leq \frac{45n^3}{r^4} \left( \frac{3n^2}{2} + \frac{nr}{4} + 6r^2 \right) + \frac{4n^3}{r(4r - (n+1))}. \quad (2.3)$$

It is notable that the upper bound of Corollary 6 does not depend on  $\lambda$ , and holds for all  $\lambda > -1$ . It follows immediately from the upper bound (2.2) when  $\lambda > 0$ . Corollary 5 shows that (2.3) also holds for  $\lambda = 0$ . In the case  $\lambda \in (-1, 0)$ , we still have that  $c_\lambda = d_\lambda = 0$ , and (2.3) follows directly from Theorem 1.

We also note that this converges to zero, with rate  $O(1/n)$ , when  $r$  is of order  $n^2$ . This is not surprising in light of the conditions (i)–(iii) above for Poisson convergence due to [13]. In this setting of uniform allocations, condition (i) asserts that we must have  $n/r \rightarrow 0$ , condition (ii) is trivially satisfied since  $d_\lambda = 0$ , and condition (iii) asserts that that  $n^2/r$  has a finite and non-zero limit. We conjecture that  $O(1/n)$  is the correct rate of convergence here, noting that the convergence of the underlying multinomial statistics to Poisson, which yields the first term of the upper bound (2.3), is known to be of this order in total variation distance in this case; see the discussion on page 121 of [3]. Whether our Theorem 1 provides optimal or near-optimal rates of convergence in other settings is an open problem.

As well as a rate of convergence, Corollary 6 provides an explicit error bound, which we illustrate in Table 1 for various choices of  $n$  and  $r$ . As expected from the conditions for Poisson convergence here, this bound is useful only when  $r$  is very large compared to  $n$ . In other



| $r \backslash n$ | 5      | 10     | 20     | 30     | 40     | 50     |
|------------------|--------|--------|--------|--------|--------|--------|
| 300              | 0.3767 | —      | —      | —      | —      | —      |
| 500              | 0.1356 | —      | —      | —      | —      | —      |
| 700              | 0.0692 | 0.5534 | —      | —      | —      | —      |
| 1000             | 0.0339 | 0.2711 | —      | —      | —      | —      |
| 3000             | 0.0038 | 0.0301 | 0.2410 | 0.8134 | —      | —      |
| 5000             | 0.0014 | 0.0108 | 0.0867 | 0.2928 | 0.6940 | —      |
| 7000             | 0.0007 | 0.0055 | 0.0443 | 0.1494 | 0.3540 | 0.6915 |
| 10000            | 0.0003 | 0.0027 | 0.0217 | 0.0732 | 0.1735 | 0.3388 |

Table 1: The upper bound (2.3) to 4 d.p., for various values of  $n$  and  $r$ . The symbol ‘—’ indicates that the bound is greater than 1, and so not informative.

regimes a chi-square or Gaussian approximation may be more appropriate than our Poisson approximation. Nevertheless, in such a very sparse regime with very large  $r$ , the results of Table 1 show that our error bound can be small and a Poisson approximation of practical use. We also expect that, while the rate of convergence we obtain is reasonable, there is room for improvement in the leading constants in our error bound, which would improve its usefulness.

### 3 A Gaussian approximation theorem

Our Poisson approximation result may be used to derive a corresponding error bound in the Gaussian approximation of  $\tilde{T}_\lambda$ , as we show here. Letting  $\mu$  and  $Z$  be as in Theorem 1, and  $N \sim N(0, 1)$  have a standard Gaussian distribution, we use the triangle inequality to write

$$d_K \left( \frac{\tilde{T}_\lambda - \mu}{\sqrt{\mu}}, N \right) \leq d_K \left( \frac{\tilde{T}_\lambda - \mu}{\sqrt{\mu}}, \frac{Z - \mu}{\sqrt{\mu}} \right) + d_K \left( \frac{Z - \mu}{\sqrt{\mu}}, N \right) \leq d_K(\tilde{T}_\lambda, Z) + \frac{0.4748}{\sqrt{\mu}},$$

where the final inequality follows from a standard application of the Berry–Esseen inequality, with constant 0.4748 due to [15]. This gives the following Gaussian approximation result.

**Theorem 7.** *Let  $\lambda > -1$  and  $n \geq 4$ . Assume that  $p_j \leq 0.13$  and  $(n + 1)p_j < 4$  for each  $j = 1, \dots, r$ . Let  $\pi_j$ ,  $\mu$ ,  $c_\lambda$  and  $d_\lambda$  be as in Theorem 1, and let  $N \sim N(0, 1)$ . Then*

$$d_K \left( \frac{\tilde{T}_\lambda - \mu}{\sqrt{\mu}}, N \right) \leq 45 \min\{1, \mu^{-1}\} \left( 6n \left[ \sum_{j=1}^r p_j \pi_j \right]^2 + \sum_{j=1}^r \pi_j^2 + \frac{24\mu^2}{n} \right) + 18 (5.55\mu c_\lambda)^{0.49/c_\lambda} + \frac{8n\mu \max_j p_j}{4 - (n + 1) \max_j p_j} + \min \left\{ 8.1d_\lambda, 2.15 \left( \frac{d_\lambda}{\mu} \right)^{1/3} \right\} + \frac{0.4748}{\sqrt{\mu}}.$$

With the addition of this final term in our upper bound, we now work in a regime where  $\mu$  is large (so that this final term is small), but not so large that the other terms in our upper bound become significant. We are not aware of other explicit Gaussian error bounds for the

family of power divergence statistics with which we can compare our Theorem 7, but we will use the remainder of this section to compare our bound with conditions under which convergence to a limiting Gaussian distribution is known to hold. For concreteness, we will do this in the particular example of uniform allocations, as considered in Section 2.3 above. Given that our bound is derived via a Poisson approximation result, it is perhaps not surprising that we will see that conditions under which our upper bound goes to zero are sub-optimal. Nevertheless, we stress again that we are not aware of other bounds that cover the family of power divergence statistics, and that our result gives explicit error bounds in a metric of practical importance.

In the uniform allocation setting with  $p_j = 1/r$ , we have  $\pi_j = \binom{n}{2} r^{-2} (1 - r^{-1})^{n-2}$  and  $c_\lambda = d_\lambda = 0$ . The upper bound of Theorem 7 thus gives

$$d_K \left( \frac{\tilde{T}_\lambda - \mu}{\sqrt{\mu}}, N \right) \leq \frac{90n^3}{(n-1)^2 r^3} \left(1 - \frac{1}{r}\right)^{2-n} \left( \frac{3n^2}{2} + \frac{nr}{4} + 6r^2 \right) + \frac{4n^3}{r(4r - (n+1))} + \frac{0.4748}{n-1} \sqrt{2r \left(1 - \frac{1}{r}\right)^{2-n}}, \quad (3.1)$$

where, for simplicity of presentation, we write  $\mu \leq \frac{n^2}{2r}$  and

$$\min\{1, \mu^{-1}\} \leq \mu^{-1} \leq \frac{2r}{(n-1)^2} \left(1 - \frac{1}{r}\right)^{2-n}.$$

Recall that in Corollary 6 we wrote  $\min\{1, \mu^{-1}\} \leq 1$ , since there we were not in a regime where we expect  $\mu$  to be large, hence the differences in the first terms of these two upper bounds for this example.

Considering  $r$  of order  $n^\alpha$ , the upper bound of (3.1) goes to zero as  $n \rightarrow \infty$  if  $\alpha \in (3/2, 2)$ . This can be compared with the rate of convergence in the Wasserstein distance (based on smooth test functions, unlike the Kolmogorov distance we use here) given by Döbler [6] for the special case of Pearson's statistic. In his Example 4.6, Döbler shows that we expect a Gaussian limit for a suitably normalised Pearson's statistic if  $\alpha < 2$ . While our bound is clearly significantly inferior in this respect, it applies to a wider family of statistics and gives an explicit error bound, not only a rate of convergence.

In the more general setting of the family of power divergence statistics with  $\lambda > -1$ , Cressie and Read [5] show in their Corollary 2.2 that in the uniform allocation setting with  $n/r$  converging to a finite and non-zero limit as  $n \rightarrow \infty$  a suitably normalised version of  $T_\lambda$  has a Gaussian limiting distribution; this fact is not reflected in our upper bound (3.1). We leave the problem of giving improved error bounds in Kolmogorov distance for the family of power divergence statistics open for future work.

## 4 Proof of Theorem 1

We use this section to prove our main result, the upper bound (1.4) of Theorem 1. Our starting point is the following representation of  $T_\lambda$ , given in Proposition 2.4 of [13]:

$$T_\lambda = \frac{1}{n^\lambda} \sum_{j=1}^r p_j^{-\lambda} g_\lambda(N_j) + n \sum_{k=1}^n p_{X_k} g_\lambda([np_{X_k}]^{-1}), \quad (4.1)$$

where  $g_\lambda$  is given by (1.1), and we define the random variable  $X$  by  $\mathbb{P}(X = j) = p_j$  and let  $X, X_1, X_2, \dots, X_n$  be an IID sequence, so that we can write  $N_j = \sum_{k=1}^n I(X_k = j)$ . The representation (4.1) follows from the observation that, for any  $a, b \geq 0$ , we have  $g_\lambda(ab) = b^{\lambda+1}g_\lambda(a) + ag_\lambda(b)$ . Proposition 2.4 of [13] is stated only for  $\lambda \geq 0$ , but the same argument and result hold for any  $\lambda > -1$ . Combining (4.1) with (1.3), we write  $\tilde{T}_\lambda = W_\lambda + R_\lambda$ , where

$$W_\lambda = \frac{1}{g_\lambda(2)\mathbb{E}[P^{-\lambda}]} \sum_{j=1}^r p_j^{-\lambda} g_\lambda(N_j), \quad \text{and}$$

$$R_\lambda = \frac{n^{\lambda+1}}{g_\lambda(2)\mathbb{E}[P^{-\lambda}]} \left( \sum_{k=1}^n p_{X_k} g_\lambda([np_{X_k}]^{-1}) - n\mathbb{E}[Pg_\lambda([nP]^{-1})] \right).$$

Now, for some  $\beta_0 > 0$  to be determined later, we define the event  $E_0 = \{|R_\lambda| \leq \beta_0\}$  with  $\delta_0 = \delta_0(\beta_0) = \mathbb{P}(E_0^c)$ . Chebyshev's inequality gives

$$\begin{aligned} \delta_0 &= \mathbb{P} \left( \left| \sum_{k=1}^n p_{X_k} g_\lambda([np_{X_k}]^{-1}) - n\mathbb{E}[Pg_\lambda([nP]^{-1})] \right| > \beta_0 |g_\lambda(2)|\mathbb{E}[P^{-\lambda}]n^{-(\lambda+1)} \right) \\ &\leq \frac{\text{Var}(Pg_\lambda([nP]^{-1}))n^{2\lambda+3}}{\beta_0^2 g_\lambda(2)^2 \mathbb{E}[P^{-\lambda}]^2} = \frac{d_\lambda}{\beta_0^2}, \end{aligned} \quad (4.2)$$

where the final equality follows from the definition of  $g_\lambda$ , checking the cases  $\lambda = 0$  and  $\lambda \neq 0$  separately.

Using the definition (1.2), we have  $d_K(\tilde{T}_\lambda, Z) = \sup_{y \in \mathbb{R}} |\mathbb{P}(\tilde{T}_\lambda \geq y - \beta_0) - \mathbb{P}(Z \geq y - \beta_0)|$ , where the triangle inequality gives

$$\begin{aligned} |\mathbb{P}(\tilde{T}_\lambda \geq y - \beta_0) - \mathbb{P}(Z \geq y - \beta_0)| &\leq |\mathbb{P}(W_\lambda \geq y - \beta_0) - \mathbb{P}(Z \geq y - \beta_0)| + \mathbb{P}(\tilde{T}_\lambda \geq y - \beta_0, E_0^c) \\ &\quad + |\mathbb{P}(\tilde{T}_\lambda \geq y - \beta_0, E_0) - \mathbb{P}(W_\lambda \geq y - \beta_0)| \\ &\leq d_K(W_\lambda, Z) + \delta_0 + |\mathbb{P}(\tilde{T}_\lambda \geq y - \beta_0, E_0) - \mathbb{P}(W_\lambda \geq y - \beta_0)|. \end{aligned} \quad (4.3)$$

For this final term, we write

$$\begin{aligned} &|\mathbb{P}(\tilde{T}_\lambda \geq y - \beta_0, E_0) - \mathbb{P}(W_\lambda \geq y - \beta_0)| \\ &\leq |\mathbb{P}(\tilde{T}_\lambda \geq y - \beta_0, E_0) - \mathbb{P}(W_\lambda \geq y - \beta_0, E_0)| + \mathbb{P}(W_\lambda \geq y - \beta_0, E_0^c) \\ &\leq |\mathbb{P}(W_\lambda \geq y - R_\lambda - \beta_0, |R_\lambda| \leq \beta_0) - \mathbb{P}(W_\lambda \geq y - \beta_0, |R_\lambda| \leq \beta_0)| + \delta_0 \\ &= \mathbb{P}(y - \beta_0 - R_\lambda \leq W_\lambda < y - \beta_0, 0 \leq R_\lambda \leq \beta_0) \\ &\quad + \mathbb{P}(y - \beta_0 \leq W_\lambda < y - \beta_0 - R_\lambda, -\beta_0 \leq R_\lambda < 0) + \delta_0 \\ &\leq \mathbb{P}(y - 2\beta_0 \leq W_\lambda < y) + \delta_0 \\ &\leq 2d_K(W_\lambda, Z) + \lfloor 2\beta_0 \rfloor \sup_x \mathbb{P}(Z = x) + \delta_0, \end{aligned} \quad (4.4)$$

where  $\lfloor \cdot \rfloor$  is the floor function.

Combining (4.2)–(4.4), we thus have

$$d_K(\tilde{T}_\lambda, Z) \leq 3d_K(W_\lambda, Z) + \frac{2d_\lambda}{\beta_0^2} + \lfloor 2\beta_0 \rfloor \sup_x \mathbb{P}(Z = x), \quad (4.5)$$

for any  $\beta_0 > 0$ . We bound the final two terms on the right-hand side of (4.5) in two different ways. Firstly, choosing  $\beta_0 \approx 0.497$ , so that  $1/\beta_0^2 = 4.05$  and  $\lfloor 2\beta_0 \rfloor = 0$ , we have  $d_K(\tilde{T}_\lambda, Z) \leq 3d_K(W_\lambda, Z) + 8.1d_\lambda$ . Secondly, we may use Proposition A.2.7 of [3] to get that  $\mathbb{P}(Z = x) \leq (2e\mu)^{-1/2}$ , so that

$$d_K(\tilde{T}_\lambda, Z) \leq 3d_K(W_\lambda, Z) + 2 \left( \frac{d_\lambda}{\beta_0^2} + \frac{\beta_0}{\sqrt{2e\mu}} \right).$$

We now choose  $\beta_0 = (2d_\lambda\sqrt{2e\mu})^{1/3}$  to obtain  $d_K(\tilde{T}_\lambda, Z) \leq 3d_K(W_\lambda, Z) + 2.15 \left( \frac{d_\lambda}{\mu} \right)^{1/3}$ , and hence

$$d_K(\tilde{T}_\lambda, Z) \leq 3d_K(W_\lambda, Z) + \min \left\{ 8.1d_\lambda, 2.15 \left( \frac{d_\lambda}{\mu} \right)^{1/3} \right\}. \quad (4.6)$$

Now, noting that  $g_\lambda(0) = g_\lambda(1) = 0$ , we may write  $W_\lambda = V_\lambda + R'_\lambda$ , where

$$V_\lambda = \frac{1}{\mathbb{E}[P^{-\lambda}]} \sum_{j: N_j=2} p_j^{-\lambda}, \quad \text{and} \quad R'_\lambda = \frac{1}{g_\lambda(2)\mathbb{E}[P^{-\lambda}]} \sum_{j: N_j \geq 3} p_j^{-\lambda} g_\lambda(N_j).$$

We now argue similarly to above: for some  $\beta_1 > 0$  to be determined later, we define the event  $E_1 = \{R'_\lambda \leq \beta_1\}$  with  $\delta_1 = \delta_1(\beta_1) = \mathbb{P}(E_1^c)$ , noting that  $R'_\lambda \geq 0$  almost surely for any  $\lambda > -1$ . We have

$$\begin{aligned} \delta_1 &= \mathbb{P} \left( \frac{1}{g_\lambda(2)\mathbb{E}[P^{-\lambda}]} \sum_{j: N_j \geq 3} p_j^{-\lambda} g_\lambda(N_j) > \beta_1 \right) \leq \mathbb{P}(\exists j \in \{1, \dots, r\} \text{ such that } N_j \geq 3) \\ &\leq \sum_{j=1}^r \mathbb{P}(N_j \geq 3) \leq 4 \binom{n}{3} \sum_{j=1}^r \frac{p_j^3 (1-p_j)^{n-2}}{4 - (n+1)p_j} \leq \frac{4n\mu \max_j p_j}{3[4 - (n+1) \max_j p_j]}, \end{aligned} \quad (4.7)$$

where the penultimate inequality uses Proposition A.2.5(ii) of [3] and the assumption that  $(n+1)p_j < 4$  for each  $j$ , since  $N_j \sim \text{Bin}(n, p_j)$  has a binomial distribution.

We now write  $d_K(W_\lambda, Z) = \sup_{y \in \mathbb{R}} |\mathbb{P}(W_\lambda \geq y - \beta_1) - \mathbb{P}(Z \geq y - \beta_1)|$ , where the triangle inequality gives

$$\begin{aligned} |\mathbb{P}(W_\lambda \geq y - \beta_1) - \mathbb{P}(Z \geq y - \beta_1)| &\leq |\mathbb{P}(V_\lambda \geq y - \beta_1) - \mathbb{P}(Z \geq y - \beta_1)| + \mathbb{P}(W_\lambda \geq y - \beta_1, E_1^c) \\ &\quad + |\mathbb{P}(W_\lambda \geq y - \beta_1, E_1) - \mathbb{P}(V_\lambda \geq y - \beta_1)| \\ &\leq d_K(V_\lambda, Z) + \delta_1 + |\mathbb{P}(W_\lambda \geq y - \beta_1, E_1) - \mathbb{P}(V_\lambda \geq y - \beta_1)|. \end{aligned} \quad (4.8)$$

For this final term, we write

$$\begin{aligned} &|\mathbb{P}(W_\lambda \geq y - \beta_1, E_1) - \mathbb{P}(V_\lambda \geq y - \beta_1)| \\ &\leq |\mathbb{P}(W_\lambda \geq y - \beta_1, E_1) - \mathbb{P}(V_\lambda \geq y - \beta_1, E_1)| + \mathbb{P}(V_\lambda \geq y - \beta_1, E_1^c) \\ &\leq |\mathbb{P}(V_\lambda \geq y - R'_\lambda - \beta_1, R'_\lambda \leq \beta_1) - \mathbb{P}(V_\lambda \geq y - \beta_1, R'_\lambda \leq \beta_1)| + \delta_1 \\ &= \mathbb{P}(y - R'_\lambda - \beta_1 \leq V_\lambda < y - \beta_1, R'_\lambda \leq \beta_1) + \delta_1 \\ &\leq \mathbb{P}(y - 2\beta_1 \leq V_\lambda < y - \beta_1) + \delta_1 \\ &\leq 2d_K(V_\lambda, Z) + \delta_1, \end{aligned} \quad (4.9)$$

for any  $\beta_1 < 1/2$ .

Combining (4.7)–(4.9), we thus have

$$d_K(W_\lambda, Z) \leq 3d_K(V_\lambda, Z) + \frac{8n\mu \max_j p_j}{3[4 - (n+1) \max_j p_j]}. \quad (4.10)$$

We now write  $V_\lambda = U_\lambda + R''_\lambda$ , where

$$U_\lambda = \sum_{j=1}^r I(N_j = 2), \quad \text{and} \quad R''_\lambda = \sum_{j=1}^r \left( \frac{1}{p_j^\lambda \mathbb{E}[P^{-\lambda}]} - 1 \right) I(N_j = 2),$$

and argue similarly again. For some  $\beta_2 > 0$  to be determined later, we define the event  $E_2 = \{|R''_\lambda| \leq \beta_2\}$  with  $\delta_2 = \delta_2(\beta_2) = \mathbb{P}(E_2^c)$ . We have

$$\delta_2 = \mathbb{P} \left( \left| \sum_{j=1}^r \left( \frac{1}{p_j^\lambda \mathbb{E}[P^{-\lambda}]} - 1 \right) I(N_j = 2) \right| > \beta_2 \right) \leq \mathbb{P}(c_\lambda U_\lambda > \beta_2) \leq \mathbb{P}(Z > \beta_2/c_\lambda) + d_K(U_\lambda, Z), \quad (4.11)$$

where Bennett's inequality gives

$$\begin{aligned} \mathbb{P}(Z > \beta_2/c_\lambda) &\leq \exp \left\{ -\mu \left[ \frac{\beta_2}{\mu c_\lambda} \log \left( \frac{\beta_2}{\mu c_\lambda} \right) - \frac{\beta_2}{\mu c_\lambda} + 1 \right] \right\} \\ &= \left[ \left( \frac{\beta_2}{\mu c_\lambda} \right)^{\beta_2/\mu c_\lambda} \exp \left\{ 1 - \frac{\beta_2}{\mu c_\lambda} \right\} \right]^{-\mu} \leq \left( \frac{e\mu c_\lambda}{\beta_2} \right)^{\beta_2/c_\lambda}, \end{aligned} \quad (4.12)$$

and  $\mathbb{P}(c_\lambda U_\lambda > \beta_2)$  should be interpreted as zero if  $c_\lambda = 0$ .

Writing  $d_K(V_\lambda, Z) = \sup_{y \in \mathbb{R}} |\mathbb{P}(V_\lambda \geq y - \beta_2) - \mathbb{P}(Z \geq y - \beta_2)|$ , the triangle inequality gives

$$\begin{aligned} |\mathbb{P}(V_\lambda \geq y - \beta_2) - \mathbb{P}(Z \geq y - \beta_2)| &\leq |\mathbb{P}(U_\lambda \geq y - \beta_2) - \mathbb{P}(Z \geq y - \beta_2)| + \mathbb{P}(V_\lambda \geq y - \beta_2, E_2^c) \\ &\quad + |\mathbb{P}(V_\lambda \geq y - \beta_2, E_2) - \mathbb{P}(U_\lambda \geq y - \beta_2)| \\ &\leq d_K(U_\lambda, Z) + \delta_2 + |\mathbb{P}(V_\lambda \geq y - \beta_2, E_2) - \mathbb{P}(U_\lambda \geq y - \beta_2)|. \end{aligned} \quad (4.13)$$

For the final term here, we have

$$\begin{aligned} &|\mathbb{P}(V_\lambda \geq y - \beta_2, E_2) - \mathbb{P}(U_\lambda \geq y - \beta_2)| \\ &\leq |\mathbb{P}(V_\lambda \geq y - \beta_2, E_2) - \mathbb{P}(U_\lambda \geq y - \beta_2, E_2)| + \mathbb{P}(U_\lambda \geq y - \beta_2, E_2^c) \\ &\leq |\mathbb{P}(U_\lambda \geq y - R''_\lambda - \beta_2, |R''_\lambda| \leq \beta_2) - \mathbb{P}(U_\lambda \geq y - \beta_2, |R''_\lambda| \leq \beta_2)| + \delta_2 \\ &= \mathbb{P}(y - \beta_2 - R''_\lambda \leq U_\lambda < y - \beta_2, 0 \leq R''_\lambda \leq \beta_2) \\ &\quad + \mathbb{P}(y - \beta_2 \leq U_\lambda < y - \beta_2 - R''_\lambda, -\beta_2 \leq R''_\lambda < 0) + \delta_2 \\ &\leq \mathbb{P}(y - 2\beta_2 \leq U_\lambda < y) + \delta_2 \\ &\leq 2d_K(U_\lambda, Z) + [2\beta_2] \sup_x \mathbb{P}(Z = x) + \delta_2. \end{aligned} \quad (4.14)$$

Combining (4.11)–(4.14) and choosing  $\beta_2 = 0.49$ , we have

$$d_K(V_\lambda, Z) \leq 5d_K(U_\lambda, Z) + 2(5.55\mu c_\lambda)^{0.49/c_\lambda}. \quad (4.15)$$

To bound  $d_K(U_\lambda, Z)$  we may use Theorem 6.F of [3], which gives an error bound in the Poisson approximation of the multinomial occupancy statistic  $U_\lambda$ . Under the assumptions that  $n \geq 4$  and  $p_j \leq 1 - \sqrt{3}/2 \approx 0.13397$ , that result gives us that

$$d_K(U_\lambda, Z) \leq \min \{1, \mu^{-1}\} \left( 6n \left[ \sum_{j=1}^r p_j \pi_j \right]^2 + \sum_{j=1}^r \pi_j^2 + \frac{24\mu^2}{n} \right). \quad (4.16)$$

Combining (4.6), (4.10), (4.15) and (4.16) completes the proof of Theorem 1.

## 5 A generalisation of Theorem 1

We note finally that our proof of Theorem 1 may be easily adapted to give a Poisson approximation bound in the following slightly more general setting: let  $T = W + R$ , where

$$W = \frac{1}{g(m)} \sum_{j=1}^r h(p_j) g(N_j) \geq 0 \text{ almost surely,}$$

$\mathbb{E}[R] = 0$ , and  $m \in \{0, 1, \dots\}$  is such that  $g(0) = g(1) = \dots = g(m-1) = 0$ . The power divergence statistic  $\tilde{T}_\lambda$  we consider here is the special case with  $m = 2$ ,  $g = g_\lambda$ , and  $h$  given by  $h(p_j) = (p_j \mathbb{E}[P^{-\lambda}])^{-1}$ . Following the same argument as in the proof of our Theorem 1 but using the full generality of Theorem 6.F and Proposition A.2.5(ii) of [3] which we applied only for  $m = 2$  in the above, we have the following result.

**Theorem 8.** *Let  $T$  be as above and  $n \geq 2m$ . Assume that  $(1-p_j)^m \geq 3/4$  and  $(n+1)p_j < m+2$  for each  $j = 1, \dots, r$ . Let  $\pi_j = \binom{n}{m} p_j^m (1-p_j)^{n-m}$  and  $\mu = \pi_1 + \dots + \pi_r$ , and let  $Z$  have a Poisson distribution with mean  $\mu$ . Then*

$$d_K(T, Z) \leq 45 \min\{1, \mu^{-1}\} \left( 6n \left[ \sum_{j=1}^r p_j \pi_j \right]^2 + \sum_{j=1}^r \pi_j^2 + \frac{6m^2 \mu^2}{n} \right) \\ + 18(5.55\mu c)^{0.49/c} + \frac{6(m+2)n\mu \max_j p_j}{(m+1)(m+2-(n+1)\max_j p_j)} + 8.1 \text{Var}(R),$$

where  $c = \max_j |h(p_j) - 1|$ .

Recalling the conditions (i), (ii), (iv) and (v) given in Section 1 under which the upper bound of Theorem 1 goes to zero as  $n \rightarrow \infty$ , we note the analogous conditions here: we require  $n \max_j p_j \rightarrow 0$ ,  $\text{Var}(R) \rightarrow 0$ ,  $n^m \mathbb{E}[P^{m-1}] \rightarrow \eta$  for some  $0 < \eta < \infty$ , and  $c \rightarrow 0$ .

**Acknowledgement:** The author thanks Robert Gaunt for several helpful comments, discussions and suggestions. Thanks are also due to an anonymous referee for their careful reading of the manuscript and suggestions which improved its presentation.

## References

- [1] Z. A. Assylebekov (2010). Convergence rate of multinomial goodness-of-fit statistics to chi-square distribution. *Hiroshima Math. J.* **40**: 115–131.
- [2] Z. A. Assylebekov, V. N. Zubov and V. V. Ulyanov (2011). On approximating some statistics of goodness-of-fit tests in the case of three-dimensional discrete data. *Sib. Math. J.* **52**(4): 571–584.
- [3] A. D. Barbour, L. Holst and S. Janson (1992). *Poisson Approximation*. Oxford University Press, Oxford.
- [4] S. X. Chen and Y.-L. Qin (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *Ann. Statist.* **38**(2): 808–835.
- [5] N. Cressie and T. R. C. Read (1984). Multinomial goodness-of-fit tests. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **46**(3): 440–464.
- [6] C. Döbler (2023). Normal approximation via non-linear exchangeable pairs. *ALEA Lat. Am. J. Probab. Math. Stat.* **20**: 167–224.
- [7] R. E. Gaunt (2022). Bounds for the chi-square approximation of the power divergence family of statistics. *J. Appl. Probab.* **59**(4): 1059–1080.
- [8] J. Janková, R. D. Shah, P. Bühlmann and R. J. Samworth (2020). Goodness-of-fit testing in high dimensional generalized linear models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **82**(3): 773–795.
- [9] I. Kim, S. Balakrishnan and L. Wasserman (2020). Robust multivariate nonparametric tests via projection averaging. *Ann. Statist.* **48**(6): 3417–3441.
- [10] C. Morris (1975). Central limit theorems for multinomial sums. *Ann. Statist.* **3**(1): 165–188.
- [11] N. Puchkin and V. Ulyanov (2023). Inference via randomized test statistics. *Ann. Inst. Henri Poincaré Probab. Stat.* **59**(3): 1508–1529.
- [12] T. R. C. Read and N. A. C. Cressie (1988). *Goodness-of-Fit Statistics for Discrete Multivariate Data*. Springer, New York.
- [13] G. Rempala and J. Wesolowski (2023). Poisson limit theorems for the Cressie–Read statistics. *J. Statist. Plann. Inference* **223**: 15–32.
- [14] B. Roos (2003). Improvements in the Poisson approximation of mixed Poisson distributions. *J. Statist. Plann. Inference* **113**: 467–483.
- [15] I. Shevtsova (2011). On the absolute constants in the Berry–Esseen type inequalities for identically distributed summands. Preprint, available at [arXiv:1111.6554](https://arxiv.org/abs/1111.6554).
- [16] G. P. Steck (1957). Limit theorems for conditional distributions. *Univ. California Publ. Statist.* **2**(12): 237–284.

- [17] V. V. Ulyanov and V. N. Zubov (2009). Refinement on the convergence of one family of goodness-of-fit statistics to chi-squared distribution. *Hiroshima Math. J.* **39**: 133–161.