Statistics and Probability Letters 201 (2023) 109900

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Ultra log-concavity of discrete order statistics

Llorenç Badiella [a,b,*], Joan del Castillo [b], Pedro Puig [b,c]

[a] Servei d'Estadística Aplicada, Universitat Autònoma de Barcelona, Cerdanyola, Spain
[b] Departament de Matemàtiques, Universitat Autònoma de Barcelona, Cerdanyola, Spain
[c] Centre de Recerca Matemàtica, Cerdanyola, Spain

ARTICLE INFO

Article history:
Received 8 November 2022
Received in revised form 12 June 2023
Accepted 23 June 2023
Available online 26 June 2023

Keywords:
Poisson distribution
Underdispersion

ABSTRACT

In this work we show that discrete order statistics preserve log-concavity and ultra log-concavity. We use a recursive expression for discrete order statistics and the concept of synchronized sequences. This finding allows to conclude that Poisson order statistics are underdispersed.

1. Introduction

Order statistics are a subject of interest in statistics and applied probability. Sometimes data come from order statistics of counts, for instance, in the study of avian fauna some researchers use the maximum of a series of replicated observations (Chamberlain et al., 2009; Hartill et al., 2011); in sports such as soccer, the amount of points achieved by a team at the end of the season is modeled according to their final position (Emparanza and Núnez-Antón, 2010); in citometry, the median of individual counts is frequently used to measure the presence of various cell types (Lesko et al., 2013; Nielson et al., 1991), and in discrete process control, it is common to monitorize certain quantiles (Jiang, 2010; Wu et al., 2014).

Discrete and continuous log-concave distributions play an increasingly important role in probability, statistics, optimization theory, econometrics and other areas of applied mathematics. Log-concavity is connected to different branches of mathematics and statistics, including concentration of measure, log-Sobolev inequalities, MCMC algorithms, Laplace approximations, and machine learning (Saumard and Wellner, 2014). The preservation of log-concavity and ultra log-concavity under different operations such as marginalization, convolution, formation of products, and limits in distribution has been object of study by a number of authors (Saumard and Wellner, 2014). The class of ultra log-concave discrete distributions plays a fundamental role in the characterization of the Poisson distribution as a maximum entropy distribution and in the study of the Law of Small Numbers (Harremoës, 2001; Johnson, 2007).

Definition 1. Let $A = (a_k)_{k=-\infty}^{\infty}$ be a sequence of non-negative real numbers. Then

(a) A is said to be log-concave if $a_k^2 \geq a_{k-1}a_{k+1}$ for all k.
(b) A is said to be ultra log-concave if $ka_k^2 \geq (k+1)a_{k-1}a_{k+1}$ for all k.

* Corresponding author at: Servei d'Estadística Aplicada, Universitat Autònoma de Barcelona, Cerdanyola, Spain.
E-mail address: badiella@mat.uab.cat (L. Badiella).

https://doi.org/10.1016/j.spl.2023.109900

Let X be a discrete distribution with $f_k = P(X = k)$. We identify the sequence $(f_k)_{k=0}^{\infty}$, with the infinite sequence $(f'_k)_{k=-\infty}^{\infty}$, where $f'_k = f_k$ for $k \geq 0$ and $f'_k = 0$ otherwise. For simplicity, hereafter, the sequence (f_k) will be denoted as f_k. In this sense, considering the sequences given by the probability function of different discrete variables, some distributions including Bernoulli, Bernoulli sums, hypergeometric, Poisson, and truncated Poisson, have the property of being ultra log-concave. Other distributions, including the geometric and the negative binomial distributions, are log-concave but not ultra log-concave. The logarithmic distribution and in general, bimodal distributions are not log-concave.

Log-concavity of order statistics has been extensively studied in the continuous case, however, due to the presence of ties, the discrete case presents more difficulties. Recent works by Alimohammadi et al. (2015) and Kim et al. (2018) have focused on studying the strong unimodality of discrete sequences, being this notion equivalent to log-concavity (Keilson and Gerber, 1971). They showed that the probability mass function (p.m.f.) of order statistics preserve the log-concavity of the original discrete distribution.

The aim of the present work is to show that order statistics of an ultra log-concave distribution are also ultra log-concave. Since ultra log-concavity is connected to underdispersion (del Castillo and Pérez-Casany, 2005; Johnson, 2007), it is shown as a corollary that Poisson order statistics are underdispersed. Statisticians should take this property into consideration when modeling this type of data.

2. Discrete order statistics

The distribution of order statistics for discrete distributions does not have a simple formulation due to the presence of ties. Let X be a discrete distribution with $f_k = P(X = k)$, $F_k = \sum_{i=0}^{k} f_i$, and $S_k = 1 - F_k$, denoting the p.m.f., the cumulative distribution function, and the survival function respectively. Given a sample of size n, let $X_{r:n}$ be the rth order statistic. Its p.m.f. can be expressed using the beta integral form (Arnold et al., 2008):

$$P(X_{r:n} = k) = I_{F_k}(r, n+1-r) - I_{F_{k-1}}(r, n+1-r) \tag{1}$$

where $I_x(a, b) = \int_0^x t^{a-1}(1-t)^{b-1}dt/\mathrm{B}(a, b)$ is the incomplete beta function and $\mathrm{B}(a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}$ is the beta function.

Alternatively, we propose an expression for order statistics using a recursive approach which will be used in proving the main theorem. First of all, define $P(X_{r:n} = k) = 0$ whenever $r < 1$ or $r > n$ and $(n-r)\mathrm{B}(r, n-r) = 1$ if $n = r$.

Theorem 1. *Let X be a discrete distribution with $X_{r:n}$, the rth order statistic. Then, for $n > 1$,*

$$P(X_{r:n} = k) = P(X_{r-1:n-1} = k)\, F_k + P(X_{r:n-1} = k)\, S_k + f_k \frac{F_{k-1}^{r-1} S_{k-1}^{n-r}}{(n-r)\mathrm{B}(r, n-r)}$$

Proof. Using expression (1), different properties of the incomplete beta function and taking into account that $F_k = F_{k-1} + f_k$,

$$\begin{aligned}
P(X_{r:n} = k) &= I_{F_k}(r, n+1-r) - I_{F_{k-1}}(r, n+1-r)\\
&= F_k I_{F_k}(r-1, n+1-r) + S_k I_{F_k}(r, n-r)\\
&\quad - F_{k-1} I_{F_{k-1}}(r-1, n+1-r) - S_{k-1} I_{F_{k-1}}(r, n-r)\\
&= F_k \left(I_{F_k}(r-1, n+1-r) - I_{F_{k-1}}(r-1, n+1-r)\right) +\\
&\quad S_k \left(I_{F_k}(r, n-r) - I_{F_{k-1}}(r, n-r)\right) +\\
&\quad f_k I_{F_{k-1}}(r-1, n+1-r) - f_k I_{F_{k-1}}(r, n-r)\\
&= P(X_{r-1:n-1} = k) F_k + P(X_{r:n-1} = k) S_k +\\
&\quad f_k (I_{F_{k-1}}(r-1, n+1-r) - I_{F_{k-1}}(r, n-r))
\end{aligned}$$

Finally, we find that

$$P(X_{r:n} = k) = P(X_{r-1:n-1} = k) F_k + P(X_{r:n-1} = k) S_k + f_k \frac{F_{k-1}^{r-1} S_{k-1}^{n-r}}{(n-r)\mathrm{B}(r, n-r)} \quad \square$$

3. Log-concave and synchronized series

We start by reviewing basic properties and notation for log-concave sequences. Let $A + B$, $A \times B$ denote the sequences with coefficients $(a_k + b_k)$ and $(a_k \times b_k)$ respectively, whereas uA denotes the sequence with coefficients (ua_k), for any constant $u \geq 0$. The convolution of A and B, denoted as $A * B$, is defined to be the sequence with coefficients: $\sum_{i=-\infty}^{\infty} a_{k-i} b_i$. For any sequence $A = (a_k)$, we define the associated offset sequence $A^- = (a_k^-)$ by $a_k^- = a_{k-1}$ for all k. We highlight the following properties of log-concave sequences:

1. Log-concavity of discrete sequences is preserved by products. If A and B are log-concave sequences then the sequence $A \times B$ is also log-concave.
2. The convolution of two log-concave sequences is log-concave.

3. Given a log-concave sequence A, the offset sequence A^- is log-concave.
4. If f_k is a log-concave sequence, so are the sequences F_k and S_k.
5. If f_k is a log-concave sequence, $F_k^i F_{k-1}^j S_k^l S_{k-1}^m$ for any $i, j, l, m \geq 0$ is a log-concave sequence.

Properties (1) and (3) are straightforward using the definition of a log-concave sequence. For a proof of the second property see Theorem 4.7.1 in Prékopa (2013) . To assess property (4), it suffices to consider the convolution between the log-concave sequence f_k with the constant sequence $q_k = 1$, and the convolution between f_k with the constant sequence $q'_k = 0$ respectively (see Theorem 4.7.2. in Prékopa, 2013). An alternative proof of property (4) can also be found in Theorem 2.2 in Alimohammadi et al. (2015). Property (5) is clear using properties (1) and (3).

In a series of recent works focused in the study of topological graph theory and combinatorics, and the Genus distribution of a graph, Gross et al. (2015) introduced new tools to deal with sums and convolutions of log-concave sequences; in particular, the concept of synchronized sequences.

Definition 2. Let A and B be two log-concave series. They are said to be synchronized, denoted as $A \sim B$, if they satisfy

$$a_k b_k \geq a_{k-1} b_{k+1} \quad \text{and} \quad a_k b_k \geq a_{k+1} b_{k-1} \quad \text{for all } k.$$

Proposition 1. *The following properties hold:*

1. *The sequence resulting from a linear combination of synchronized sequences is log-concave, i.e. let A and B be synchronized sequences, and let $u, v > 0$; then $uA + vB$ is log-concave.*
2. *Given a set of n pairwise synchronized sequences denoted as $(A_i)_{i=1}^n$, for any numbers $u_1, v_1, \ldots, u_n, v_n \geq 0$, we have $\sum_{i=1}^n u_i A_i \sim \sum_{i=1}^n v_i A_i$.*
3. *If $A \sim B$ and $C \sim D$, then $(A \times C) \sim (B \times D)$.*

and if f_k is a log-concave sequence,

4. $F_k \sim F_{k-1}$ *and* $S_k \sim S_{k-1}$
5. $F_k^i \sim F_{k-1}^i$ *and* $S_k^i \sim S_{k-1}^i$ *for* $i > 0$.
6. $F_k S_k \sim F_k S_{k-1} \sim F_{k-1} S_k \sim F_{k-1} S_{k-1}$.
7. $F_k^i F_{k-1}^j \sim F_k^{i'} F_{k-1}^{j'}$ *with* $i + j = i' + j'$; *and* $S_k^l S_{k-1}^m \sim S_k^{l'} S_{k-1}^{m'}$ *with* $l + m = l' + m'$.
8. $F_k^i F_{k-1}^j S_k^l S_{k-1}^m \sim F_k^{i'} F_{k-1}^{j'} S_k^{l'} S_{k-1}^{m'}$ *with* $i + j = i' + j'$ *and* $l + m = l' + m'$.

Proof. Property (1) is a particular case of property (2). See Theorem 2.3 from Gross et al. (2015) for a proof of property (2). Property (3) can be assessed using the definition of synchronicity. The first condition for synchronicity in property (4) is obvious. The second condition becomes $F_k F_{k-1} \geq F_{k+1} F_{k-2}$, which is true since F_k is a log-concave sequence. Properties (5) and (6) can be proven using (3) and (4). In order to prove property (7), without loss of generality we assume that $i > i'$. Thus, the property becomes $F_k^{i-i'}(F_k^{i'} F_{k-1}^j) \sim F_{k-1}^{j'-j}(F_k^{i'} F_{k-1}^j)$, which is true by (5) and (3). Finally (8) is a consequence of (7) and (3). □

4. Ultra log-concavity of discrete order statistics

Theorem 2. *Discrete order statistics preserve log-concavity and ultra log-concavity.*

Proof. Consider a discrete random variable X with f_k as its p.m.f. and assume that f_k is a log-concave or ultra log-concave sequence. In order to demonstrate the theorem we will first show that $P(X_{r:n} = k)/f_k$ can be expressed as a linear combination of log-concave synchronized terms:

$$P(X_{r:n} = k)/f_k = \sum_{i=0}^{r-1} \sum_{m=0}^{n-r} c(i, m) F_k^i F_{k-1}^{r-1-i} S_k^m S_{k-1}^{n-r-m} \tag{2}$$

i.e. a linear combination of terms where the sum of the exponents in F_k and F_{k-1} is $r - 1$ and the sum of the exponents in S_k and S_{k-1} is $n - r$ for all $k \geq 1$ and $r \geq 1$.

The proof of this property follows by induction:

- For $r = 1$ $n = 1$ expression (2) holds:

$$P(X_{1:1} = k)/f_k = 1$$

- Suppose that for some $n_0 \in \mathbb{N}, n_0 \geq 1$ and $\forall r \in \mathbb{N}, r \leq n_0$, the property is true. Using the representation given by Theorem 1 and choosing an arbitrary r_0 $(\leq n_0 + 1)$, we can write:

$$P(X_{r_0:n_0+1} = k)/f_k = F_k P(X_{r_0-1:n_0} = k)/f_k$$

$$+S_k P(X_{r_0:n_0} = k)/f_k + \frac{F_{k-1}^{r_0-1} S_{k-1}^{n_0-r_0+1}}{(n_0 - r_0)\mathrm{B}(r_0, n_0 + 1 - r_0)}$$

Each one of the three terms can be expressed as a linear combination of terms fulfilling the desired property. Using properties (2) and (7) from Proposition 1 it follows that $P(X_{r:n} = k)/f_k$ is a log-concave sequence. Thus, if f_k is a log-concave sequence so are their order statistics because $P(X_{r:n} = k)$ can be expressed as a product of log-concave sequences and log-concavity is preserved by products. On the other hand, the ultra log-concavity of the p.m.f. sequence for a random variable is equivalent to its log-concavity with respect to the Poisson distribution, given p_k the p.m.f. of the Poisson distribution, f_k is ultra log-concave if and only if f_k/p_k is log-concave. Thus, if f_k is ultra log-concave, $P(X_{r:n} = k)/p_k$ is log-concave, and $P(X_{r:n} = k)$ is ultra log-concave. These results lead us to conclude that discrete order statistics preserve log-concavity and ultra log-concavity. □

Remark 1. The fact that discrete order statistics preserve log-concavity was also proved by Kim et al. (2018) using a totally different approach in terms of strong unimodality.

Corollary 1. *Order statistics of a Poisson distribution are under-dispersed distributions (i.e. variance is strictly smaller than the mean).*

Proof. According to Johnson (2007), given X a discrete ultra log-concave random variable,

$$\mathrm{E}[X(X-1)] \leq (\mathrm{E}[X])^2 \tag{3}$$

Since the Poisson distribution has the maximum entropy property within ultra log-concave distributions (Johnson, 2007) this is the only ultra log-concave distribution fulfilling equality in (3). On the other hand, Poisson order statistics are not Poisson distributed, as can be assessed using formulation (1). □

Remark 2. Corollary 1 can also be proved using Corollary 4 in del Castillo and Pérez-Casany (2005) using the fact that the distribution of Poisson order statistics can be expressed as a weighted Poisson distribution, and the function defining the weights is log-concave.

Funding

This work was partially funded by the grants RTI2018-096072-B-I00 from the Spanish Ministry of Science, Innovation and Universities and CEX2020-001084-M from the Spanish State Research Agency, through the Severo Ochoa and María de Maeztu Program for Centers and Units of Excellence in R&D.

Data availability

No data was used for the research described in the article.

References

Alimohammadi, M., Alamatsaz, M.H., Cramer, E., 2015. Discrete strong unimodality of order statistics. Statist. Probab. Lett. 103, 176–185.
Arnold, B.C., Balakrishnan, N., Nagaraja, H.N., 2008. A First Course in Order Statistics. SIAM.
Chamberlain, D.E., Glue, D.E., Toms, M.P., 2009. Sparrowhawk accipiter nisus presence and winter bird abundance. J. Ornithol. 150 (1), 247–254.
del Castillo, J., Pérez-Casany, M., 2005. Overdispersed and underdispersed Poisson generalizations. J. Statist. Plann. Inference 134 (2), 486–500.
Emparanza, I.D., Núnez-Antón, V., 2010. On the use of simulation methods to compute probabilities: application to the first division spanish soccer league. SORT-Stat. Oper. Res. Trans. 34, 181–200.
Gross, J.L., Mansour, T., Tucker, T.W., Wang, D.G., 2015. Log-concavity of combinations of sequences and applications to genus distributions. SIAM J. Discrete Math. 29 (2), 1002–1029.
Harremoës, P., 2001. Binomial and Poisson distributions as maximum entropy distributions. IEEE Trans. Inform. Theory 47 (5), 2039–2041.
Hartill, B.W., Watson, T.G., Bian, R., 2011. Refining and applying a maximum-count aerial-access survey design to estimate the harvest taken from New Zealand's largest recreational fishery. North Am. J. Fish. Manag. 31 (6), 1197–1210.
Jiang, R., 2010. Discrete competing risk model with application to modeling bus-motor failure data. Reliab. Eng. Syst. Saf. 95 (9), 981–988.
Johnson, O., 2007. Log-concavity and the maximum entropy property of the Poisson distribution. Stochastic Process. Appl. 117 (6), 791–802.
Keilson, J., Gerber, H., 1971. Some results for discrete unimodality. J. Amer. Statist. Assoc. 66 (334), 386–389.
Kim, B., Kim, J., Lee, S., 2018. Strong unimodality of discrete order statistics. Statist. Probab. Lett. 140, 48–52.
Lesko, C.R., Cole, S.R., Zinski, A., Poole, C., Mugavero, M.J., 2013. A systematic review and meta-regression of temporal trends in adult CD4+ cell count at presentation to HIV care, 1992–2011. Clin. Infect. Dis. 57 (7), 1027–1037.
Nielson, L., Smyth, G., Greenfield, P., 1991. Hemacytometer cell count distributions: implications of non-Poisson behavior. Biotechnol. Prog. 7 (6), 560–563.
Prékopa, A., 2013. Stochastic Programming, Vol. 324. Springer Science and Business Media.
Saumard, A., Wellner, J.A., 2014. Log-concavity and strong log-concavity: A review. Stat. Surv. 8 (45).
Wu, H., Gao, L., Zhang, Z., 2014. Analysis of crash data using quantile regression for counts. J. Transp. Eng. 140 (4), 04013025.