Moments Determine the Tail of a Distribution (But Not Much Else)

Bruce G. LINDSAY and Prasanta BASAK

How much information does a finite collection of moments carry about the underlying distribution? We revive an old bound, give a new, simple formula for its calculation, and demonstrate that although very little can be said about the central part of the distribution, the tail is much more sharply defined.

KEY WORDS: Distribution bounds; Moments.

1. INTRODUCTION

Recent articles in this journal have focused on the subtle relationships between the moments of a distribution and distributional properties. McCullagh (1994) demonstrated by example that two distributions having nearly identical cumulant generating functions can have very different density functions. On the other hand, the article by Philips and Nelson (1995) shows that the positive tail of a distribution is better determined by certain moment bounds than by Chernoff's bound.

In fact, one can quite precisely answer the following question: given a target distribution $G(x)$ with $2p$ moments $m_0 = 1$, $m_1 = \int x dG(x)$, ..., $m_{2p} = \int x^{2p} dG(x)$, and another distribution $F(x)$ with the same first $2p$ moments, then how big can $|F(x) - G(x)|$ be at any particular x?

The answer depends very much on the value of x. For example, if the target distribution $G(x)$ is normal, denoted $\Phi(x)$, then even for F with the same first 60 moments in common, the best we can say at $x = 0$ is

$$|F(0) - \Phi(0)| \leq .2233.$$

Moreover, because $\Phi(x)$ is continuous at 0, we know that there will exist at least one distribution $F_0(x)$ with $|F_0(0) - \Phi(0)| \geq \frac{1}{2}(.2233)$. On the other hand,

$$|F(3) - \Phi(3)| \leq .00597.$$

So even in the least favorable case, F_0 with $|F_0(3) - \Phi(3)| \geq \frac{1}{2}(.00597)$, the tail probability $1 - F_0(3)$ cannot be grossly different from $1 - \Phi(3)$.

To derive this result, we go back to a classical bounding result, found in Akhiezer (1965)

$$|F(x) - G(x)| \leq w_p(x),$$

and then we prove that the "window" function $w_p(x)$ has the remarkably simple representation

$$w_p(x) = \left\{\mathbf{V}_p'(x) M_p^{-1} \mathbf{V}_p(x)\right\}^{-1},$$

where $\mathbf{M}_p$ is the moment matrix

$$\mathbf{M}_p = \begin{pmatrix} 1 & m_1 & m_2 & \cdots & m_p \\ m_1 & m_2 & m_3 & \cdots & m_{p+1} \\ m_2 & m_3 & m_4 & \cdots & m_{p+2} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ m_p & m_{p+1} & m_{p+2} & \cdots & m_{2p} \end{pmatrix}, \qquad (1)$$

and the $\mathbf{V}_p(x)$ is the power vector $\mathbf{V}_p(x) = (1, x, x^2, \ldots, x^p)'$. One consequence of this is that the bound is the reciprocal of a polynomial of degree $2p$ in the variable x. Thus, the bound goes to zero at the rate x^{-2p} as $x \to \infty$. As we will see, this behavior gives relatively sharp tail information. In Section 2, we define the matrices necessary for our result and briefly state the known results about differences between two distribution functions. In Section 3, the main result is given, and in Section 4 we illustrate graphically the window bound with the help of moments of the normal and lognormal distribution.

2. PRELIMINARIES

Let the distribution function $G(\cdot)$ have the set of moments given by

$$m_i(G) = m_i, i = 0, 1, 2, \ldots, \; m_0 = 1.$$

The pth *moment matrix* of the distribution function $G(\cdot)$ is then defined by the $(p+1) \times (p+1)$ matrix defined in (1). Let $D_p = \det \mathbf{M}_p$. Also define the $(p+1) \times (p+1)$ matrix

$$\mathbf{S}_p(x) = \begin{pmatrix} 1 & m_1 & m_2 & \cdots & m_{p-1} & 1 \\ m_1 & m_2 & m_3 & \cdots & m_p & x \\ m_2 & m_3 & m_4 & \cdots & m_{p+1} & x^2 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ m_p & m_{p+1} & m_{p+2} & \cdots & m_{2p-1} & x^p \end{pmatrix},$$

which is obtained by replacing the $(p+1)$th column of $\mathbf{M}_p$ by the vector $\mathbf{V}_p(x)$, and let $E_p(x) = \det \mathbf{S}_p(x)$, which is polynomial of degree p.

Consider the following orthogonal polynomials, $P_k(x)$

Bruce G. Lindsay is Distinguished Professor of Statistics, Pennsylvania State University, University Park, PA 16801. Prasanta Basak is Associate Professor of Mathematics, Penn State–Altoona, Altoona, PA 16601 (E-mail: fkv@psu.edu). The research of Bruce G. Lindsay was partially supported by NSF Grant DMS–9403847.

$$P_0(x) = 1$$

$$P_1(x) = \frac{x - m_1}{\sqrt{D_1}}$$

$$\vdots$$

$$P_k(x) = \frac{E_k(x)}{\sqrt{D_{k-1}D_k}}; k = 2, 3, \ldots,$$

and define the window functions

$$w_p(x) = \left[\sum_{k=0}^{p} |P_k(x)|^2\right]^{-1}.$$

The following theorem (Akhiezer 1965, corollary 2.5.4, p. 66) establishes the relevant relationships of the difference between two distributions $F(\cdot)$ and $G(\cdot)$ and the functions $w_p(x)$.

Theorem 1. Let any two arbitrary distributions $F(\cdot)$ and $G(\cdot)$ have the same first $2p$ moments: $m_i(F) = m_i(G) = m_i, i = 0, 1, 2, \ldots, 2p$ with $m_0 = 1$. Then, for all the values of x,

$$|F(x) - G(x)| \leq w_p(x).$$

3. MAIN RESULT

Theorem 2. Let any two arbitrary distributions $F(\cdot)$ and $G(\cdot)$ have the same first $2p$ moments: $m_i(F) = m_i(G) = m_i, i = 0, 1, 2, \ldots, 2p$ with $m_0 = 1$. Then, for all the values of x,

$$|F(x) - G(x)| \leq \frac{1}{\mathbf{V}_p'(x)\mathbf{M}_p^{-1}\mathbf{V}_p(x)}. \quad (2)$$

Proof: In light of Theorem 1 it is enough to show that $\mathbf{V}_p'(x)\mathbf{M}_p^{-1}\mathbf{V}_p(x) = [w_p(x)]^{-1}$.

The polynomial $\sum_{k=0}^{p} |P_k(x)|^2$, called the *kernel polynomial of degree* p, can be shown to be equal to (see Akhiezer 1965, p. 9)

$$-\frac{1}{D_p} \det \begin{pmatrix} 0 & 1 & x & x^2 & \cdots & x^p \\ 1 & 1 & m_1 & m_2 & \cdots & m_p \\ x & m_1 & m_2 & m_3 & \cdots & m_{p+1} \\ x^2 & m_2 & m_3 & m_4 & \cdots & m_{p+2} \\ \vdots & \vdots & \vdots & \vdots & \ldots & \vdots \\ x^p & m_p & m_{p+1} & m_{p+2} & \cdots & m_{2p} \end{pmatrix},$$

which is nothing but

$$-\frac{1}{D_p} \det \begin{pmatrix} 0 & \mathbf{V}_p'(x) \\ \mathbf{V}_p(x) & \mathbf{M}_p \end{pmatrix}.$$

The above determinant can be evaluated using results on partitioned matrices (see Rao 1973, p. 32) to get the desired result.

Further insights can be gained by understanding the construction of the window upper bound. First, given any x_0, one can construct a $p+1$ point distribution $F_p(x)$ with the following properties:

1. It has mass $w_p(x_0)$ at x_0.
2. It has the same first $2p$ moments as that of G.
3. Every distribution function $G(x)$ with the given moments satisfies

$$F_p(\zeta^-) \leq G(\zeta) \leq F_p(\zeta),$$

at every mass point ζ of F_p, including x_0.

Property (3) together with property (1) give the window bound at x_0. But we can also note that if the target $G(x)$ is continuous at x_0, then

$$\max\left\{G(x_0^-) - F_p(x_0^-), F_p(x_0) - G(x_0)\right\}$$

is no smaller than $w_p(x_0)/2$. Such a discrete distribution clearly creates the worst case departure in distribution functions from G at x_0.

Although the calculation of the discrete distribution function $F_p(x)$ is numerically challenging for large p, relative to the window calculation, it can be used to provide an upper and lower bound for G itself:

$$F_p(x_0^-) \leq G(x_0) \leq F_p(x_0).$$

4. EXAMPLES

The results are shown graphically in Figures 1a and 1b using the moments of the standard normal distribution for $p = 10$, $p = 20$, $p = 30$ and for different values of x. We match first $2p$ moments of any distribution with that of standard normal distribution and compute the upper bound. Figure 1 shows the bound on the difference between two distributions, both of which have the same first $2p$ moments as the standard normal distribution. Here the "window" means the right side of the Equation (2). Figure 1a shows window for the middle of the distribution and Figure 1b shows window for the tail.

Similarly, Figures 2a and 2b show how the method works for the standard lognormal distribution, which is not determined by its moments. Here we standardize the usual standard lognormal distribution, the probability density function of which is given by

$$\frac{1}{x\sqrt{2\pi}} e^{-(\ln x)^2/2}; \quad 0 < x < \infty.$$

The mean and variance of standard lognormal distribution are $\sqrt{e}$, and $(e^2 - e)$, respectively. In this case, the window bounds cannot decrease to zero as $p \to \infty$. However, the bound still goes to zero as $x \to \infty$ as explained in the introduction. Thus, as expected, the window becomes smaller at the tail even though it is extremely large near zero. We note parenthetically that even though the graphs for different values of p look identical, the windows are monotonically decreasing as p increases. The differences are of the order of 10^{-5} between $p = 10$ and $p = 20$, and of the order of 10^{-10} between $p = 20$ and $p = 30$.

[Received November 1995. Revised March 2000.]

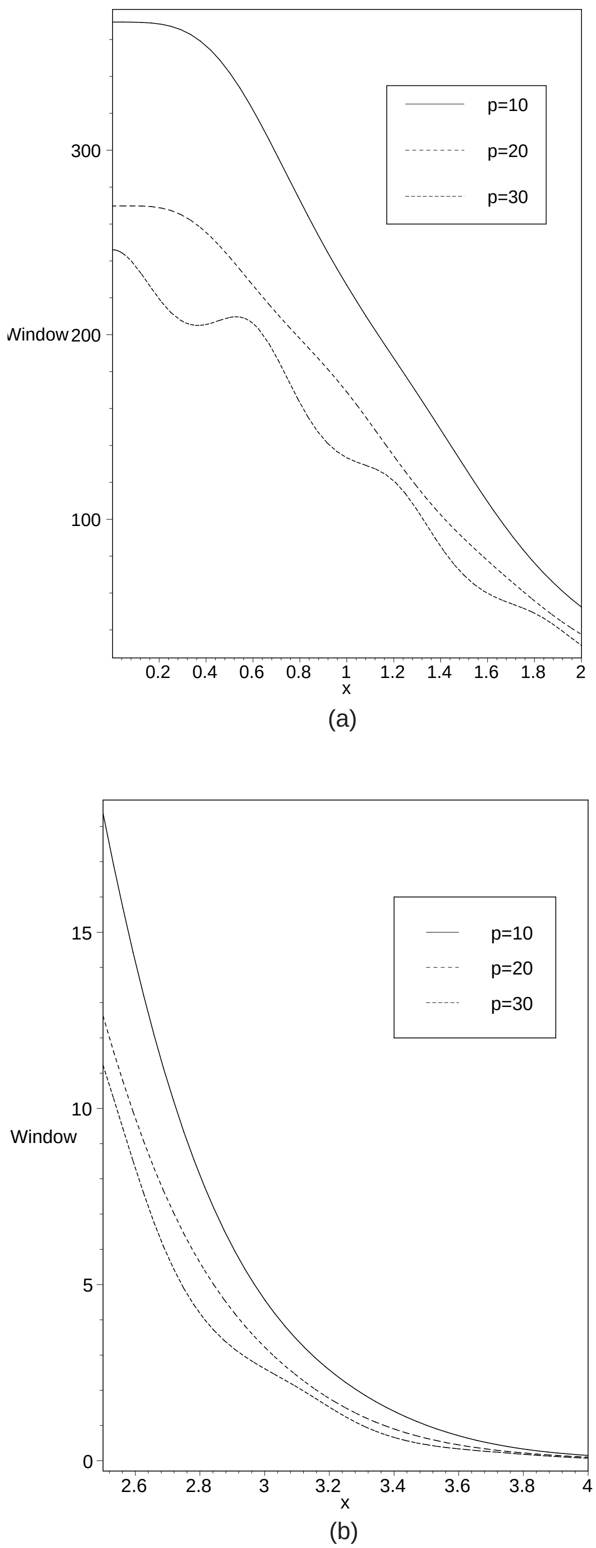

Figure 1. (a) Upper Bound (x 1000) of the Difference Between Two Distributions Having Same First 2p Moments as Standard Normal. (b) Upper Bound (x 1000) of the Difference Between Two Distributions Having Same First 2p Moments as Standard Normal.

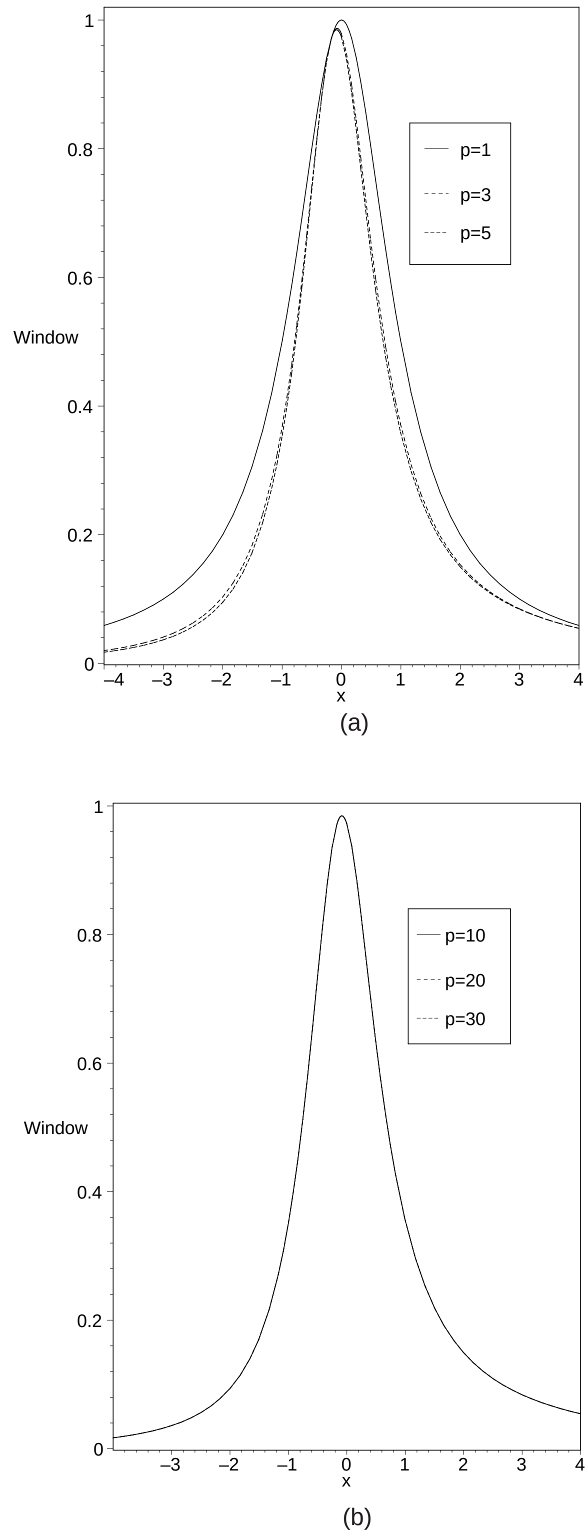

Figure 2. (a) Upper Bound of the Difference Between Two Distributions Having Same First 2p Moments as (Standardized) Standard Lognormal. (b) Upper Bound of the Difference Between Two Distributions Having Same First 2p Moments as (Standardized) Standard Lognormal.

REFERENCES

Akhiezer, N. I. (1965), *The Classical Moment Problem and Some Related Questions in Analysis*, (translated by N. Kemmer), London: Oliver & Boyd.

McCullagh, P. (1994), "Does the Moment Generating Function Characterize a Distribution?" *The American Statistician*, 48, 208.

Philips, T. K., and Nelson, R. (1995), "The Moment Bound is Tighter Than Chernoff's Bound for Positive Tail Probabilities," *The American Statistician*, 49, 175–178.

Rao, C. R. (1973), *Linear Statistical Inference and its Applications*, New York: Wiley.