On Two Normal Mixture Models of the Classical Method of Moments

Ray-Shine Lee*

Abstract
This research is concerned with an implicit merit of analyzing multiple solutions in application of the classical method of moments. A system-wide computation on one of Karl Pearson's biometric samples is performed, specifically to search for additional parameter estimates by extra participation of the sixth moment measurement and equation. An alternative investigation is also made for the three-component normal mixture model, conditional on a specification with five regular parameters.

Key Words: Normal mixture, Nonic equation, Initial estimation, Efficient estimation

1. Introduction

At present, studying the mixture of normal random variables seems not so particularly difficult as contrasted to the beginning of the twentieth century. Back a few then years to the date of the innovative Pearson (1894), the immediate next decade did not attain substantial progress, either to generalize the underlying method of solving equations or effectively to disseminate the contributed method of moments. Increasing the transparency of the methodology appeared much later; for example, see Charlier and Wicksell (1924), Pollard (1934), Cohen (1967), among others.

The concerned mixed normality could be practically described as a sum of weighted normal random variables, but the weight is stochastic following the zero-one Bernoulli variability. We can write $x = Iu + (1 - I)v$ and assign that u is normally distributed with mean μ_1(or α) and standard deviation σ_1(or σ), v is also normally distributed with mean μ_2(or β) and standard deviation σ_2(or δ), but I is only binary with $\Pr(I = 0) = \lambda$ versus $\Pr(I = 1) = 1 - \lambda$; the three random variables have zero stochastic dependence. Let $\phi(z)$ denote the standard normal probability density function $\frac{1}{\sqrt{2\pi}}\exp(-\frac{z^2}{2})$. It follows that a convenient formula of x's probability density is

$$f(x) = \frac{\lambda}{\sigma_1}\phi(\frac{x-\mu_1}{\sigma_1}) + \frac{1-\lambda}{\sigma_2}\phi(\frac{x-\mu_2}{\sigma_2}) \text{ or } \frac{\lambda}{\sigma}\phi(\frac{x-\alpha}{\sigma}) + \frac{1-\lambda}{\delta}\phi(\frac{x-\beta}{\delta}). \qquad (1)$$

When a random sample of x is available as $\{x_t\}_1^N$, equating the five sample moments $(m_1, m_2, \cdots, m_5)$ to the corresponded ordinal moment integrals could generate a nonlinear and simultaneous equation system with parameters as the variables. For simplicity of symbols, the second set of parameters $(\alpha, \beta, \sigma, \delta, \lambda)$ is preferred in displaying the equation system:

$$\begin{aligned}
m_1 =& \lambda\alpha + (1-\lambda)\beta,\\
m_2 =& \lambda(\alpha^2 + \sigma^2) + (1-\lambda)(\beta^2 + \delta^2),\\
m_3 =& \lambda\alpha(\alpha^2 + 3\sigma^2) + (1-\lambda)\beta(\beta^2 + 3\delta^2),\\
m_4 =& \lambda(\alpha^4 + 6\alpha^2\sigma^2 + 3\sigma^4) + (1-\lambda)(\beta^4 + 6\beta^2\delta^2 + 3\delta^4),\\
m_5 =& \lambda\alpha(\alpha^4 + 10\alpha^2\sigma^2 + 15\sigma^4) + (1-\lambda)\beta(\beta^4 + 10\beta^2\delta^2 + 15\delta^4). \qquad (2)
\end{aligned}$$

*Shine-In Quantitative Research, Irvine, California

When the equation relationship of m_6 is placed to the above system, the method of moments would require a different research process to increase the performance of the parameter estimation. The purpose of this paper is to present an example for imitating the sketched process and related decisions. For compatibility of technical contents, and other familiar benefits, the computations in this paper mostly apply Pearson's first sample of 1000 biometric measurements–breadth of forehead of crabs. It should be concerned that since 1970s the normal mixture model has evolved significantly, via the use of regression functions or/and generalized normal distributions to more complicated data. For examples of this advancement, in part, the reader is referred to Quandt and Ramsey (1978), Ball and Torous (1983), Cosslett and Lee (1985), McLachlan and Peel (2000), Dai et al. (2018).

This paper is planned as follows. Section 2 performs an interpretation on Pearson's equation finding and parameter estimates. Section 3 exhibits an adverse consequence of applying the measurement m_6 into the modified estimation. The problem is diminished in Section 4, by changing the model to a three-component specification. Section 5 contains the summary and conclusion of the research.

2. Pearson's Estimator and Application

Pearson demonstrated that a nonic polynomial equation could indirectly solve the equation system of his moment estimator. The equation's argued parameter (variable) was selected after implementing a particular transform of three parameters, which causes that the proper roots in solution should be negative.

In modern notation, especially with the sample cumulants (k_3, k_4, k_5), the nonic equation is written

$$a_9 p^9 + a_8 p^8 + \cdots + a_2 p^2 + a_1 p + a_0 = 0,$$

where the ten coefficients are as follows:

$$\begin{aligned}
&a_9 = 24,\\
&a_8 = 0,\ a_7 = 84k_4,\ a_6 = 36k_3^2,\ a_5 = 9(8k_3k_5 + 10k_4^2),\\
&a_4 = 6(74k_3^2k_4 - 3k_5^2),\ a_3 = 9(32k_3^4 - 12k_3k_4k_5 + 3k_4^3),\\
&a_2 = -9(7k_3^2k_4^2 + 8k_3^3k_5),\ a_1 = -96k_3^4k_4,\ a_0 = -24k_3^6.
\end{aligned} \tag{3}$$

It is helpful to review the following cumulant-to-moment conversions

$$\begin{aligned}
k_3 &= m_3 - 3m_1m_2 + 2m_1^3,\\
k_4 &= m_4 - 3m_2^2 - 4m_1m_3 + 12m_1^2m_2 - 6m_1^4,\\
k_5 &= m_5 - 5m_1m_4 - 10m_2m_3 + 20m_1^2m_3 + 30m_1m_2^2 - 60m_1^3m_2 + 24m_1^5.
\end{aligned} \tag{4}$$

Notice that the sum of coefficients in each case equation is zero. These three conversion formulas reveal that directly deriving the nonic equation from the moment conditions was a very difficult task; Pearson applied both raw and central moments to reach his classic solution.

Through laborious calculations, Pearson obtained the first five sample moments:

$$m_1 = 16.799, m_2 = 304.923, m_3 = 5831.759, m_4 = 116061.435, m_5 = 2385609.719;$$

his notation to m_j was μ'_j. He found three real roots: -8.757, -6.724, 4.170, which after contemporary computation are improved to -8.51646, -6.72225, 4.16932. Pearson delineated that the first root is better than the second, since the sample's sixth central moment (his notation μ_6) is equal to 177004 approximately, and the two negative roots have the respective predictions 188099 and 192446. It is believed that Pearson was constrained by the calculation of higher sample moments for providing stronger evidence to support the aforementioned decision making.

3. Taking m_6 and Its Equation into Computation

To examine the parameter estimates retrievable from the two algebraic roots, with the calculated moments, it is convenient to use the abbreviation θ as $(\mu_1, \mu_2, \sigma_1, \sigma_2, \lambda)$. The two vector of parameter estimates in upper precision are written

$$\hat{\theta}_1 = (13.39801, 19.30311, 4.51164, 3.10942, 0.42406),$$

$$\hat{\theta}_2 = (14.37198, 19.56875, 4.77025, 2.87726, 0.53298).$$

They could individually make predictions on the sixth moment and the sixth central moment. The outcome is (50399505.747, 189500.094) by $\hat{\theta}_1$ and (50402599.907, 192420.887) by $\hat{\theta}_2$. In contrast to the corresponded sample averages (50392382.883, 181943.441), the acceptance of $\hat{\theta}_1$ is better supported.

The above decision could be further agreed, if the moment estimator of solving the equation system of $\{m_1, m_2, m_3, m_4\}$ and m_6 has only one feasible solution. Switching the notation, the new moment equation is

$$m_6 = \lambda[\alpha^6 + 15(\alpha^4\sigma^2 + 3\alpha^2\sigma^4 + \sigma^6)] + (1-\lambda)[\beta^6 + 15(\beta^4\delta^2 + 3\beta^2\delta^4 + \delta^6)]. \quad (5)$$

However, the modified equation system has empty solution, maybe because the largeness of $m_6 = 50392382.883$ has affected the process of iterative computation. The four analogous exchanges between m_6 and one of $\{m_1, m_2, m_3, m_4\}$ have the same problem of empty solution. The author has conjectured that Pearson's method of deriving the nonic equation is useful in distinguishing the common source of the computational puzzle. This conjecture is posed for future research, in part because of anticipating a more practical analysis.

4. The Three-Component Model

In this section an interesting perspective is inquired. If it is possible to find a similar model as well as its major moment estimator, will the former problem of two roots and/or empty solution disappear? To depict this expectation, the function form and parameterization of $f(x)$ is adjusted to

$$g(x) = \frac{\lambda}{\sigma}\phi(\frac{x-\mu+\gamma}{\sigma}) + \frac{\omega}{\sigma}\phi(\frac{x-\mu}{\sigma}) + \frac{1-\lambda-\omega}{\sigma}\phi(\frac{x-\mu-\gamma}{\sigma}). \quad (6)$$

Subsequently, the moment equation system is changed to

$$\begin{aligned}
m_1 =& \lambda(\mu-\gamma) + \omega\mu + (1-\lambda-\omega)(\mu+\gamma), \\
m_2 =& \lambda[(\mu-\gamma)^2 + \sigma^2] + \omega(\mu^2+\sigma^2) + (1-\lambda-\omega)[(\mu+\gamma)^2 + \sigma^2], \\
m_3 =& \lambda(\mu-\gamma)[(\mu-\gamma)^2 + 3\sigma^2] + \omega\mu(\mu^2 + 3\sigma^2) + (1-\lambda-\omega)\cdots, \\
m_4 =& \lambda[(\mu-\gamma)^4 + 6(\mu-\gamma)^2\sigma^2 + 3\sigma^4] + \omega(\mu^4 + 6\mu^2\sigma^2 + 3\sigma^4) + \cdots, \\
m_5 =& \lambda(\mu-\gamma)[(\mu-\gamma)^4 + 10(\mu-\gamma)^2\sigma^2 + 15\sigma^4] + \omega\mu(\mu^4 + \cdots) + \cdots. \quad (7)
\end{aligned}$$

When solving the above equation system for the same data set, the problem of two roots occurs again. The two vector of estimates for $(\mu, \gamma, \sigma, \lambda, \omega) \equiv \vartheta$ are expressed as

$$\hat{\vartheta}_1 = (13.56181, 5.77191, 3.09148, 0.07435, 0.29045),$$

$$\hat{\vartheta}_2 = (10.91176, 7.79145, 3.30948, 0.00426, 0.23646).$$

The simple prediction method suggests accepting $\hat{\vartheta}_1$ and rejecting $\hat{\vartheta}_2$, whereas the respective predictions are 185987.432 and 193949.263, in contrast to the central moment $c_6 = 181943.441$, an average from the sample. Notice that $\hat{\vartheta}_1$'s prediction is better than that of $\hat{\theta}_1$

When m_5 is replaced by m_6, the moment method still have two roots, unlike the empty solution in the two-component model. Moreover, the replacement is permissible for three more cases, with m_4, m_3, m_2 but not possible with m_1. Though m_1 cannot be replaced by m_6, the research emerges to find that if m_1 is replaced by m_6/m_1^5 then there are two roots, like the other three cases. Notice that m_1 and m_6/m_1^5 has approximate order and the implemented equation is

$$\frac{m_6}{m_1^5} = \frac{\lambda[(\mu-\gamma)^6 + 15(\mu-\gamma)^4\sigma^2 + 45(\mu-\gamma)^2\sigma^4 + 15\sigma^6] + \cdots}{[\lambda(\mu-\gamma) + \omega\mu + (1-\lambda-\omega)(\mu+\gamma)]^5}. \tag{8}$$

The three-component model was further estimated with a variation of replacement. The modified estimation of replacing m_5 by c_6, instead of by m_6, has two roots in solution. Moreover, when the replacement was advanced to the next higher-order central moments persistently, the exchange with c_{10} found single root. The vector of retrieved parameter estimates is

$$\breve{\vartheta} = (18.52952, 8.41952, 3.23462, 0.21032, 0.78489).$$

Thus, the numerical use of extra higher moments should be guarded if possible.

Relied on $\hat{\theta}_1$ and $\hat{\vartheta}_1$, the author continued to realize the respective maximum likelihood estimation of the two models. The interesting inference made in this section is a consequence of carefully examining the moment-based and the likelihood-based estimates listed in Table 1, Table 2A and 2B.

5. Summary and Conclusion

This research has two interesting implications: (i) Regarding the classical method of moments, if the two-component mixture model has encountered the problem of multiple roots or empty solution, one should consult analogous estimation on the three-component model for remedy. (ii) Pearson's method of deriving the nonic equation was conjectured to be able to diagnose a category of algebraic problems in modern application of the method of moment. In addition, while editing this paper, the author has further reviewed the expositions in Lee (2016) and Weisstein (2019) to recognize that the three-component model's moment estimator could be formulated into a quartic equation under a very plausible condition.

Table 1. Method-of-Moments Estimates of $g(x)$ with m_j vs m_6 Replacement

m_j	$\hat{\mu}$	$\hat{\gamma}$	$\hat{\sigma}$	$\hat{\lambda}$	$\hat{\omega}$	$\mathcal{L}(\hat{\vartheta})$
m_5	13.5145	5.8373	3.0801	0.0711	0.2951	−2953.0408
	11.0806	7.6940	3.2778	0.0057	0.2454	−2956.3255
m_4	13.4455	5.9187	3.0713	0.0668	0.2998	−2953.0508
	11.3272	7.5321	3.2410	0.0083	0.2570	−2956.1164
m_3	13.3226	6.0394	3.0679	0.0602	0.3039	−2953.1160
	11.7174	7.2595	3.1947	0.0137	0.2726	−2956.1164
m_2	12.8662	6.4089	3.0918	0.0418	0.3027	−2953.6032
	12.5901	6.6178	3.1134	0.0332	0.2977	−2954.0133
m_1	13.6229	5.7104	3.0936	0.0780	0.2877	−2953.0883
	10.8426	7.8511	3.3156	0.0034	0.2345	−2956.4699

Note:

1. The symbol $\mathcal{L}(\hat{\vartheta})$ stands for the log-likelihood at $\vartheta = \hat{\vartheta}$.
2. The standard errors of the method-of-moments estimates could be determined according to a standard formula of the generalized method of moments. For related interpretations of the asymptotic theory, one could refer to Schmidt (1982) and Hansen (1982), among others.

Table 2A. The Standard Method-of-Moments Estimates

pdf	$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\sigma}_1$	$\hat{\sigma}_2$	$\hat{\lambda}$	$\mathcal{L}(\hat{\theta})$
$f(x)$	13.3981	19.3031	4.5116	3.1094	0.4241	−2953.9672
	14.3720	19.5688	4.703	2.8773	0.5330	−2954.3795
pdf	$\hat{\mu}$	$\hat{\gamma}$	$\hat{\sigma}$	$\hat{\lambda}$	$\hat{\omega}$	$\mathcal{L}(\hat{\vartheta})$
$g(x)$	13.5618	5.7719	3.0915	0.0744	0.2905	−2953.0544
	10.9118	7.7915	3.3095	0.0043	0.2365	−2956.3316

Table 2B. The Maximum Likelihood Estimates

pdf	$\tilde{\mu}_1$	$\tilde{\mu}_2$	$\tilde{\sigma}_1$	$\tilde{\sigma}_2$	$\tilde{\lambda}$	$\mathcal{L}(\tilde{\theta})$
$f(x)$	13.5609 (2.8546)	19.2697 (0.6260)	4.5782 (0.9231)	3.1548 (0.4050)	0.4328 (0.2731)	−2953.8820
pdf	$\tilde{\mu}$	$\tilde{\gamma}$	$\tilde{\sigma}$	$\tilde{\lambda}$	$\tilde{\omega}$	$\mathcal{L}(\tilde{\vartheta})$
$g(x)$	13.4476 (0.6092)	5.8720 (0.5240)	3.0924 (0.1758)	0.0682 (0.0296)	0.2928 (0.0391)	−2953.0326

Note:

1. The maximum likelihood estimation of this research is completed with the Gretl package; see the optimization techniques and commands documented in Cottrell and Lucchetti (2018).
2. The vector of standard errors, below the estimates, is computed from inverting the numerical Hessian matrix.

Table A-1. Attached Frequency Counts–The First Sample Estimated in Pearson (1894)

x_t	Freq.	x_t	Freq.	x_t	Freq.	x_t	Freq.
1	1	8	19	15	54	22	47
2	3	9	20	16	74	23	43
3	5	10	25	17	84	24	24
4	2	11	40	18	86	25	19
5	7	12	31	19	96	26	9
6	10	13	60	20	85	27	5
7	13	14	62	21	75	28	0

Note:

1. The frequency of $x_t = 29$ being only one should be added to the table to have $N = 1000$.
2. The sample mean and variance are 16.799 and 22.717. The sample skewness and kurtosis are -0.498 and 3.055.

REFERENCES

Ball, C. A. and W. N. Torous (1983): "A Simplified Jump Process for Common Stock Returns," *Journal of Financial and Quantitative Analysis,*18, 53–65.

Charlier, C. V. L. and S. D. Wicksell (1924): "On the Dissection of Frequency Functions," *Arkiv for Mathematik, Astronomi och Fysik,*18.

Cohen, A. C. (1967): "Estimation in Mixtures of Two Normal Distributions," *Technometrics,* 9, 15–28.

Cosslett, S. R. and L.-F. Lee (1985): "Serial Correlation in Latent Discrete Variable Models," *Journal of Econometrics*, 27, 79–97.

Cottrell, A. and R. Lucchetti (2018): *Gretl Users Guide*. Gnu Regression, Econometrics and Time-series Library.

Dai, L., Yin, J., Xie, Z. and Wu, L. (2018): "Robust Variable Selection in Finite Mixture of Regression Models Using the t Distribution," *Communications in Statistics-Theory and Methods*, 48, 5370–5386.

Hansen, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica,* 50, 1029–1054.

Lee, R.-S. (2016): "A Regulative Normality Application of Classical Moment Estimators," Conference paper presented at a joint meeting of the Taiwan Economic Association and Chinese Economic Association in North America.

McLachlan, G. and D. Peel (2000): *Finite Mixture Models*. New York: John Wiley & Sons, Inc.

Pearson, K. (1894): "Contributions to the Mathematical Study of Evolution," *Philosophical Transactions of the Royal Society of London,* Series A, 185, 71–110.

Pollard, H. S. (1934): "On the Relative Stability of the Median and Arithmetic Mean, with Particular Reference to Certain Frequency Distributions Which Can Be Dissected into Normal Distributions," *Annals of Mathematical Statistics,* 5, 227–262.

Quandt, R. E., and J. B. Ramsey (1978): "Estimating Mixtures of Normal Distributions and Switching Regressions," *Journal of the American Statistical Association*, 73, 730–738.

Schmidt, P. (1982): "An Improved Version of the Quandt-Ramsey MGF Estimator for Mixtures of Normal Distributions and Switching Regressions," *Econometrica*, 50, 501–516.

Weisstein, E. W. (2019): "Quartic Equation," *MathWorld–A Wolfram Web Resource*.