

A characterization of normality via convex likelihood ratios

Royi Jacobovic ^{*†} Offer Kella ^{*‡}

March 3, 2022

Abstract

This work includes a new characterization of the multivariate normal distribution. In particular, it is shown that a positive density function f is Gaussian if and only if the $f(x+y)/f(x)$ is convex in x for every y . This result has implications to recent research regarding inadmissibility of a test studied by Moran (1973).

Keywords: Characterization of probability distributions, multivariate normal, Gaussian, convex likelihood ratio.

AMS Subject Classification (MSC2020): 62H05, 62E10.

1 Introduction

There are many characterizations of the normal distribution. For some surveys see, *e.g.*, [4, 9, 13, 15] and more recently, [1, 3, 6, 12, 16]. The present study is about a new characterization, which is that a positive (possibly multivariate) density f is Gaussian if and only if $f(x+y)/f(x)$ is convex in x for every y . Note that the latter is equivalent to the convexity of the likelihood ratio $f(x-\theta_1)/f(x-\theta_0)$ in x for every $\theta_0 \neq \theta_1$ where θ_1, θ_2 are (possibly multivariate) location parameters. A characterization of the one parameter

^{*}Department of Statistics and data science; The Hebrew University of Jerusalem; Jerusalem 9190501; Israel. royi.jacobovic@mail.huji.ac.il, offer.kella@gmail.com

[†]Supported by the GIF Grant 1489-304.6/2019

[‡]Supported by grant No. 1647/17 from the Israel Science Foundation and the Vigevani Chair in Statistics.

exponential family via monotonicity of (more general) likelihood ratios may be found in [2].

The motivation for this follows from a statistical hypothesis problem. Assume that one observes an i.i.d. sample. The distribution is known up to a location parameter y and the need arises to test whether y is the null vector or not. In [11] Moran suggested a test which is based on a single split of the data. The rationale is to use the first part of the data in order to estimate in what direction some estimator of y diverges from the null. Then, the second part is applied in order to test whether the true value of y diverges from the null in this direction. Theorem 4 in [7] asserts that once the observations are normally distributed, then this test is inadmissible. A careful reading of the proof of this theorem yields that the same conclusion can be made for every density function $f(\cdot)$ such that $f(x+y)/f(x)$ is convex in x for every y . This issue, has been the initial motivation to characterize the space of density functions satisfying this property, which as announced can unfortunately only be Gaussian.

One potential direction for further research is the possible application of the current characterization to devising new normality tests. For various examples of normality tests see, *e.g.*, [5, 8, 10] and references therein.

2 Main results

We begin with the following.

Lemma 1 *$g : \mathbb{R}^n \rightarrow \mathbb{R}$ is convex, strictly positive and $g(x+y)/g(x)$ is convex in x for every y if and only if for some symmetric positive semidefinite $A \in \mathbb{R}^{n \times n}$, $b \in \mathbb{R}^n$ and $c \in \mathbb{R}$,*

$$g(x) = \exp \left(\frac{1}{2} x^T A x + b^T x + c \right) . \quad (1)$$

Proof: If (1) holds with a positive semidefinite A , then g is log-convex. Similarly $g(x+y)/g(x)$ is log-linear in x for every y , hence also log-convex. log-convex functions are convex.

To show the converse denote

$$g(x; y) = \lim_{t \downarrow 0} \frac{g(x+ty) - g(x)}{t} \quad (2)$$

Since $g(x+ty)$ is convex in $t \in \mathbb{R}$, this limit exists and is finite for every x, y . Since $g(x+ty)/g(x)$ is convex in x for every t, y then so is

$$\frac{g(x+ty) - g(x)}{g(x)} = \frac{g(x+ty)}{g(x)} - 1 \quad (3)$$

and thus, letting $t \downarrow 0$ implies that $g(x; y)/g(x)$ is convex and therefore also continuous in x for every y . Since g is convex then it is also continuous and thus $g(x; y)$ is continuous in x for every y .

By Theorem 25.5 of [14], g is differentiable on a dense subset D of \mathbb{R}^n . Therefore, for every $x \in D$ we have that $g(x; y) = \nabla g(x)^T y$. Since $g(x; y)$ is continuous in x then for every sequence $x_n \in D$ such that $x_n \rightarrow x$ we have that

$$g(x; y) = \lim_{n \rightarrow \infty} \nabla g(x_n)^T y \quad (4)$$

so that necessarily $g(x; y)$ is linear in y . Theorem 25.2 of [14] implies that g is (necessarily continuously) differentiable on \mathbb{R}^n . Moreover $\nabla g(x)^T y/g(x)$ is convex in x for every y . Therefore, so is $-\nabla g(x)^T y/g(x) = \nabla g(x)^T (-y)/g(x)$ which implies that $\nabla g(x)^T y/g(x)$ is affine in x for every y . In particular, taking $y_i = 1$ and $y_j = 0$ for $j \neq i$ implies that $\frac{\partial g(x)}{\partial x_i}/g(x) = a_i^T x + b_i$ for some $a_i \in \mathbb{R}^n$ and $b_i \in \mathbb{R}$. Therefore, if A is a matrix with rows a_i and $b = (b_1, \dots, b_n)^T$ we have that

$$\nabla \log(g(x)) = \frac{\nabla g(x)}{g(x)} = Ax + b. \quad (5)$$

Hence,

$$\begin{aligned} \log(g(x)) - \log(g(0)) &= \int_0^1 \frac{d \log(g(tx))}{dt} dt = \int_0^1 (A(tx) + b)^T x dt \\ &= \frac{1}{2} x^T A^T x + b^T x = \frac{1}{2} x^T Ax + b^T x. \end{aligned} \quad (6)$$

Without loss of generality A is symmetric. If not then we replace it with $(A + A^T)/2$ and the right hand side remains the same. Taking $c = \log(f(0))$ implies (1). It remains to show that A is positive semidefinite. If it is not then there exists $\lambda < 0$ and $y \neq 0$ such that $Ay = \lambda y$. Taking $x = -b/\lambda$ implies that $y^T(Ax + b) = 0$ and thus we have that

$$\begin{aligned} y^T \nabla^2 g(x) y &= y^T (g(x)(A + (Ax + b)(Ax + b)^T)) y \\ &= g(x) y^T A y = g(x) \lambda \|y\|^2 < 0 \end{aligned} \quad (7)$$

contradicting the convexity of g . ■

The following is our main result.

Theorem 1 *$f : \mathbb{R}^n \rightarrow \mathbb{R}$ is positive, Borel and integrates to 1 (a positive density function). The following are equivalent.*

- (i) $f(x + y)/f(x)$ is convex in x for every y .

- (ii) $f(x+y)/f(x)$ is log-convex in x for every y .
- (iii) $f(x+y)/f(x)$ is log-concave in x for every y .
- (iv) f is a multivariate normal density (necessarily with positive definite covariance matrix).

Proof:

(iv) \Rightarrow (iii) $f(x+y)/f(x)$ is log-linear in x for every y .

(iii) \Rightarrow (ii) If $h(x, y) = f(x+y)/f(x)$ is log-concave in x for every y , then so is $h(x+y, -y) = f(x)/f(x+y)$. Thus $f(x+y)/f(x)$ is log-convex in x for every y .

(ii) \Rightarrow (i) Any log-convex function is convex.

(i) \Rightarrow (iv) Since f is a density then integrating $f(x+y)/f(x)$ with respect to y implies that $g(x) = 1/f(x)$ is convex. Also, if $h(x, y) = f(x+y)/f(x)$ is convex in x for every y then so is $h(x+y, -y) = f(x)/f(x+y)$ and thus $g(x+y)/g(x)$ is convex in x for each y . By Lemma 1 we must have that for some symmetric positive semidefinite $A \in \mathbb{R}^{n \times n}$, $b \in \mathbb{R}^n$ and $c \in \mathbb{R}$ we have that

$$f(x) = 1/g(x) = \exp \left(-\frac{1}{2}x^T A x - b^T x - c \right). \quad (8)$$

Let $O\Lambda O^T$ be the spectral decomposition of A with $O^T O = O O^T = I$ and $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$ where $\lambda_i \geq 0$ for all i . Then, $f(Ox)|O| = f(Ox)$ is also a density satisfying

$$f(Ox) = \exp \left(-\frac{1}{2}x^T \Lambda x - (O^T b)^T x - c \right). \quad (9)$$

If $\lambda_i = 0$ for any i then the integral with respect to x_i is infinite and thus $f(Ox)$ cannot be a density. Therefore $\lambda_i > 0$ for all i and thus A is positive definite. This implies that f is a multivariate normal density with positive definite covariance matrix $\Sigma = A^{-1}$ and mean vector $\mu = -A^{-1}b$.

■

3 Some discussion

It would be interesting to check whether the conditions of Theorem 1 can be weakened. We do this with a series of remarks.

Remark 1 Given the equivalence of (ii) and (iii) of Theorem 1 a natural question is why did we not consider concavity of $f(x+y)/f(x)$ in x for all y ? This is because there is no density satisfying this. The reason is that if $f(x+y)/f(x)$ is concave in x for each y , then by integrating with respect to y , $1/f(x)$ is concave and positive and thus must be a constant (since it is bounded below by zero). Therefore $f(x)$ must be a positive constant which is impossible since f is a density.

Remark 2 In Lemma 1 if instead of convexity of g we assume that g is differentiable (all else being the same), then we would instantly arrive at the conclusion that $\nabla g(x)^T y/g(x)$ is convex (and thus concave and therefore affine) in x for each y . We note that in this case it would even suffice to assume that $g(x+y)/g(x)$ is convex in x for all y in a neighborhood of zero. In this case one would have that A that appears in this lemma need not be positive semidefinite. Although this is true, this would not help us in the proof of Theorem 1 for which Lemma 1 was needed. Note that only the convexity of $f(x+y)/f(x)$ in x for each y is assumed in (i) of Theorem 1. Nothing about the properties of f (such as differentiability/convexity/concavity) other than it being a positive density is assumed.

We note that if we assume that f is differentiable, then the condition that $f(x+y)/f(x)$ is convex in x for “all y ” can be weakened to “for all y in a neighborhood of the origin”. Our view is that assuming differentiability on top of the convexity in x of $f(x+y)/f(x)$, rather than inferring it, substantially weakens the result.

Remark 3 Can the convexity of $f(x+y)/f(x)$ in x for each y be replaced with (the weaker) quasi-convexity? The answer is no. For example take $f(x) = e^{-|x|}/2$ ($x \in \mathbb{R}$) which is, of course, not Gaussian. It is easy to check that with $y^+ = \max(y, 0)$ and $y^- = -\min(y, 0)$,

$$\log \frac{f(x+y)}{f(x)} = \begin{cases} y & x \in (-\infty, -y^+] \\ -y - 2x & x \in (-y^+, 0] \\ y + 2x & x \in (0, -y^-] \\ -y & x \in [y^-, \infty). \end{cases} \quad (10)$$

Therefore, noting that either $(-y^+, 0]$ or $(0, -y^-]$ is empty, $\log(f(x+y)/f(x))$ and thus $f(x+y)/f(x)$ is monotone, hence quasi-convex in x for each y .

Remark 4 In Remark 2 we mentioned that when f is differentiable, then convexity of $f(x+y)/f(x)$ in x for all y in a neighborhood of zero (rather than for all y) implies that f is Gaussian. Are there non-Gaussian differentiable densities for which $f(x+y)/f(x)$ is convex in x for all $y \notin (-\epsilon, \epsilon)$ for some $\epsilon > 0$? The answer is yes.

Let $f(x) = ce^{-x^4}$, where c is a normalizing constant. Then it is straightforward to check that the second derivative of $h(x, y) = f(x+y)/f(x)$ with respect to x is given by

$$h_{xx}(x, y) = h(x, y) (-12(2x+y)y + y^2(3(2x+y)^2 + y^2)^2). \quad (11)$$

Since $h_{xx}(y)/y \rightarrow -24x$ as $y \rightarrow 0$ we clearly have that for any $x \neq 0$ there are y sufficiently close to zero for which $h_{xx}(x, y) < 0$ and thus $h(x, y)$ is not convex in x for such y . This is, of course, expected for any (differentiable) non-Gaussian density.

On the other hand, since $2ab \leq a^2 + b^2$,

$$2(2x+y)y \leq (2x+y)^2 + y^2 \leq 3(2x+y)^2 + y^2 \quad (12)$$

and thus for all y such that $y^2 \geq 6$ we clearly have that

$$12(2x+y)y \leq y^2(3(2x+y)^2 + y^2) \leq y^2(3(2x+y)^2 + y^2)^2 \quad (13)$$

and thus, with $\epsilon = \sqrt{6}$, $f(x, y)/f(x)$ is convex in x for all $y \notin (-\epsilon, \epsilon)$.

References

- [1] Azzalini, A. and Genton, M. G. (2007). On Gauss's characterization of the normal distribution. *Bernoulli*, **13**, 169-174.
- [2] Borges, R. and Pfanzagl, J. A. (1963). A characterization of the one parameter exponential family of distributions by monotonicity of likelihood ratios. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*. **2**, 111-117.
- [3] Ejsmont, W. (2016). A characterization of the normal distribution by the independence of a pair of random vectors. *Statistics and Probability Letters*. **114**, 1-5.

- [4] Hamedani, G. G. (1992). Bivariate and multivariate normal characterizations: a brief survey. *Communications in Statistics-Theory and Methods*, **21**, 2665-2688.
- [5] Henze, N. (2002). Invariant tests for multivariate normality: a critical review. *Statistical Papers* **43**, 467-506.
- [6] Henze, N., Jiménez-Gamero, M. D. and Meintanis, S. G. (2019). Characterizations of multinormality and corresponding tests of fit, including for GARCH models. *Econometric Theory*, **35**, 510-546.
- [7] Jacobovic, R. (2021). Simple sufficient condition for inadmissibility of Moran's single-split test. <https://arxiv.org/pdf/2103.11205>
- [8] Jelito, D. and Pitera, M. (2021). New fat-tail normality test based on conditional second moments with applications to finance. *Statistical Papers*, **62**, 2083-2108.
- [9] Kagan, A. M., Linnik, Y. V., Linnik, Ū. V., Linnik, I. V., Linnik, I. V. and Rao, C. R. (1973). *Characterization problems in mathematical statistics*. Wiley-Interscience.
- [10] Khatun, N. (2021). Applications of normality test in statistical analysis. *Open Journal of Statistics*, **11**, 113.
- [11] Moran, P. A. (1973). Dividing a sample into two parts a statistical dilemma. *Sankhya: The Indian Journal of Statistics, Series A*, **35**, 329-333.
- [12] Novak, S. Y. (2007). A new characterization of the normal law. *Statistics and probability letters*, **77**, 95-98.
- [13] Patel, J. K. and Read, C. B. (1996). *Handbook of the normal distribution* (Vol. 150). CRC Press.
- [14] T. R. Rockafellar. (1997). *Convex analysis*. Princeton landmarks in mathematics. Princeton University Press.
- [15] Tong, Y. L. (2012). *The multivariate normal distribution*. Springer Science and Business Media.
- [16] Volkmer, H. (2013). A characterization of the normal distribution. *Journal of Statistical Theory and Applications*, **13**, 83-85.