



HAL
open science

Testing convexity of a discrete distribution

Fadoua Balabdaoui, Cécile Durot, François Koladjo

► **To cite this version:**

Fadoua Balabdaoui, Cécile Durot, François Koladjo. Testing convexity of a discrete distribution. 2018. hal-01693857

HAL Id: hal-01693857

<https://hal.science/hal-01693857>

Preprint submitted on 20 Feb 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Testing convexity of a discrete distribution

Fadoua Balabdaoui^{1,2}, Cécile Durot^{3*} and Babagnidé François Koladjo⁴

¹Seminar für Statistik, E.T.H. Zürich, 8092, Zürich, Schweiz

²CEREMADE, Université Paris-Dauphine, PSL Research University, 75775, Paris, France

³Modal'x, Université Paris Nanterre, F-92001, Nanterre, France

⁴ENSPD, Université de Parakou, BP 55 Tchaourou, Bénin

Abstract

Based on the convex least-squares estimator, we propose two different procedures for testing convexity of a probability mass function supported on \mathbb{N} with an unknown finite support. The procedures are shown to be asymptotically calibrated.

1 The testing problem

Modeling count data is an important issue in statistical research, see e.g. Gómez-Déniz and Calderín-Ojeda (2011). A popular parametric model for such data is the Poisson model. While non-parametric extensions are conceivable, those that only assume a shape constraint of the underlying probability mass function (pmf) may offer more flexibility. Recent papers on estimating a pmf under a shape constraint are Jankowski and Wellner (2009), Durot et al. (2013, 2015), Balabdaoui et al. (2013), Giguelay (2016), Chee and Wang (2016). In

*Corresponding author. Email address: cecile.durot@gmail.com

any case, it is sensible to validate the chosen model using a goodness-of-fit test. Goodness-of-fit tests to validate a model connected to the Poisson distribution are given in Karlis and Xekalaki (2000), Meintanis and Nikitin (2008), Ledwina and Wyłupek (2016) and the references therein. However, to the best of our knowledge, no goodness-of-fit test has been proposed to validate an assumed shape constraint for discrete data.

Motivated by the biological application in Durot et al. (2015), where the number of species is estimated assuming a convex abundance distribution, we develop here a goodness-of-fit test for convexity of the underlying pmf on \mathbb{N} . To the best of our knowledge, this is the first attempt to build a convexity test for count data. Precisely, based on i.i.d. observations X_1, \dots, X_n from some pmf p_0 on \mathbb{N} , we test the null hypothesis H_0 : “ p_0 is convex on \mathbb{N} ” (i.e. $p_0(k+1) - p_0(k) \geq p_0(k) - p_0(k-1)$ for all integers $k \geq 1$) versus H_1 : “ p_0 is not convex”. The test is based on the convex least-squares estimator $\hat{p}_n := \operatorname{argmin}_{p \in \mathcal{C}_1} \|p_n - p\|$, where \mathcal{C}_1 is the set of all convex pmfs on \mathbb{N} , $\|q\|^2 = \sum_{j \in \mathbb{N}} (q(j))^2$ for any sequence $q = \{q(j), j \in \mathbb{N}\}$, and $p_n(j) = n^{-1} \sum_{i=1}^n \mathbb{1}_{\{X_i=j\}}$, $j \in \mathbb{N}$, is the empirical pmf. It is proved in Durot et al. (2013, Sections 2.1 to 2.3) that \hat{p}_n exists, is unique, and can be implemented with an appropriate algorithm. We reject H_0 if $\{T_n > t_{\alpha,n}\}$ where $T_n = \sqrt{n} \|p_n - \hat{p}_n\|$ and $t_{\alpha,n}$ is an appropriate quantile, chosen in such a way that the test has asymptotic level α .

In the sequel, we assume that p_0 has a finite support in $\{0, \dots, S\}$ with an unknown integer $S > 0$ and we consider two different constructions of $t_{\alpha,n}$. First, we define $t_{\alpha,n}$ as the $(1-\alpha)$ -quantile of a random variable whose limiting distribution coincides with the limiting distribution of T_n under H_0 . Next, we calibrate the test under a least favorable hypothesis (when the true pmf is triangular). Theoretical justification requires knowledge of the limiting distribution of T_n under H_0 . This needs some notation. For all $p = \{p(j), j \in \mathbb{N}\}$ and $k \in \mathbb{N} \setminus \{0\}$ we set $\Delta p(k) = p(k+1) - 2p(k) + p(k-1)$ (hence p is convex on \mathbb{N} iff $\Delta p(k) \geq 0$ for all k) and a given $k \in \mathbb{N} \setminus \{0\}$ is called a knot of p if $\Delta p(k) > 0$. For all $s > 0$

and $u = (u(0), \dots, u(s+1)) \in \mathbb{R}^{s+2}$, we set $\|u\|_s^2 = \sum_{k=0}^{s+1} (u(k))^2$. Also, let g_0 be a $(S+2)$ centered Gaussian vector whose dispersion matrix Γ_0 has component $(i+1, j+1)$ equal to $\mathbb{1}_{\{i=j\}}p_0(i) - p_0(i)p_0(j)$ for all $i, j = 0, \dots, S+1$, and \widehat{g}_0 the minimizer of $\|g - g_0\|_S$ over

$$g \in \mathcal{K}_0 := \left\{ g = (g(0), \dots, g(S+1)) \in \mathbb{R}^{S+2} : \Delta g(k) \geq 0 \text{ for all } k \in \{1, \dots, S\} \right. \\ \left. \text{such that } \Delta p_0(k) = 0 \right\}.$$

Existence, uniqueness and characterization of \widehat{g}_0 are given in Balabdaoui et al. (2017, Theorem 3.1). The asymptotic distribution of T_n under H_0 is given below.

Theorem 1.1. (i) *The distribution function of $\widehat{T}_0 := \|\widehat{g}_0 - g_0\|_S$ is continuous on $(0, \infty)$.*
(ii) *Under H_0 , $T_n \xrightarrow{d} \widehat{T}_0$ and $\sup_{t \geq 0} |P(T_n \leq t) - P(\widehat{T}_0 \leq t)| \rightarrow 0$, as $n \rightarrow \infty$.*

2 Calibrating by estimating the limiting distribution

Here, we build a random variable that weakly converges to \widehat{T}_0 and can be approximated via Monte-Carlo simulations. Let $S_n = \max\{X_1, \dots, X_n\}$, and let g_n be a random vector which, conditionally on (X_1, \dots, X_n) , is a S_n+2 centered Gaussian vector whose dispersion matrix Γ_n has component $(i+1, j+1)$ equal to $\mathbb{1}_{\{i=j\}}p_n(i) - p_n(i)p_n(j)$ for all $i, j = 0, \dots, S_n+1$. Now, let $\widehat{g}_n = \operatorname{argmin}_{g \in \mathcal{K}_n} \|g - g_n\|_{S_n}$, the least squares projection of g_n on \mathcal{K}_n , where \mathcal{K}_n “approaches” \mathcal{K}_0 as $n \rightarrow \infty$:

$$\mathcal{K}_n = \left\{ g = (g(0), \dots, g(S_n+1)) \in \mathbb{R}^{S_n+2} : \Delta g(k) \geq 0 \text{ for all } k \in \{1, \dots, S_n\} \right. \\ \left. \text{such that } \Delta \widehat{p}_n(k) \leq v_n \right\}$$

for an appropriate positive sequence $(v_n)_n$. Choosing $v_n = 0$ would make \mathcal{K}_n to be the largest possible and hence $\|\widehat{g}_n - g_n\|_{S_n}$ the smallest possible; this distance would be stochas-

tically smaller than the actual limit of T_n , yielding a large probability of rejection. In fact, choosing $v_n = 0$ amounts to estimate the knots of p_0 by those of \hat{p}_n , which is not desirable since \hat{p}_n has typically more knots than p_0 . The conditions required on v_n are given below.

Theorem 2.1. *Let g_n , \mathcal{K}_n , and \hat{g}_n be as above, and take $v_n > 0$ such that $v_n = o(1)$ and $v_n \gg n^{-1/2}$. (i) Then, \hat{g}_n uniquely exists, both \hat{g}_n and $\hat{T}_n := \|\hat{g}_n - g_n\|_{S_n}$ are measurable. (ii) Under H_0 , conditionally on X_1, \dots, X_n we have $\hat{T}_n \xrightarrow{d} \hat{T}_0$ in probability as $n \rightarrow \infty$.*

By (i) in Theorem 1.1, the conditional convergence $\hat{T}_n \xrightarrow{d} \hat{T}_0$ in probability means that

$$\sup_{t \in \mathbb{R}} \left| P(\hat{T}_n \leq t | X_1, \dots, X_n) - P(\hat{T}_0 \leq t) \right| = o_p(1). \quad (2.1)$$

In Balabdaoui et al. (2017, Theorem 3.3) it is shown that $\lim_{n \rightarrow \infty} P(\mathcal{K}_n \neq \mathcal{K}_0) = 0$ for any $(v_n)_n$ satisfying the conditions of the theorem. The intuition behind is as follows: when k is a knot of p_0 and $\Delta \hat{p}_n(k) \leq v_n$, then $\sqrt{n}(\Delta \hat{p}_n(k) - \Delta p_0(k)) < -\sqrt{n}\epsilon_0$ for some $\epsilon_0 > 0$ and n large enough. Weak convergence of \hat{p}_n to p_0 implies that this happens with probability converging to zero. In case k is not a knot; i.e., $\Delta p_0(k) = 0$ such that $\Delta \hat{p}_n(k) > v_n$ then $\sqrt{n}\Delta \hat{p}_n(k) > \sqrt{nv_n} \rightarrow \infty$, which again happens with decreasing probability. We now state the main result of the section, which is proven in the supplement.

Theorem 2.2. *Let \hat{T}_n as in Theorem 2.1. Let $\alpha \in (0, 1)$ and $t_{\alpha, n}$ the conditional $(1 - \alpha)$ -quantile of \hat{T}_n given X_1, \dots, X_n . Under H_0 , $\limsup_{n \rightarrow \infty} P(T_n > t_{\alpha, n}) \leq \alpha$.*

Hence, the test is asymptotically calibrated. In fact, it can be shown that the asymptotic Type I error is precisely equal to α for appropriate range of α , i.e. $\lim_{n \rightarrow \infty} P(T_n > t_{\alpha, n}) = \alpha$, see the supplementary material. An approximative value of $t_{\alpha, n}$ can be computed using Monte-Carlo simulations as follows. Having observed X_1, \dots, X_n , draw independent sequences $(Z_i^{(b)})_{0 \leq i \leq S_n+1}$ for $b \in \{1, \dots, B\}$, of i.i.d. $\mathcal{N}(0, 1)$ variables $Z_i^{(b)}$, where $B > 0$

is the number of Monte-Carlo runs. For all b , compute $g_n^{(b)} = \Gamma_n^{1/2}(Z_0^{(b)}, \dots, Z_{S_n+1}^{(b)})^T$ and $\widehat{g}_n^{(b)}$ the minimizer of $\|g_n^{(b)} - g\|_{S_n}$ over \mathcal{K}_n using Dykstra's algorithm, see Balabdaoui et al. (2017). If $(s_j)_j$ is the sequence of successive knots of \widehat{p}_n such that $\Delta\widehat{p}_n(s_j) > v_n$, then the algorithm performs iterative projections on the cones $\{g \in \mathbb{R}^{S_n+2} : \Delta g(k) \geq 0 \text{ for all } k \in \{s_j + 1, \dots, s_{j+1} - 1\}\}$, whose intersection is precisely \mathcal{K}_n . See Dykstra (1983) for more details and a proof of convergence. Then, $t_{\alpha,n}$ can be approximated by the $(1 - \alpha)$ -quantile of the empirical distribution corresponding to $\|g_n^{(b)} - \widehat{g}_n^{(b)}\|_{S_n}$, with $b \in \{1, \dots, B\}$.

3 Calibrating under the least favorable hypothesis

We consider below an alternative calibration that is easier to implement than the first one since it does not involve a sequence (v_n) . In what follows we denote by \mathcal{T}_a the triangular pmf supported on $\{0, \dots, a - 1\}$ for a given integer $a \geq 1$; i.e., $\mathcal{T}_a(i) = 2(a - i)_+ [a(a + 1)]^{-1}$. Consider $\widetilde{\mathcal{K}}_0$ the set of all vectors $g = (g(0), \dots, g(S + 1)) \in \mathbb{R}^{S+2}$ such that $\Delta g(k) \geq 0$ for all $k \in \{1, \dots, S\}$. Similarly, let $\widetilde{\mathcal{K}}_n$ be the set of all vectors $g = (g(0), \dots, g(S_n + 1)) \in \mathbb{R}^{S_n+2}$ such that $\Delta g(k) \geq 0$ for all $k \in \{1, \dots, S_n\}$. Let \widetilde{g}_0 be the least squares projection of g_0 onto $\widetilde{\mathcal{K}}_0$ and \widetilde{g}_n that of g_n onto $\widetilde{\mathcal{K}}_n$, with g_0 and g_n as in Sections 1 and 2. Finally, let $\widetilde{t}_{\alpha,n}$ be the conditional $(1 - \alpha)$ -quantile of $\widetilde{T}_n := \|\widetilde{g}_n - g_n\|_{S_n}$ given (X_1, \dots, X_n) . Recall that $\widehat{T}_0 = \|\widehat{g}_0 - g_0\|_S$ is the weak limit of T_n . We start with the following proposition.

Proposition 3.1. *We have $\|\widetilde{g}_0 - g_0\|_S \geq \widehat{T}_0$. Moreover under H_0 , conditionally on X_1, \dots, X_n , $\widetilde{T}_n \xrightarrow{d} \|\widetilde{g}_0 - g_0\|_S$ in probability, as $n \rightarrow \infty$.*

Hence, \widetilde{T}_n is typically larger than T_n , which provides the intuition for the following theorem. A proof of the theorem is given in the Supplement.

Theorem 3.2. *Let $\alpha \in (0, 1)$. Under H_0 , $\limsup_{n \rightarrow \infty} P(T_n > \widetilde{t}_{\alpha,n}) \leq \alpha$ with equality if p_0 is \mathcal{T}_{S+1} and $\alpha < \widetilde{\alpha}_0$, with $\widetilde{\alpha}_0 = P(\|\widetilde{g}_0 - g_0\|_S \neq 0) \in [1/2, 1)$.*

The test is asymptotically calibrated since the Type I error does not exceed α . It reaches precisely α when p_0 is triangular, which can be viewed as the least favorable hypothesis for testing convexity. The test is thus referred to as the LFH test in the sequel. The theorem above does not exclude existence of other least favorable hypotheses. Since with an appropriate sequence (v_n) , the test of Section 2 has the exact asymptotic Type I error α , it is typically less conservative and more powerful than the LFH test. The performance could be however poor with a bad choice of (v_n) . On the other hand, the LFH test has the great advantage of not requiring the choice of such a sequence, which makes it easier to implement.

4 Simulations

To illustrate the theory, we have considered four pmf's supported on $\{0, \dots, 5\}$. It follows from Theorem 7 in Durot et al. (2013) that any convex pmf on \mathbb{N} can be written as $\sum_{k \geq 1} \pi_k \mathcal{T}_k$ where $\pi_k \in [0, 1]$, $\sum_{k \geq 1} \pi_k = 1$ and \mathcal{T}_k is the triangular pmf supported on $\{0, \dots, k-1\}$ already defined above. Under H_0 , we considered the triangular pmf $p_0^{(1)} = \mathcal{T}_6$ and $p_0^{(2)} = \sum_{k=1}^6 \pi_k \mathcal{T}_k$ with $\pi_1 = 0, \pi_2 = \pi_3 = 1/6, \pi_4 = 0$ and $\pi_5 = \pi_6 = 1/3$, which has knots at 2, 3, 5 and 6. Under H_1 we considered $p_1^{(1)}$ the pmf of a truncated Poisson on $\{0, \dots, 5\}$ with rate $\lambda = 1.5$ and $p_1^{(2)}$ the pmf equal to $p_0^{(1)}$ on $\{2, \dots, 5\}$ such that $(p_1^{(2)}(0), p_1^{(2)}(1)) = (p_0^{(1)}(0) + 0.008, p_0^{(1)}(1) - 0.008)$. To investigate the asymptotic type I error and power of our tests, we have drawn $n \in \{50, 100, 500\}$ rv's from the aforementioned pmf's. Here, $\alpha = 5\%$, $t_{\alpha, n}$ and $\tilde{t}_{\alpha, n}$ were estimated for each drawn sample using $B = 1000$ i.i.d. copies of g_n . The rejection probability was estimated using $N = 500$ replications of the whole procedure. For the first convexity test, we considered the sequences $v_n = \sqrt{\log(\log n)}n^{-1/2}$ and $n^{-1/4}$. We also added the sequence $v_n = 0$ to compare our approach

PMF	$n = 50$			$n = 100$			$n = 500$		
	0	$\frac{(\log(\log n))^{1/2}}{n^{1/2}}$	$n^{-1/4}$	0	$\frac{(\log(\log n))^{1/2}}{n^{1/2}}$	$n^{-1/4}$	0	$\frac{(\log(\log n))^{1/2}}{n^{1/2}}$	$n^{-1/4}$
$p_0^{(1)}$	0.208	0.082	0.08	0.224	0.160	.066	0.226	0.106	0.054
$p_0^{(2)}$	0.102	0.054	0.036	0.096	0.038	0.012	0.190	0.046	0.020
$p_1^{(1)}$	0.618	0.512	0.488	0.836	0.774	0.740	1	1	1
$p_1^{(2)}$	0.192	0.108	0.044	0.206	0.120	0.036	0.234	0.102	0.038

Table 1: Type I error for the pmfs $p_0^{(1)}$ and $p_0^{(2)}$ and power for $p_1^{(1)}$ and $p_1^{(2)}$ of the convexity test based on the sequence $(v_n)_n$, with $v_n \in \{0, \sqrt{\log(\log n)}n^{-1/2}, n^{-1/4}\}$. The asymptotic level is 5%.

PMF	$n = 50$	$n = 100$	$n = 500$
$p_0^{(1)}$	0.040	0.052	0.048
$p_0^{(2)}$	0.014	0.014	0.014
$p_1^{(1)}$	0.444	0.714	1
$p_1^{(2)}$	0.052	0.036	0.042

Table 2: Type I error for the pmfs $p_0^{(1)}$ and $p_0^{(2)}$ and power for $p_1^{(1)}$ and $p_1^{(2)}$ of the test based on the least favorable hypothesis. The asymptotic level is 5%.

with the naive one where no knot extraction is attempted. The results are reported in Tables 1 and 2. Some more results for pmfs with a larger support are given in the Supplement.

The simulations confirm that the first test depends on the choice of v_n and that choosing $v_n = 0$ does not yield a valid test. The choice $v_n = \sqrt{\log(\log n)}n^{-1/2}$ is appropriate for $p_0^{(2)}$ whereas $v_n = n^{-1/4}$ is appropriate for $p_0^{(1)}$. Small sequences can have the same effect as choosing $v_n = 0$. On the contrary, large sequences can yield a conservative test in case the true pmf has small slope changes so that the estimated conditional quantile is too large to be exceeded: in such a case it may happen that no knot is extracted up so that the distance $\|\widehat{g}_n - g_n\|_{S_n}$ is too large. The question is then open as to how to choose such a sequence. To select among possible candidates of sequences $(v_n)_n$, it would be possible to resort to a cross-validation approach. This is however beyond the main scope of this article.

The second testing approach is, as expected, conservative when the true pmf is not

triangular. For $p_1^{(1)}$, which strongly violates the convexity constraint, the power is equal to 1 for $n = 500$. For $p_1^{(2)}$, which has only a small flaw at 2 with a change of slope equal to -0.008 , the power is comparable to that obtained with the first testing method and $v_n \equiv n^{-1/4}$. This is somehow expected as it is the largest sequence among the ones considered, yielding the largest distance $\|\widehat{g}_n - g_n\|_{S_n}$. As for any result based on asymptotic theory, it is difficult to know in advance the range of the sample size which would guarantee the applicability of this theory, since this would require a theoretical statement on the convergence of the involved processes at the second order. Our numerical investigation above shows that $n = 500$ is enough to obtain the correct type I error for both tests.

5 Proofs

We begin with two preparatory lemmas and then go to the proof of the main results.

Lemma 5.1. *Let $s \in \mathbb{N} \setminus \{0\}$ and $\mathcal{K} \subset \mathbb{R}^{s+2}$ a non-empty closed convex set. Then, for $u = (u(0), \dots, u(s+1)) \in \mathbb{R}^{s+2}$, the minimizer of $\|g - u\|_s$ over $g \in \mathcal{K}$ uniquely exists. Denoting by $\Phi(u)$ the minimizer, the applications Φ and $u \mapsto \|u - \Phi(u)\|_s$ are measurable.*

Proof: It follows from standard results on convex optimization that $\Phi(u)$ uniquely exists for all u , and $\|\Phi(u) - \Phi(v)\|_s \leq \|u - v\|_s$ for all u and v in \mathbb{R}^{s+2} . Hence, Φ is continuous, whence it is measurable. Now, the function $u \mapsto (u, \Phi(u))$ is continuous, whence measurable. By continuity of the norm, this ensures that $u \mapsto \|u - \Phi(u)\|_s$ is measurable. \square

Lemma 5.2. *If p_0 has support $\{0, \dots, S\}$, with $g_0 = (g_0(0), \dots, g_0(S+1))$ as in Section 1, $(\Delta g_0(1), \dots, \Delta g_0(S))$ is a centered Gaussian vector with invertible dispersion matrix.*

Proof: For notational convenience, we assume in the sequel that $S \geq 3$. The case $S \leq 2$ can be handled likewise. Let B be the $S \times (S+1)$ -matrix which j -th line has

components $j, j + 1$ and $j + 2$ equal to $1, -2$ and 1 respectively while the other components are zero, for $j = 1, \dots, S - 1$, and S -th line has components equal to zero except the penultimate and the last one, which are equal respectively to 1 and -2 . We have $p_n(S + 1) = p_0(S + 1) = 0$ almost surely so that in the limit, $g_0(S + 1) = 0$ almost surely and $(\Delta g_0(1), \dots, \Delta g_0(S))^T = B(g_0(0), \dots, g_0(S))^T$. Hence, $(\Delta g_0(1), \dots, \Delta g_0(S))$ is a centered Gaussian vector with dispersion matrix $V = B\Sigma_0 B^T$ where Σ_0 is the dispersion matrix of $(g_0(0), \dots, g_0(S))^T$, *i.e.* Σ_0 is obtained by deleting a line and a column of zeros in Γ_0 , the dispersion matrix of g_0 . It remains to prove that V is invertible.

Let $\sqrt{p_0}$ be the column vector with components $\sqrt{p_0(0)}, \dots, \sqrt{p_0(S)}$ and let $\text{diag}(\sqrt{p_0})$ be the $(S + 1) \times (S + 1)$ diagonal matrix with diagonal vector $\sqrt{p_0}$. Denoting by I the identity matrix on \mathbb{R}^{S+1} , the matrix (in the canonical basis) associated with the orthogonal projection from \mathbb{R}^{S+1} onto the orthogonal supplement of the linear space generated by $\sqrt{p_0}$ is given by $I - \Pi_0 = I - \sqrt{p_0}\sqrt{p_0}^T$. The linear subspace of \mathbb{R}^{S+1} generated by $\sqrt{p_0}$ has dimension 1, so its orthogonal supplement in \mathbb{R}^{S+1} has dimension S , whence $\text{rank}(I - \Pi_0) = S$. Now, $\Sigma_0 = \text{diag}(\sqrt{p_0})(I - \Pi_0)\text{diag}(\sqrt{p_0})$ where $\text{diag}(\sqrt{p_0})$ is invertible, so that $\text{rank}(\Sigma_0^{1/2}) = \text{rank}(\Sigma_0) = S$. Hence, the kernel of $\Sigma_0^{1/2}$ is a linear subspace of \mathbb{R}^{S+1} of dimension 1. Let λ be the column vector in \mathbb{R}^{S+1} whose components are all equal to 1. Using that $\sum_{k=0}^S p_0(k) = 1$, it is easy to see that $\Sigma_0\lambda$ is the null vector in \mathbb{R}^{S+1} , whence $\lambda^T \Sigma_0 \lambda = 0$. This means that $\|\Sigma_0^{1/2}\lambda\|^2 = 0$ with $\|\cdot\|$ the Euclidean norm in \mathbb{R}^{S+1} . Hence, $\Sigma_0^{1/2}\lambda$ is the null vector in \mathbb{R}^{S+1} , so the kernel of $\Sigma_0^{1/2}$ is the linear subspace of \mathbb{R}^{S+1} generated by λ .

Next, let us determine the kernel of V . Let $\mu = (\mu_1, \dots, \mu_S) \in \mathbb{R}^S$ with $V\mu = 0$. Then, $\mu^T V \mu = 0$ where $V = B\Sigma_0 B^T$, so that $\|\Sigma_0^{1/2} B^T \mu\|^2 = \mu^T B \Sigma_0 B^T \mu = 0$. Hence, $\Sigma_0^{1/2} B^T \mu = 0$. Since the kernel of $\Sigma_0^{1/2}$ is the linear subspace of \mathbb{R}^{S+1} spanned by λ , we conclude that $B^T \mu = a\lambda$ for some $a \in \mathbb{R}$. Hence, $\mu_1 = a$, $\mu_2 - 2\mu_1 = a$, $\mu_{k-2} - 2\mu_{k-1} + \mu_k = a$ for all $k \in$

$\{3, \dots, S\}$ and $\mu_{S-1} - 2\mu_S = a$. This is equivalent to $2\mu_k = ak(k+1)$ for all $k \in \{1, \dots, S\}$ and $2\mu_S = \mu_{S-1} - a$. Combining the first equation with $k = S, S-1$ to the second equation yields $aS(S+1) = -a + a(S-1)S/2$. Therefore, $a(2 + S^2 + 3S)/2 = 0$, whence $a = 0$. Hence, μ is the null vector in \mathbb{R}^S and therefore, $\text{rank}(V) = S$, and hence V is invertible. \square

Proof of Theorem 1.1: Let F be the c.d.f. of $\widehat{T}_0 := \|\widehat{g}_0 - g_0\|_S$. Note that \widehat{T}_0 is a properly defined random variable by the measurability proved in Lemma 5.1. The proof of Theorem 1.1 is inspired by that of Lemma 1.2 in Gaenssler et al. (2007). First, we will prove that

$$F(t) > 0 \quad \text{for all } t \geq 0. \quad (5.2)$$

To this end, note that $F(0) = P(g_0 \in \mathcal{K}_0)$, hence $F(0) \geq P(\delta \in A)$ with A the set of all vectors $(u_1, \dots, u_S) \in \mathbb{R}^S$ such that $u_k \geq 0$ for all $k \in \{1, \dots, S\}$, and $\delta = (\Delta g_0(1), \dots, \Delta g_0(S))$. From Lemma 5.2, the dispersion matrix of the centered Gaussian vector δ is invertible, so δ has a strictly positive density with respect to the Lebesgue measure on \mathbb{R}^S . This implies that the probability that δ belongs to a Borel set whose Lebesgue measure is not equal to zero, is strictly positive, whence $P(\delta \in A) > 0$. Thus, $F(0) \geq P(\delta \in A) > 0$. Combining this with the monotonicity of F yields (5.2).

Next, we prove that the function $\log(F)$ (which is well defined on $[0, \infty)$ thanks to (5.2)) is concave on $[0, \infty)$. For $u = (u(0), \dots, u(S+1)) \in \mathbb{R}^{S+2}$, let us write \widehat{u} the minimizer of $\|g - u\|_S$ over $g \in \mathcal{K}_0$. For all $t \in [0, \infty)$, let A_t be the set of all $u = (u(0), \dots, u(S+1)) \in \mathbb{R}^{S+2}$ such that $\|\widehat{u} - u\|_S \leq t$. Note that A_t is a Borel set in $\mathcal{B}(\mathbb{R}^{S+2})$ for all t since, according to Lemma 5.1, the application $u \mapsto \|\widehat{u} - u\|_S$ is measurable. Finally, let $\mu = P \circ g_0^{-1}$ be the distribution of g_0 on \mathbb{R}^{S+2} endowed with the Borel σ -algebra $\mathcal{B}(\mathbb{R}^{S+2})$. This means that $F(t) = \mu(A_t)$. Fix $\lambda \in (0, 1)$, $t, t' \in [0, \infty)$, and consider an arbitrary $x \in \lambda A_t + (1 - \lambda)A_{t'}$. Then, x takes the form $x = \lambda u + (1 - \lambda)v$ for some (non necessarily unique) $u \in A_t$

and $v \in A_{t'}$. By definition, both \hat{u} and \hat{v} belong to the convex set \mathcal{K}_0 and therefore, $\lambda\hat{u} + (1 - \lambda)\hat{v} \in \mathcal{K}_0$. Since \hat{x} minimizes $\|g - x\|_S$ over $g \in \mathcal{K}_0$, we conclude that

$$\|\hat{x} - x\|_S \leq \|\lambda\hat{u} + (1 - \lambda)\hat{v} - x\|_S \leq \|\lambda(\hat{u} - u) + (1 - \lambda)(\hat{v} - v)\|_S,$$

using that $x = \lambda u + (1 - \lambda)v$. It then follows from the triangle inequality that $\|\hat{x} - x\|_S \leq \lambda\|\hat{u} - u\|_S + (1 - \lambda)\|\hat{v} - v\|_S$. Since $u \in A_t$ and $v \in A_{t'}$, we have $\|\hat{u} - u\|_S \leq t$ and $\|\hat{v} - v\|_S \leq t'$, which implies that $\|\hat{x} - x\|_S \leq \lambda t + (1 - \lambda)t'$. Hence, $x \in A_{\lambda t + (1 - \lambda)t'}$ and

$$\lambda A_t + (1 - \lambda)A_{t'} \subset A_{\lambda t + (1 - \lambda)t'}. \quad (5.3)$$

Now, $\mu = P \circ g_0^{-1}$ is a Gaussian probability measure on $\mathcal{B}(\mathbb{R}^{S+2})$, so it follows from Lemma 1.1 in Gaenssler et al. (2007) that μ is log-concave in the sense that $\mu_\star(\lambda A + (1 - \lambda)B) \geq \mu(A)^\lambda \mu(B)^{1 - \lambda}$ for all $\lambda \in (0, 1)$ and $A, B \in \mathcal{B}(\mathbb{R}^{S+2})$, with μ_\star the inner measure pertaining to μ . Applying this with $A = A_t$ and $B = A_{t'}$, and combining with (5.3) yields $\mu_\star(A_{\lambda t + (1 - \lambda)t'}) \geq \mu(A_t)^\lambda \mu(A_{t'})^{1 - \lambda}$ for all $t, t' \in [0, \infty)$ and $\lambda \in (0, 1)$. The same inequality remains true with μ_\star replaced by μ . Using the equality $F(t) = \mu(A_t)$, $t \in [0, \infty)$, and taking the logarithm on both sides of the inequality, we conclude that $\log(F(\lambda t + (1 - \lambda)t')) \geq \lambda \log(F(t)) + (1 - \lambda) \log(F(t'))$ for all $\lambda \in (0, 1)$, and $t, t' \in [0, \infty)$. This means that the function $\log(F)$ is concave on $[0, \infty)$. Recalling (5.2), this implies (i).

Now, with probability one, \hat{p}_n is supported on $\{0, \dots, S + 1\}$ for sufficiently large n (see Balabdaoui et al. (2017, Proposition 3.5)), and p_n also is supported on that set by definition, so $T_n = \sqrt{n}\|p_n - \hat{p}_n\|_S$. It can be proved in the same manner as (Balabdaoui et al., 2017, Theorem 3.2) that under H_0 , $\sqrt{n}(\hat{p}_n - p_0, p_n - p_0) \xrightarrow{d} (\hat{g}_0, g_0)$ as $n \rightarrow \infty$ as a joint weak convergence on $\{0, \dots, S + 1\}$, so $T_n \xrightarrow{d} \hat{T}_0$ as $n \rightarrow \infty$. The last assertion follows from the latter weak convergence together with (i), since $T_n \geq 0$ and $\hat{T}_0 \geq 0$. \square

Proof of Theorem 2.1: Clearly, \mathcal{K}_n is a non-empty closed convex subset of \mathbb{R}^{S_n+2} . Hence, we consider the specific case of $s = S_n$ and $\mathcal{K} = \mathcal{K}_n$ in Lemma 5.1. In the notation of the lemma, we have $\hat{g}_n := \Phi(g_n)$, so that \hat{g}_n is uniquely defined. Moreover, since both Φ and g_n are measurable, we conclude that $\hat{g}_n = \Phi(g_n)$ is measurable. Likewise, $\|g_n - \hat{g}_n\|_{S_n}$ is measurable. This proves the first two assertions in Theorem 2.1. Next, similar to Balabdaoui et al. (2017, Theorem 3.3), the following joint weak convergence on $\{0, \dots, S + 1\}$ can be proved under H_0 : conditionally on X_1, \dots, X_n , $(\hat{g}_n, g_n) \xrightarrow{d} (\hat{g}_0, g_0)$ in probability as $n \rightarrow \infty$. The result follows, since $S_n = S$ with probability that tends to one. \square

Proof of Proposition 3.1: The first assertion follows from the inclusion $\tilde{\mathcal{K}}_0 \subset \mathcal{K}_0$. Next, similar to Balabdaoui et al. (2017, Theorem 3.3), the following joint weak convergence on $\{0, \dots, S + 1\}$ can be proved under H_0 : conditionally on X_1, \dots, X_n , $(\tilde{g}_n, g_n) \xrightarrow{d} (\tilde{g}_0, g_0)$ in probability as $n \rightarrow \infty$. The second assertion follows since $\lim_{n \rightarrow \infty} P(S_n = S) = 1$. \square

References

- BALABDAOUI, F., DUROT, C. and KOLADJO, F. (2017). On asymptotics of the discrete convex LSE of a p.m.f. *Bernoulli* **23** 1449–1480.
- BALABDAOUI, F., JANKOWSKI, H., RUFIBACH, K. and PAVLIDES, M. (2013). Asymptotic distribution of the discrete log-concave mle and some applications. *JRSS-B*. **75** 769–790.
- CHEE, C.-S. and WANG, Y. (2016). Nonparametric estimation of species richness using discrete k -monotone distributions. *Comput. Statist. Data Anal.* **93** 107–118.
- DUROT, C., HUET, S., KOLADJO, F. and ROBIN, S. (2013). Least-squares estimation of a convex discrete distribution. *Comput. Statist. Data Anal.* **67** 282–298.

- DUROT, C., HUET, S., KOLADJO, F. and ROBIN, S. (2015). Nonparametric species richness estimation under convexity constraint. *Environmetrics* **26** 502–513.
- DYKSTRA, R. L. (1983). An algorithm for restricted least squares regression. *Journal of the American Statistical Association* **78** 837–842.
- GAENSSLER, P., MOLNÁR, P. and ROST, D. (2007). On continuity and strict increase of the cdf for the sup-functional of a gaussian process with applications to statistics. *Results in Mathematics* **51** 51–60.
- GIGUELAY, J. (2016). Estimation of a discrete probability under constraint of k-monotony. *arXiv preprint arXiv:1608.06541* .
- GÓMEZ-DÉNIZ, E. and CALDERÍN-OJEDA, E. (2011). The discrete lindley distribution: properties and applications. *Journal of Statistical Computation and Simulation* **81** 1405–1416.
- JANKOWSKI, H. K. and WELLNER, J. A. (2009). Estimation of a discrete monotone distribution. *Electronic journal of statistics* **3** 1567.
- KARLIS, D. and XEKALAKI, E. (2000). A simulation comparison of several procedures for testing the poisson assumption. *Journal of the Royal Statistical Society: Series D (The Statistician)* **49** 355–382.
- LEDWINA, T. and WYŁUPEK, G. (2016). On charlier polynomials in testing poissonity. *Communications in Statistics-Simulation and Computation* 1–15.
- MEINTANIS, S. and NIKITIN, Y. Y. (2008). A class of count models and a new consistent test for the poisson distribution. *Journal of Statistical Planning and Inference* **138** 3722–3732.