Stoyanov J.

[Inference Problems Involving Moment Determinacy of Distributions.](#)

Communications in Statistics: Theory and Methods 2012, 41(16-17), 2864-2878.

Copyright:

This is an Accepted Manuscript of an article published by Taylor & Francis in *Communications in Statistics: Theory and Methods* on 9th July 2012, available online: <http://dx.doi.org/10.1080/03610926.2011.605237>

DOI link to article:

<http://dx.doi.org/10.1080/03610926.2011.605237>

Date deposited:

14/08/2017

This work is licensed under a

[Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International licence](#)

This is the final author's version of the paper. It appeared in:
Communications in Statist.: Theory and Methods **41** (2012), pp. 2864–2878.
DOI 10.1080/03610926.2011.605237

Inference Problems Involving Moment Determinacy of Distributions

JORDAN STOYANOV¹

School of Mathematics & Statistics, Newcastle University, UK

We describe diverse stochastic inference problems whose solution essentially depends on the moment determinacy of some distributions involved. For a variety of stochastic models we ask questions such as ‘how to identify a distribution if knowing its moments?’, ‘how asymmetric can be a distribution with zero odd order moments?’, ‘is any mixture model identifiable?’. For specific models we provide answers, motivating arguments and illustrations. Some challenging open questions are outlined.

Keywords Distributions; Problem of moments; Stieltjes class; Index of dissimilarity; Asymmetry; Mixture; Identifiability

2010 Mathematics Subject Classification 60E05, 62E10, 44A60

1. Introduction

Since important properties of any distribution are expressed in terms of the moments, it is natural to involve the moments when solving diverse inference problems. One of our goals is to show that the solution of some inference problems for complex stochastic models essentially depends on whether or not the distribution under study is, or is not, uniquely determined by the moments.

Suppose we want to use the two coefficients, skewness and kurtosis, of a distribution F . Imagine that F is non-unique by the moments. Then there will be infinitely many distributions, continuous and discrete, all with the same moments as F , hence the same skewness and kurtosis. A good reason to worry.

Let us emphasise that in this paper we are not dealing with the method of moments. We are interested in the uniqueness or non-uniqueness of distributions in terms of their moments. This is one of the aspects of the classical problem of moments. In general, in order to follow the method of moments it is necessary

¹Received February 22, 2011. Revised June 04, 2011.

Address correspondence to Jordan Stoyanov, School of Mathematics & Statistics, Newcastle University, Newcastle upon Tyne NE1 7RU, U.K.; E-mail: stoyanovj@gmail.com

the distribution to be uniquely determined by its moments. Otherwise, for non-unique distributions, the method of moments cannot work at all.

We describe briefly existing criteria for uniqueness and non-uniqueness and relate them to inference problems. We discuss topics such as identifying a distribution from known moments, measuring the difference between distributions with the same moments, asymmetry and identifiability of mixtures.

We follow the ‘Questions-Answers’ style which allows to clearly see the role in inference problems of the moment uniqueness or non-uniqueness of the distributions involved.

2. Moment Determinacy of Distributions. Some Criteria

Suppose that X is a random variable (r.v.) with distribution function (d.f.) F . We write $X \sim F$ and assume that all positive integer order moments are finite, i.e. $m_k = E[X^k] = \int x^k dF(x)$, $k = 1, 2, \dots$ are all in the interval $(-\infty, \infty)$. Thus the moment sequence $\{m_k, k = 1, 2, \dots\}$ of X , and F , is well-defined.

Assume that we know the whole moment sequence $\{m_k\}$ and want to establish if F obeys some useful property. We are interested in one of the classical questions originally discussed in works by P.L. Chebyshev and A.A. Markov and developed by T.J. Stieltjes: Is F the only d.f. with the moments $\{m_k\}$? If “yes”, we say that F is uniquely determined by its moments, or that F is *M-determinate* (M-det). If “no”, F is non-unique in terms of the moments, or F is *M-indeterminate* (M-indet) in which case there must be at least one d.f., say G , such that:

$$F \neq G, \text{ but } m_k(F) = m_k(G) \text{ for all } k = 1, 2, \dots$$

Any distribution with finite moments is either M-det or M-indet. The criteria listed below contain workable conditions guaranteeing such a property.

Nowadays problems of moments and related topics form a well-developed branch of mathematics, see the classical books by Shohat and Tamarkin (1943) and Akhiezer (1965). More recent sources are Landau (1987) and Stoyanov (1997). There are many works on inference problems for which the moment determinacy is essential. Some relevant papers and books are included in our references list.

The criteria for uniqueness or non-uniqueness depend on the set of values of X , equivalently, on the support of the d.f., which can be a bounded interval (Hausdorff problem of moments), $\mathbb{R}^+ = [0, \infty)$ (Stieltjes problem of moments), or $\mathbb{R}^1 = (-\infty, \infty)$ (Hamburger problem of moments).

Cramér Criterion. Let the moment generating function (m.g.f.) of X exist, i.e. the function $M(t) = E[e^{tX}]$ is well-defined for all $t \in (-t_0, t_0)$ with some $t_0 > 0$. Then all moments of X are finite, and moreover, F is M-det.

In this case, when there is a m.g.f., we say that F has *light tails*. Hence, any light tailed d.f. is M-det. If there is no m.g.f., F has *heavy tails*. Still, we may have all m_k finite and then either F is M-det, or F is M-indet.

Carleman Criterion. Suppose we know the moments $m_k = E[X^k]$ for all $k = 1, 2, \dots$. For X in $\mathbb{R}^1$ or $\mathbb{R}^+$, define the following *Carleman quantity*:

$$C = \sum_{k=1}^{\infty} \frac{1}{(m_{2k})^{1/2k}} \quad \text{or} \quad C = \sum_{k=1}^{\infty} \frac{1}{(m_k)^{1/2k}}.$$

In both cases, the condition $C = \infty$ is sufficient for F to be M-det.

Note that the Carleman condition is expressed in terms of all moments (hence we have to know them, or at least their asymptotic behavior). In the case of a finite Carleman quantity, $C < \infty$, we may only suggest that F is M-indet. Pakes (2001) proved that if $C < \infty$ and there is a density f of F such that $f > 0$ and the function $-\ln f(e^x)$ is ultimately concave, then F is M-indet.

Krein Criterion. Let $X \sim F$ be a r.v. with finite moments and density $f > 0$. For X in $\mathbb{R}^1$ or $\mathbb{R}^+$, define the logarithmic normalized integral, *Krein quantity*:

$$K[f] \equiv \int_{-\infty}^{\infty} \frac{-\ln f(y)}{1+y^2} dy \quad \text{or} \quad K[f] \equiv \int_{x_0}^{\infty} \frac{-\ln f(y^2)}{1+y^2} dy, \quad \text{for some } x_0 \geq 0.$$

In both cases, the condition $K < \infty$ is sufficient for F to be M-indet.

The Krein condition is expressed in terms of the density, only assuming that all moments are finite (we do not need to know explicitly the moments). There are, however, distributions with positive densities and infinite Krein quantity. In order to make a conclusion about the moment determinacy, we need a condition which is called *Lin condition*, see Lin (1997). It requires the density f to be symmetric, differentiable and the ratio $L(x) = -x f'(x)/f(x)$ to tend to infinity ultimately, i.e., for some $x_0 > 0$, $L(x) \nearrow \infty$ for $x_0 \leq x \rightarrow \infty$. Two conditions, $K[f] = \infty$ and the Lin condition, imply that the d.f. F is M-det.

Explicit Stieltjes Classes. While to apply the Carleman and Krein criteria is relatively easy, their proofs are quite sophisticated. We still can ask, how $C = \infty$ implies uniqueness, and why $K < \infty$ implies non-uniqueness? Such questions do not arise if we are able to construct a *Stieltjes class*, which is a parametrized family of different d.f.s, all having the same moments as a fixed d.f., say F . If F has density f , the Stieltjes class for F is defined as follows:

$$\mathbf{S} = \mathbf{S}(f, h) = \{f_{\varepsilon}(x) = f(x)[1 + \varepsilon h(x)], \quad x \in \mathbb{R}^1, \quad \varepsilon \in [-1, 1]\}.$$

Here h is called a *perturbation function*: $|h(x)| \leq 1$ for all $x \in \mathbb{R}^1$ and the product $f(x)h(x)$, $x \in \mathbb{R}^1$, has vanishing ‘moments’, $\int x^k f(x)h(x) dx = 0$, $k = 0, 1, 2, \dots$. If $L^2[f]$ is the Hilbert space with ‘weight function’ f , then in this space the perturbation h is orthogonal to all polynomials; see Stoyanov (2004).

If the perturbation h is a proper function, i.e. not identically zero, then for any $\varepsilon \in [-1, 1]$, f_{ε} is a density; $f_0 = f$. Denote by $X_{\varepsilon} \sim F_{\varepsilon}$ the corresponding r.v. and d.f., so $F_0 = F$ and $X_0 = X$. Thus we have that F_{ε} are different for different ε , however the moments are the same: $E[X_{\varepsilon}^k] = E[X^k]$ for $k = 1, 2, \dots$

Remark. The non-triviality when dealing with distributions which are non-unique in terms of the moments is seen from the following result of C. Berg and co-authors (the exact references are given in Stoyanov (2000)).

General Result: (C. Berg et al.) Suppose that F is a d.f. with all moments finite and such that F is M-indet. Then there are infinitely many absolutely continuous distributions and infinitely many discrete distributions, all are different and all have the same moments as F .

3. Given the Moment Sequence, What Is the Distribution?

Suppose that $X \sim F$ is a r.v. for which the moments are calculated and the moment sequence $\{m_k, k = 1, 2, \dots\}$ recorded. Then this sequence is the only information provided to us and we are asked to recognize, or identify F . We will see in the specific models below that sometimes it is possible to give a correct answer, but not always. We will make use of the criteria described in Section 2.

Question 1. We know that a r.v. $X \sim F$ has the following moment sequence:

$$m_k = E[X^k] = k!, \quad k = 1, 2, \dots \quad \text{What is } F? \quad (1)$$

Answer. We easily recognize that $F = \text{Exp}(1)$, the exponential distribution with parameter 1, its density is e^{-x} , $x > 0$. Since Exp has a m.g.f., hence a light tail, it follows by the Cramér criterion that $\text{Exp}(1)$ is the only distribution with these moments. Indeed, if ξ is a r.v., $\xi \sim \text{Exp}(1)$, then $E[\xi^k] = k!$. We also say that Exp is M-det, and that ξ is M-det. The M-det property is valid for $\text{Exp}(\lambda)$ for any λ .

Similarly we conclude that the Laplace distribution (double-exponential), its density is $\frac{1}{2}e^{-|x|}$, $x \in \mathbb{R}^1$, is also M-det. And, e.g., M-det is any distribution with bounded support.

Question 2. Suppose $X \sim F$ has the following moments:

$$m_k = E[X^k] = (2k)!, \quad k = 1, 2, \dots \quad \text{What is } F? \quad (2)$$

Answer. With the knowledge of the exponential distribution, we notice that $\{(2k)!\}$ is the moment sequence of the square ξ^2 of the r.v. $\xi \sim \text{Exp}(1)$. In this case $X = \xi^2$ will have the density $f(x) = \frac{1}{2}x^{-1/2}e^{-\sqrt{x}}$, $x > 0$. Hence the d.f. F with this density is a ‘candidate’ to be the solution. Clearly, the r.v. X does not have a m.g.f., i.e. the tail is heavy; still all moments are finite. By using the Stirling formula for $n!$ and its asymptotic as $n \rightarrow \infty$, we find that in this case the Carleman condition is satisfied: $C = \infty$. Hence the moment sequence $\{(2k)!\}$ comes from a distribution which is M-det. In other words, the d.f. F with density $f(x) = \frac{1}{2}x^{-1/2}e^{-\sqrt{x}}$, $x > 0$ is the only one with the moment sequence $\{(2k)!\}$.

Question 3. What is F , if we know that $X \sim F$ has the moments:

$$m_k = E[X^k] = (3k)!, \quad k = 1, 2, \dots? \quad (3)$$

Answer. These moments grow ‘very’ fast, there is no m.g.f., and, as in the previous question, the Cramér criterion does not work. Again, by using the Stirling formula we find that $C < \infty$, so the Carleman criterion also does not apply. We notice now that $\{(3k)!\}$ is the moment sequence of the r.v. $X = \xi^3$, the cube of $\xi \sim \text{Exp}(1)$. The density of X is $f(x) = \frac{1}{3} x^{-2/3} e^{-x^{1/3}}$, $x > 0$. We easily find that $K[f] < \infty$ and conclude by the Krein criterion that the distribution with the moments $\{(3k)!\}$ is M-indet. Referring to the above Berg’s general result we infer that the d.f. F with density $f(x) = \frac{1}{3} x^{-2/3} e^{-x^{1/3}}$, $x > 0$ is only one possible answer out of infinitely many.

Another approach is to take this density $f(x) = \frac{1}{3} x^{-2/3} e^{-x^{1/3}}$, choose the following perturbation function $h(x) = \sin\left(\frac{\pi}{6} - \sqrt{3} x^{1/3}\right)$, $x > 0$, and write the Stieltjes class:

$$\mathbf{S}(f, h) = \left\{ \frac{1}{3} x^{-2/3} e^{-x^{1/3}} \left[1 + \varepsilon \sin\left(\frac{\pi}{6} - \sqrt{3} x^{1/3}\right) \right], x > 0, \varepsilon \in [-1, 1] \right\}.$$

Clearly, the densities in this class are different, however they all have the same moments $\{(3k)!\}$.

Open Question. How to construct a purely discrete distribution with the moment sequence $\{(3k)!, k = 1, 2, \dots\}$?

Question 4. What is F , if $X \sim F$ is such that

$$m_{2k-1} = E[X^{2k-1}] = 0, \quad m_{2k} = E[X^{2k}] = (6k-1)!!, \quad k = 1, 2, \dots? \quad (4)$$

Answer. We easily guess that if η is a r.v., $\eta \sim N(0, 1)$, then $X = \eta^3$ has exactly the moments as given in (4). The density of η^3 is $f(x) = \frac{1}{3\sqrt{2\pi}} x^{-2/3} \exp[-\frac{1}{2} x^{2/3}]$ and hence the corresponding d.f. F is an answer to the above question. Is this the only possibility? Clearly, F does not have a m.g.f., so the Cramér criterion does not apply. Then we can check that $C < \infty$. Still no conclusion by the Carleman criterion. The next is to see that $K[f] < \infty$, hence the Krein criterion tells that F is M-indet. In other words, there are infinitely many distributions with the same moments as in (4). We can write explicit Stieltjes classes, for details see Berg (1988), Stoyanov (2004) or Stoyanov and Tolmatz (2005).

Open Question. How to find a discrete distribution on $\mathbb{R}^1$ whose moments are as in (4)?

Question 5. Suppose a r.v. $X \sim F$ has the following moments:

$$m_k = E[X^k] = e^{k^2/2}, \quad k = 1, 2, \dots \quad \text{What is } F? \quad (5)$$

Answer. The numbers $e^{k^2/2}$ are well-known as being the moments of a r.v. $X \sim LN(0, 1)$, the lognormal distribution with parameters 0 and 1. Its density is $f(x) = \frac{1}{\sqrt{2\pi}} \frac{1}{x} \exp\left[-\frac{1}{2}(\ln x)^2\right]$, $x > 0$. This distribution is heavy tailed, so no m.g.f., the moments grow very fast and Cramér and Carleman criteria do not apply. However, we easily check that the Krein condition is satisfied: $K[f] < \infty$.

Hence $LN(0, 1)$ is M-indet. This fact is well-known; in an analytic form it was suggested by T.J. Stieltjes in 1894. Heyde (1963) described an infinite family, or, in our terminology, a Stieltjes class, of distributions all having the same moments as $LN(0, 1)$.

Here is a remarkable case of an explicit family of purely discrete distributions with the same moments as $LN(0, 1)$. Indeed, for any $a > 0$, define the r.v. Y_a as follows:

$$P[Y_a = ae^k] = \frac{1}{C} a^{-k} e^{-k^2/2}, \quad k = 0, \pm 1, \pm 2, \dots,$$

where C is the normalizing constant. It can be shown that for any $a > 0$ we have $E[Y_a^k] = e^{k^2/2}$, the moments of $LN(0, 1)$. This is called Chihara-Leipnik example, details and references can be seen in Stoyanov (1997).

Hence, the correct answer to Question 5 should be: $LN(0, 1)$ is only one possible choice; there are infinitely many other distributions, they are all different, but all have the same moments $\{e^{k^2/2}\}$.

Question 6. We have a r.v. $X \sim F$ such that for any real λ the moments are

$$m_k(\lambda) = E[X^k] = 2e^{k^2/2} \Phi\left(\frac{k\lambda}{\sqrt{1+\lambda^2}}\right), \quad k = 1, 2, \dots, \quad (6)$$

where Φ is the standard normal d.f. What is F ?

Answer. Since for $\lambda = 0$, $\Phi(0) = \frac{1}{2}$, this is the case already discussed above in Question 5. The reader may suggest to consider a r.v. $X_\lambda \sim G_\lambda$, where G_λ is the logarithmic skew-normal distribution usually denoted by $LSN(\lambda)$. If so, the density g_λ and the moments $m_k(\lambda)$ of X_λ are as follows:

$$g_\lambda(x) = \frac{2}{y} \varphi(\ln x) \Phi(\lambda \ln x), \quad x > 0;$$

$$m_k(\lambda) = E[X_\lambda^k] = 2e^{k^2/2} \Phi\left(\frac{k\lambda}{\sqrt{1+\lambda^2}}\right), \quad k = 1, 2, \dots$$

Thus the moments of X_λ are the ones in (6) and hence $LSN(\lambda)$ is a ‘candidate’ as the answer to the question. One of the results in Lin and Stoyanov (2009) is that for any λ , $K[g_\lambda] < \infty$, implying by the Krein criterion that the distribution G_λ is M-indet. Therefore the correct answer to Question 6 is that the logarithmic skew-normal distribution $LSN(\lambda)$ is only one possibility; there are infinitely many other distributions, all with the same moments. In the paper cited above we have also given an explicit Stieltjes class.

Recall that in $LSN(\lambda)$ the parameter λ regulates the two coefficients, skewness and kurtosis and the whole shape of the distribution G_λ . As mentioned above, for any fixed λ , positive or negative, G_λ is M-indet. What can we say about the determinacy of the limit of X_λ as $|\lambda| \rightarrow \infty$? It turns out that there is a difference. With a r.v. $\zeta \sim N(0, 1)$, we have $\lim_{\lambda \rightarrow \infty} X_\lambda = e^{|\zeta|}$ which is M-indet. However, $\lim_{\lambda \rightarrow -\infty} X_\lambda = e^{-|\zeta|}$ which is a bounded r.v., hence M-det. For details see Lin and Stoyanov (2009).

4. How Different Are Two M-indet Distributions?

Suppose we have a parametrized family, say $\{F_\varepsilon, \varepsilon \in [-1, 1]\}$, of M-indet distributions all with the same moments as the ‘centre’ $F_0 = F$. Thus for any choice of $\varepsilon_1 \neq \varepsilon_2$ the two distributions F_{ε_1} and F_{ε_2} are moment-equivalent, however they are different. Now the question is: How much is the difference, or how to measure/calculate the difference between them?

We can exploit different metrics in the space of distributions. One natural possibility is to use the total variation distance $|F_{\varepsilon_1} - F_{\varepsilon_2}|_{TV}$. If we take the maximum with respect to ε_1 and ε_2 , both in $[-1, 1]$, we obtain a number which characterizes the whole family of moment-equivalent distributions.

For absolutely continuous distributions this approach was proposed and discussed in Stoyanov (2004). Here are some details. Start with a d.f. F which is M-indet, use its density f and some perturbation h and write an explicit Stieltjes class $\mathbf{S} = \mathbf{S}(f, h) = \{f_\varepsilon(x) = f(x)[1 + \varepsilon h(x)], x \in \mathbb{R}^1, \varepsilon \in [-1, 1]\}$. Then for $\varepsilon_1 \neq \varepsilon_2$ the total variation distance between F_{ε_1} and F_{ε_2} is

$$|F_{\varepsilon_1} - F_{\varepsilon_2}|_{TV} = \frac{1}{2} \int |f_{\varepsilon_1}(u) - f_{\varepsilon_2}(u)| du = \frac{1}{2} |\varepsilon_1 - \varepsilon_2| \int f(u) |h(u)| du.$$

Introduce the number $D(f, h) = \max |F_{\varepsilon_1} - F_{\varepsilon_2}|_{TV}$ in $\varepsilon_1, \varepsilon_2 \in [-1, 1]$ and call $D(f, h)$ an *index of dissimilarity* of the class $\mathbf{S}(f, h)$. We easily find that

$$D = D(f, h) = E[|h(X)|] = \int |h(x)| f(x) dx.$$

Clearly, $0 \leq D \leq 1$ and $D = 0$ iff F is M-det. (If F is M-det, $h(x) \equiv 0$.) The number $D(f, h)$ can naturally be interpreted as the maximum error we make when choosing a distribution from the class $\mathbf{S}(f, h)$. Indeed, if we choose, say F_1 , but the true distribution is F_{-1} , the error we make is equal to $D(f, h)$.

For any explicit Stieltjes class $\mathbf{S}(f, h)$ we can calculate its index of dissimilarity $D(f, h)$. However, our main interest in the M-indet d.f. F , and/or its density f . Then, h can be chosen in many ways as soon as it satisfies the requirements to be a perturbation function, see Section 2. Clearly, a convex combination of perturbations is again a perturbation. For any choice of h , the index $D(f, h)$ depends on h and there are good reasons to look for the perturbation h for which $D(f, h)$ is maximum.

Open Question. For a given M-indet d.f. F with density f , let $D(f, h)$ be the index of dissimilarity for the Stieltjes class $\mathbf{S}(f, h)$, where h is some perturbation function. Find

$$D^* = D(f) = \max_h D(f, h),$$

over the set of all perturbations h . The number D^* will be the maximum error we risk to make when choosing a d.f. from all possible classes based on the density f , not just from a Stieltjes class $\mathbf{S}(f, h)$ with a specific perturbation h .

Remark. Following the suggestion of one of the Referees we add here details of how, starting with a fixed M-indet distribution, to construct a perturbation

h. In general, this is a difficult problem. One possibility is to consider this as a complex analysis problem. This approach was successfully followed by Ostrovska and Stoyanov (2005), however the same idea can be used for distributions other than powers of the inverse Gaussian. The method developed in Stoyanov and Tolmatz (2005) is another possibility; see also the papers by Pakes (2007) and Lin and Stoyanov (2009). The case of discrete M-indet distributions is even less studied. Besides the Chihara-Leipnik example given in Question 5 above, only a few more examples are available in the literature, see Stoyanov (1997).

5. Asymmetric Distributions With Zero Odd Order Moments

Recall that a r.v. $X \sim F$ is *symmetric* with respect to zero if

$$(-X) \stackrel{d}{=} X.$$

This is equivalent to the relation $F(-x) = 1 - F(x)$ for all $x \in \mathbb{R}^1$. If there is a density, say f , we have $f(-x) = f(x)$ for all x .

Easy Property: If X is symmetric and has finite moments, then the odd order moments are all equal to zero, i.e.

$$m_{2k-1} = E[X^{2k-1}] = 0 \quad \text{for } k = 1, 2, \dots$$

Converse Question: Suppose X is a r.v. with finite moments and we know that all odd order moments are zero. Is it true that X is symmetric?

Sometimes the answer is “yes”, sometimes it is “not”. The reader may guess that the moment determinacy will play a role. Here are two examples.

Example 1. Consider a r.v. $\theta \sim Lap(1)$, this is *Laplace distribution* (or *double-exponential*) with parameter 1. Its density is $\frac{1}{2}e^{-|x|}$, $x \in \mathbb{R}^1$, all moments are finite and we have

$$m_{2k-1} = E[\theta^{2k-1}] = 0, \quad m_{2k} = E[\theta^{2k}] = (2k)! \quad \text{for } k = 1, 2, \dots$$

This distribution is symmetric. Notice, however, that $Lap(1)$ is M-det; it is the only distribution with these moments. It is useful to see again the above Question 2 and the answer. The same inference can be made for $Lap(\lambda)$ for any $\lambda > 0$, all odd order moments are zero, the others are as the above with some factors. However they are all M-det because there is a m.g.f. Moreover, they are all symmetric with respect to zero.

Example 2. Consider the following power (Box-Cox) transformation, $Y = \theta^3$, where $\theta \sim Lap(1)$. The density g of Y is

$$g(x) = \frac{1}{6} |x|^{-2/3} \exp(-|x|^{1/3}), \quad x \in \mathbb{R}^1.$$

It is easy to see that this r.v. Y is symmetric and has finite moments:

$$m_{2k-1} = E[Y^{2k-1}] = 0, \quad m_{2k} = E[Y^{2k}] = (6k)! \quad \text{for } k = 1, 2, \dots \quad (7)$$

Suppose now that we are given only the moment sequence (7). What can we conclude about the ‘unknown’ distribution, say G , which is behind? We may suggest: the answer is G with the above density g . However, there is a trouble coming from the fact this G is M-indet. The latter follows from the Krein criterion. We can also write explicit Stieltjes classes either by modifying a little the construction in Question 3 or using the method described in Stoyanov and Tolmatz (2005). The moment non-uniqueness of G implies that there are infinitely many distributions with the same moments as in (7), and, notice, their all odd order moments are equal to zero. Among them only one, the distribution of the cube of $\theta \sim Lap(1)$, is symmetric. All others are non-symmetric.

Comment. If we have in our disposal empirical data of size n drawn from an unknown distribution F , we can use them and calculate the empirical k th order moments $\hat{m}_{k,n}$ for $k = 1, 2, \dots$, finitely many or more. Assume that available to us is a long series of these moments and we see that all odd order moments are equal, or ‘very close’, to zero: $\hat{m}_{2k-1,n} \approx 0$. Can we infer that the distribution F which is behind these data is symmetric? We can make a correct inference in both cases, when F is M-det and when it is M-indet. The above examples illustrate well this point.

Index of Asymmetry. Suppose that the distribution F on $\mathbb{R}^1$ is such that it has finite moments, its odd order moments are equal to zero and it is M-indet. Let us think that we have written a Stieltjes class, say $\mathbf{S}(f, h)$, assuming f is the density of F and h is some proper perturbation. If we take $F_0 = F$ to be the centre of this class, then F itself will be symmetric, however any other F_ε will be asymmetric. Question: How much is the asymmetry of F_ε ? In Section 4 we have considered the number $D(f, h)$ as the index of dissimilarity within the class $\mathbf{S}(f, h)$. The way of defining $D(f, h)$ and its meaning suggest to measure the asymmetry of F_ε as its total variation distance from the centre $F = F_0$. Hence we can adopt that the numbers

$$A(f, h) = \frac{1}{2}D(f, h) \quad \text{and} \quad A^* = \frac{1}{2}D^*(f)$$

characterize the ‘amount’ of asymmetry. Specifically, $A(f, h)$ will be the index of asymmetry within the Stieltjes class $\mathbf{S}(f, h)$, while A^* will be the index of asymmetry of the M-indet F itself. If these indices are calculated for a specific case, we can make useful conclusions.

6. Identifiability of Mixtures and Moment Determinacy

Suppose that (X, θ) is a 2-dimensional random vector for which we know the conditional distribution of $X|\theta = t \sim F(\cdot | t)$ and the unconditional distribution of $\theta \sim G$ called a *mixing distribution*. Then we find the *mixture distribution* H , the distribution of just X : $H(\cdot) = \int F(\cdot | t)dG(t)$.

We are interested in the relationship between the distributions involved. In particular, for fixed $F(\cdot | t)$, how are the mixture H and the mixing G related to each other? Given mixing G , the mixture H is well-defined, it is unique. The

converse question is about the mixing G given the mixture H . If there is only one mixing G , producing H , the mixture model is said to be *identifiable*. If, e.g., there are at least two mixing distributions, say $G_1 \neq G_2$, both producing the same mixture H , we say that the mixture model is *non-identifiable*. Thus we want to know whether or not a specific mixture model is or is not identifiable.

Because of their importance in applications, mixture models have been intensively studied by many researchers. The reader can easily find a lot of references. We have included in our list of references only two papers, Sapatinas (1995) and Stoyanov and Lin (2011), which are directly related to the discussion below.

Inference Problem for Mixture Models. Suppose that the family of conditionals $\{F(\cdot | t)\}$ is known to us and available are observations from the mixture H , i.e. on X . Can we identify, or estimate well, the mixing G ? In general, if the mixture model is identifiable, we may suggest that this can be done. However, we guess that there is no chance if the model is non-identifiable. In what follows we are going to illustrate this phenomenon by relating the identifiability of a specific model to the moment determinacy of some distribution involved.

Specific Mixture Model. Consider the 2-dimensional model (X, θ) , where X is a discrete r.v. with values in the set $\{0, 1, 2, \dots\}$, and θ is a r.v., $\theta \sim G$, with arbitrary d.f. G . Assume that we know that the conditional distribution of $X | \theta = t$ is a *power series distribution* (PSD):

$$F_t = \{f(\cdot | t)\}, \quad f(k | t) = P[X = k | t] = \frac{a_k t^k}{A(t)}, \quad k = 0, 1, 2, \dots$$

Here $A(t) = \sum_{j=0}^{\infty} a_j t^j$ is the power series function and we assume that it has a radius of convergence $\rho_A \in (0, \infty)$. It is known that PSD includes popular distributions such as Binomial, Poisson, Negative-Binomial, Logarithmic, etc.

We want to answer the question: Is the above mixture model identifiable?

The answer will come after presenting some preliminary arguments. Assume first that $\theta \sim G$ takes values in the interval $(0, \rho_A)$. Then we find the mixture H : $H = \{h_k, k = 0, 1, 2, \dots\}$, $h_k = P[X = k] = \int f(k | t) dG(t)$. The next step is to define a ‘new’ r.v. $\tilde{\theta}$, where

$$\tilde{\theta} \sim \tilde{G}, \quad d\tilde{G}(t) = \frac{a_0}{h_0} \frac{1}{A(t)} dG(t), \quad t \in (0, \rho_A).$$

We notice that all moments of $\tilde{\theta}$ are finite and can be explicitly expressed in terms of $\{a_k\}$ and $\{h_k\}$ as follows: $\tilde{m}_k = E[\tilde{\theta}^k] = \frac{a_0}{h_0} \frac{h_k}{a_k}$, $k = 1, 2, \dots$. We do not exploit this fact here, however the next statement shows that the moment determinacy of $\tilde{G}$ is decisive to infer identifiability or non-identifiability.

Since the power series function A is fixed, we see that the d.f. $\tilde{G}$ is defined uniquely by G and vice versa, G is defined by $\tilde{G}$.

Statement. For the mixture model described above we have the following:

1. If G and A are such that the d.f. $\tilde{G}$ is M-det, then G is the only distribution

producing H . Hence the mixture model (X, θ) will be identifiable.

2. Otherwise, if $\tilde{G}$ is M-indet, the mixture model is non-identifiable. There are ‘many’ mixing G producing the same H .

As shown in Stoyanov and Lin (2011), this statement allows to establish some general new results and also to provide different and transparent proofs of some known results. Let us consider a mixture model with different choices of the mixing distribution and clarify whether or not the model is identifiable.

Examples of Mixture Models. Consider (X, θ) such that

$$P[X = k | \theta = t] = \frac{1}{(3k)!} \frac{t^k}{A(t)}, \quad k = 0, 1, 2, \dots$$

By using Formula 5.2.7.8 in Prudnikov et al. (1992), we find explicitly the power series function:

$$A(t) = \sum_{j=0}^{\infty} \frac{t^j}{(3j)!} = \frac{1}{3} \left(e^{t^{1/3}} + 2e^{-t^{1/3}/2} \cos\left(\frac{1}{2}\sqrt{3}t^{1/3}\right) \right), \text{ where } \rho_A = \infty.$$

This means that we can choose any mixing d.f. G with support $\mathbb{R}^+$ or any its subset. Let us assume that G has density g and make two different choices. In each case we ask the question: Is the mixture model identifiable?

Case 1. Take a mixing G with the density $g(t) = \frac{2}{3}t^{-1/3}e^{-t^{2/3}}$, $t > 0$. The next is to analyse the ‘new’ density $\tilde{g}(t) = \frac{a_0}{h_0} \frac{1}{A(t)}g(t)$. We can show that the d.f. $\tilde{G}$ with this density $\tilde{g}$ is M-det. Hence the mixture model (X, θ) is identifiable.

Case 2. Suppose now that the mixing density is $g(t) = \frac{4}{9}t^{-5/9}e^{-t^{4/9}}$, $t > 0$. In this case the mixture model (X, θ) is non-identifiable. This follows from the fact that the ‘new’ density $\tilde{g}$ has a finite Krein quantity, $K[\tilde{g}] < \infty$, so the d.f. $\tilde{G}$ is M-indet. If we wish, we can, e.g., write a Stieltjes class for $\tilde{G}$, by using its density $\tilde{g}$ and some appropriate perturbation h . If, say $\mathbf{S} = \{\tilde{G}_\varepsilon, \varepsilon \in [-1, 1]\}$, then any member of this class defines a ‘new’ mixing G_ε , and all these will produce the same mixture H . The trouble is obvious, for any sample of observations on $X \sim H$, we do not know which of the many mixing d.f. we try to estimate.

7. Other Problems Involving Moment Determinacy

Besides the problems discussed above there are more problems which in one or another way involve the moment determinacy of some distributions. We give a brief account of some of them. We start with a fundamental result which usually is called a second central limit theorem. Briefly it is formulated as follows.

Fréchet-Shohat Theorem. Suppose that F_n , $n = 1, 2, \dots$ is an arbitrary sequence of d.f.s. Assume that for any n , F_n has finite all integer order moments $m_{k,n} = \int x^k dF_n(x)$, $k = 1, 2, \dots$. Assume further that

$$m_{k,n} \rightarrow m_k \text{ as } n \rightarrow \infty \text{ for each } k = 1, 2, \dots$$

Then: (a) $\{m_k, k = 1, 2, \dots\}$ is a moment sequence of some distribution, say F .
(b) If F is M-det, then F_n converges weakly to F as $n \rightarrow \infty$.

Here is an equivalent form of this theorem: Let X_n , $n = 1, 2, \dots$ be a sequence of r.v.s with converging moments $m_{k,n} \rightarrow m_k$ as $n \rightarrow \infty$ for $k = 1, 2, \dots$. Then there is a r.v., say X , with the moment sequence $\{m_k\}$. If X is M-det, we have that $X_n \rightarrow X$ in distribution as $n \rightarrow \infty$.

The importance of this theorem is obvious. Suppose for example, we find the limiting moments and check that the Carleman quantity for the moment sequence $\{m_k\}$ is equal infinity, hence there is only one d.f. F with these moments. To identify F is a difficult problem, even knowing that there is such a d.f. F and it is unique. Another possibility is to use the M-det limiting distribution F to build confidence intervals when estimating unknown parameters. And there are several stochastic problems for which the weak convergence is vital.

It is important to remember that there are problems whose solution requires to follow a specific limiting procedure. It may happen, however, that classical methods such as the powerful method of characteristic functions just do not work. At the same time, by using the Fréchet-Shohat theorem we can provide a short and elegant proof of some results. Here is an illustration.

Random Walk in Random Environment. Suppose that a particle D starts from a position $x_0 \in (0, 1)$ and moves randomly, up with probability p , or down with probability $q = 1 - p$. The position X_1 of D in one unit-time is a random point in the sub-interval $(x_0, 1)$ if going up, and in the sub-interval $(0, x_0)$ if going down. The motion continues independently according to the same rules. If X_n is the position of D at time n , then $X = (X_n, n = 0, 1, 2, \dots)$ is a homogeneous Markov chain. Moreover, it is ergodic. One of the interesting questions is to find explicitly the ergodic distribution. It turns out that the efficient way (maybe the only way) to do this is to use the Fréchet-Shohat theorem. Indeed, we can show that

$$E[X_n^k] \rightarrow m_k \text{ as } n \rightarrow \infty \text{ for } k = 1, 2, \dots,$$

where $m_k = p(p+1) \dots (p+k-1)/k!$. These are the moments of some distribution in $(0, 1)$, hence it is M-det. We easily recognize that m_k is exactly the k th order moment of a r.v., say η , which follows beta distribution, with parameters p, q . Thus we conclude that in this model

$$X_n \rightarrow \eta \text{ as } n \rightarrow \infty, \text{ where } \eta \sim \beta(p, q).$$

This distribution, $\beta(p, q)$, is the ergodic distribution of the Markov chain X . Details about the convergence of the moments, and essentially more, can be seen in Stoyanov and Pirinsky (2000). The method of characteristic functions does not work here. The characteristic function of beta-distribution does not have a closed form, while the moments are explicitly known.

Extended variations of the above model and explicit ergodic distributions of Markov chains can be found in Pacheco-Gonzalez and Stoyanov (2008). The

Fréchet-Shohat theorem is well combined with some distributional equations. A variety of interesting limit results entirely based on the Fréchet-Shohat theorem can be found in DasGupta (2008) and Mahmoud (2008).

Estimation of the Moments of a Distribution. It is well-known that if a d.f. F has finite moments which are of interest to us, we can use observations on $X \sim F$ and find estimators which, under some general conditions, are unbiased, consistent and asymptotically normal.

Suppose k_* is a positive integer number and we want to estimate the first k_* population moments, $m_1, \dots, m_{k_*}$, however in a stronger probabilistic sense, not just in probability which is provided by the law of large numbers. Such a problem was discussed by Kagan and Nagaev (2001) by involving a type of complete a.s. convergence. Namely, if $\hat{m}_{k,n} = \frac{1}{n}[X_1^k + \dots + X_n^k]$ is the empirical k th order moment, the estimator of m_k , how many moments k_* can be estimated simultaneously and consistently in the sense that

$$P \left[\max_{1 \leq j \leq k_*} |\hat{m}_{j,n} - m_j| > \varepsilon \right] \rightarrow 0 \text{ as } n \rightarrow \infty?$$

It turns out, not too many, even if the sample size n is large. The answer is: we can estimate in this strong way not more than k_* moments, where k_* is the integer part of the number

$$\tilde{k}_* = (1 - \delta) \frac{\ln n}{2 \ln \ln n} (1 + o(1)).$$

Here δ is an arbitrary ‘small’ positive number. In other words, the population moments m_j can be consistently estimated uniformly in j , $j = 1, 2, \dots, k_*$. It is also shown that there is no such a convergence for a larger number of moments, larger than k_* . Details can be seen in Kagan and Nagaev (2001). We have to tell here especially that all these results are established under a strong condition on the moments of F . It is assumed that for some constant $C > 0$ we have $|m_k| \leq k!H^k$, $k = 1, 2, \dots$, or its variation, $|m_k| \leq k!H^k(1 + o(1))$ as $k \rightarrow \infty$. These are usually called Bernstein conditions and they imply that the characteristic function of F is analytic, which is equivalent to the existence of a m.g.f. Hence in this study the d.f. F satisfies the Cramér condition. Thus, F is M-det by the Cramér criterion. The existence of the m.g.f. is the strongest condition guaranteeing the moment determinacy. Since there are distributions which are M-det but do not have m.g.f.s, it would be interesting to clarify if the results in Kagan and Nagaev (2001) are valid under such a weaker assumption. If the d.f. F has finite moments and F is M-indet, then it is nonsense to ask about estimating the moments.

Recovering and/or Bounding a Distribution by its Moments. Let us mention here a few more inference problems which involve explicitly or implicitly the requirement for some distributions to be M-det.

The maximum entropy approach was successfully followed in a series of papers by Tagliani (2003) and Gzyl and Tagliani (2010a,b), see also the references

therein. In the Hausdorff case the distributions are M-det, however for distributions with support in $\mathbb{R}^+$, there is a requirement related to a Stieltjes problem of moments to have a unique solution.

In a series of papers, see Mnatsakanov and Hakobyan (2009) and Mnatsakanov (2011) and the references therein, the important practical problem was discussed: how to recover a distribution, or its density, by knowing the first n moments, theoretical or empirical. Such an inference problem is meaningful only if the distribution involved is M-det.

Other useful properties such as finding one-sided or two-sided bounds for a distribution in terms of the first n moments have been recently studied by Gavrilidis and Athanassolis (2008a,b), see also Rasz et al. (2006). Not surprisingly, the moment uniqueness in any such a problem is essential.

8. Final Comments

There are more stochastic inference problems involving moment determinacy of distributions. Let us mention here two recent papers. Penson et al. (2010) show how Stieltjes classes of M-indet distributions arise when describing coherent states in quantum theory. Lasserre et al. (2006) exploit a new methodology for numerical pricing of exotic options by using the moments of some distributions. The M-det or M-indet property of the distributions is crucial.

All the problems discussed or briefly mentioned in this paper are interesting from theoretical point of view and related to applications. Works in this area are widely spread in mathematics and statistics literature. This rich area of research, called *Moment Analysis of Distributions*, is attractive for several reasons, to mention just three: (a) the arising mathematical problems are more than challenging; (b) this is a way to enhance the stochastics theory; (c) there are serious applications in areas such as econometrics, quantum theory, finance and natural sciences.

Acknowledgments

This paper is dedicated to Professor Fortunato Pesarin on the occasion of his 70th birthday. The paper is based on my talk presented at the International Workshop ‘Statistics for Complex Systems’ (University of Padova, June 2010). I am using now the opportunity to express again my great respect to Professor Pesarin and good wishes for the years to come.

I am grateful to the anonymous Referees for the careful reading of the submitted manuscript and their useful comments. Thanks to the Editors for their idea and efficient efforts in preparing this special volume.

References

Akhiezer, N.I. (1965). *The Classical Problem of Moments and Some Related Questions of Analysis*. Oliver & Boyd, Edinburgh. (Russian original 1961)

- Berg, C. (1988). The cube of the normal distribution is indeterminate. *Ann. Probab.* 16:910–913.
- DasGupta, A. (2008). *Asymptotic Theory of Statistics and Probability*. Springer, New York.
- Gavriliadis, P.N, Athanassolis, G.A. (2008a). Moment information for probability distributions, without solving the moment problem. I. Where is the mode? *Comm. Statist. Theory Methods* 37:671–681.
- Gavriliadis, P.N, Athanassolis, G.A. (2008b). Moment information for probability distributions, without solving the moment problem. II. Main mass, tails and shape approximation. *J. Comput. Appl. Math.* 229:7–15.
- Gzyl, H., Taglioni, A. (2010a). Hausdorff moment problem and fractional moments. *Appl. Math. Comput.* 216:3319–3328.
- Gzyl, H., Taglioni, A. (2010b). Stieltjes moment problem and fractional moments. *Appl. Math. Comput.* 216:3307–3318.
- Heyde, C.C. (1963). On a property of the lognormal distribution. *J. Roy. Statist. Soc. Ser. B* 25:392–393.
- Kagan, A., Nagaev, S. (2001). How many moments can be estimated from a large sample? *Statist. Probab. Letters* 55:99–105.
- Landau, H.J., ed. (1987). *Moments in Mathematics*. Amer. Math. Soc., Providence (RI).
- Lasserre, J.B., Prieto-Rumeau, T. and Zervos, M. (2006). Pricing a class of exotic options via moments and SDP relaxations. *Math. Finance* 16:469–494.
- Lin, G.D. (1992). Characterizations of distributions via moments. *Sankhya* A54:128–132.
- Lin, G.D. (1997). On the moment problem. *Statist. Probab. Letters* 35:85–90.
- Lin, G.D., Stoyanov, J. (2009). The logarithmic skew-normal distributions are moment-indeterminate. *J. Appl. Probab.* 46:909–916.
- Mahmoud, H.M. (2008). *Pólya Urn Models*. CRC, Boca Raton (FL) and Chapman & Hall, London.
- Mnatsakanov, R., Hakobyan, A. (2009). Recovery of distributions via moments. *IMS Lecture Notes–Monograph Series* 57:252–265.
- Mnatsakanov, R. (2011). Moment-recovered approximations of multivariate distributions: Laplace transform inversion. *Statist. Probab. Letters* 81:1–7.
- Ostrowska, S., Stoyanov, J. (2005). Stieltjes classes for M-indeterminate powers of inverse Gaussian distributions. *Statist. Probab. Letters* 71:165–171.

- Pacheco-Gonzalez, C., Stoyanov, J. (2008). A class of Markov chains with beta ergodic distributions. *Math. Sci.* 33:110–119.
- Pakes, A.G. (2001). Remarks on converse Carleman and Krein criteria for the classical moment problem. *J. Austral. Math. Soc.* 71:81–104.
- Pakes, A.G. (2007). Structure of Stieltjes classes of moment-equivalent probability laws. *J. Math. Anal. Appl.* 326:1268–1290.
- Pakes, A.G., Hung, W.-L., Wu, J.-W. (2001). Criteria for the unique determination of probability distributions by moments. *Austral.N.Z. J. Stat.* 43:101–111.
- Penson, K.A., Blasiak, P., Duchamp, G.H.E., Horzela, A., Solomon, A.I. (2010). On certain non-unique solutions of the Stieltjes moment problem. *Discrete Math. Theor. Comput. Sci.* 12:295–306.
- Prudnikov, A.P., Brychkov, Yu.A., Marichev, O.I. (1992). *Integrals and Series* (Vol. 1). Gordon & Breach, New York.
- Rasz, S., Tari, A., Miklos, T. (2006). A moment based distribution bounding method. *Math. Comp. Modelling* 43:1367–1382.
- Sapatinas, T. (1995). Identifiability of mixtures of power-series distributions and related characterizations. *Ann. Inst. Statist. Math.* 47:447–459.
- Shohat, J.A., Tamarkin, J.D. (1943). *The Problem of Moments*. Amer. Math. Soc., Providence (RI).
- Slud, E.V. (1993). The moment problem for polynomial forms of normal random variables. *Ann. Probab.* 21:2200–2214.
- Stoyanov, J. (1997). *Counterexamples in Probability*. 2nd ed. Wiley, Chichester.
- Stoyanov, J. (2000). Krein condition in probabilistic moment problems. *Bernoulli* 6:939–949.
- Stoyanov, J. (2004). Stieltjes classes for moment-indeterminate probability distributions. *J. Appl. Probab.* 41A:281–294.
- Stoyanov, J., Lin, G.D. (2011). Mixtures of power series distributions: identifiability via uniqueness in problems of moments. *Ann. Inst. Statist. Math.* 63:291–303.
- Stoyanov, J., Pirinsky, C. (2000). Random motions, ergodic Markov chains and beta distributions. *Statist. Probab. Letters* 50:293–304.
- Stoyanov, J., Tolmatz, L. (2005). Method for constructing Stieltjes classes for M-indeterminate probability distributions. *Appl. Math. Comput.* 165:669–685.
- Tagliani, A. (2003). Real moments from moments and vice versa. *Appl. Math. Comput.* 145:511–523.