THE METHOD OF MOMENTS APPLIED TO A MIXTURE OF TWO EXPONENTIAL DISTRIBUTIONS[1]

By Paul R. Rider

Aeronautical Research Laboratories, Wright-Patterson Air Force Base

The dissection[2] of mixed frequency distributions is often very complicated ([6], pp. 152–158). This is certainly true for a mixture of two normal distributions, which was studied by Karl Pearson [9] in perhaps the earliest investigation of the dissection problem. Pearson was led to an equation of ninth degree, the setting up and solution of which involved a tremendous amount of calculation. This calculation could doubtless be performed rather expeditiously today if a high-speed computer is available, but when his paper was published (1894) it was extremely laborious.

In the present paper a mixture of two exponential distributions is considered.

In experiments in life testing it has been found that the life, x, may often be reasonably described by a probability density function of the form

$$f(x) = \theta^{-1}e^{-x/\theta}, \qquad \theta > 0, \quad 0 \leqq x < \infty. \tag{1}$$

For example, there seems to be evidence, [2], [3], and [4], that the lives of electron tubes or the time intervals between failures of electronic systems are random variables having, at least to a first approximation, the density function given by (1). The parameter θ is the mean life or the mean time between failures.

Suppose now that two populations of the type (1), with parameters θ_1 and θ_2 respectively, have been mixed in the unknown proportions p and $1 - p$. The resulting probability density function is

$$f(x) = p\theta_1^{-1}e^{-x/\theta_1} + (1 - p)\theta_2^{-1}e^{-x/\theta_2}. \tag{2}$$

A simple method of estimating the three parameters p, θ_1, θ_2 from the first three moments of a sample is derived.

Mendenhall and Hader [8] have treated a related problem. They considered the question of estimating the parameters of a population obtained by mixing two exponential failure time distributions in unknown proportions, the population model being based upon a sample censored at a fixed test-termination time. They assumed that each unit of the population conceptually bears a tag that indicates the component, or subpopulation, from which it came. This information is available only after failure has occurred. Weiner [10] has also studied the problem under consideration in the present paper and has given maximum likelihood estimators of the parameters. He states that it is imperative to have the

Received April 10, 1959; revised August 12, 1960.

[1] The author of this paper is deeply indebted to Professor Wassily Hoeffding for valuable criticisms and suggestions.

[2] "Dissection" here means point estimation of the parameters in a parametric mixture model.

calculations of the estimates programmed on a high-speed digital computer. His paper gives formulas for the variances of the maximum likelihood estimates.

Gumbel [5] has discussed the general dissection problem and has shown how the method of moments can be used to estimate the parameters of a mixed distribution. He applied the method to mixed exponential and mixed Poisson distributions, but assumed that the proportion p is known.

Let m_1', m_2', m_3' denote the moments, about zero, of a random sample from (2), and let p^*, θ_1^*, θ_2^* denote the estimators of p, θ_1, θ_2 respectively, obtained by the method of moments. Then

$$p^*\theta_1^* + (1 - p^*)\theta_2^* = m_1', \tag{3}$$

$$p^*\theta_1^{*2} + (1 - p^*)\theta_2^{*2} = \tfrac{1}{2} m_2', \tag{4}$$

$$p^*\theta_2^{*3} + (1 - p^*)\theta_2^{*3} = \tfrac{1}{6} m_3'. \tag{5}$$

From (3) it is found that

$$p^* = (m_1' - \theta_2^*)/(\theta_1^* - \theta_2^*). \tag{6}$$

Substituting this expression for p^* in (4) and (5) leads to the following two equations:

$$(m_1' - \theta_2^*)(\theta_1^* + \theta_2^*) = \tfrac{1}{2} m_2' - \theta_2^{*2}, \tag{7}$$

$$(m_1' - \theta_2^*)(\theta_1^{*2} + \theta_1^*\theta_2^* + \theta_2^{*2}) = \tfrac{1}{6} m_3' - \theta_2^{*2}. \tag{8}$$

Equation (7) may be solved for $\theta_i^*(i = 1 \text{ or } 2)$, the solution being

$$\theta_i^* = (\tfrac{1}{2} m_2' - m_1'\theta_j^*)/(m_1' - \theta_j^*), \tag{9}$$

where $j = 2$ or 1 according as $i = 1$ or 2. When θ_i^* from (9) is substituted in (8) and the result simplified, the equation for θ_j^* is

$$6(2\, m_1'^2 - m_2')\theta_j^{*2} + 2(m_3' - 3\, m_1'm_2')\theta_j^* + 3\, m_2'^2 - 2\, m_1'm_3' = 0. \tag{10}$$

The two roots of this quadratic are θ_1^* and θ_2^*, it being immaterial which root is designated θ_1^* and which θ_2^*. That is, the estimate p^* of the proportion p, obtained by substituting θ_1^* and θ_2^* respectively in (6), will refer to the component having θ_1 as parameter and $1 - p^*$ will refer to the other component.

It is possible that the roots of (10) will not both be positive, or even real. For example, if every observation in a sample were equal to some constant $c > 0$, then it would follow that $m_1' = c$, $m_2' = c^2$, $m_3' = c^3$ and (10) could be reduced to the form

$$c^2[6(\theta - \tfrac{1}{3}c)^2 + \tfrac{1}{3}c^2] = 0, \tag{11}$$

the roots of which are imaginary. From continuity considerations it is seen that this may occur with positive probability.

However, if $\theta_1 \neq \theta_2$, the proposed estimators are consistent and the probability that $\theta_1^* > 0$, $\theta_2^* > 0$, $0 \leqq p^* \leqq 1$ approaches 1 as n tends to infinity.

This follows from the facts that, in this case, the estimators, regarded as functions of (m_1', m_2', m_3'), are continuous at the point (μ_1', μ_2', μ_3'), where the μ_i' are the population moments, and that $\theta_1^* > 0$, $0 \leqq p^* \leqq 1$ if (m_1', m_2', m_3') is sufficiently close to (μ_1', μ_2', μ_3').

If $\theta_1 = \theta_2 = \theta$, the behavior of the estimators changes radically. For in this case $\mu_1' = \theta$, $\mu_2' = 2\theta^2$, $\mu_3' = 6\theta^3$, and therefore

$$2\mu_1'^2 - \mu_2' = \mu_3' - 3\mu_1'\mu_2' = 3\mu_2'^2 - 2\mu_1'\mu_3' = 0.$$

Hence the three coefficients in the quadratic equation (10), multiplied by $n^{\frac{1}{2}}$, are normally distributed in the limit as n approaches infinity, with zero means and finite and positive variances. This can be shown to imply that the roots θ_1^* and θ_2^* have no constant limits in probability and their imaginary parts do not become negligibly small as n increases. In particular, the estimators are not consistent in this case.

Some discussion of the variances of the three estimators is of course in order. It seems that the calculation of these variances, even in asymptotic form, would not only be a difficult task but would lead to somewhat complicated expressions. However to simplify matters and to give some idea of the reliability of the estimators θ_1^* and θ_2^*, it will be temporarily assumed that p is known. With this assumption, only two sample moments are needed to estimate θ_1 and θ_2. Thus, from equations (3) and (4) it is found that

$$\theta_1^* = m_1' \pm (q/2p)^{\frac{1}{2}}(m_2' - 2\, m_1'^2)^{\frac{1}{2}}, \tag{12}$$

where, as usual, $q = 1 - p$. The upper sign is used if $\theta_1 \geqq \theta_2$, the lower sign if $\theta_1 \leqq \theta_2$. The estimator θ_2^* may be found by interchanging p and q and replacing the sign $\pm$ by $\mp$. The first pair of estimators is consistent when $\theta_1 \geqq \theta_2$, the second pair when $\theta_1 \leqq \theta_2$. Thus, here the estimators are consistent when $\theta_1 = \theta_2$, but in this case the rate of approach to the limit is $n^{-\frac{1}{4}}$ as compared with $n^{-\frac{1}{2}}$ for $\theta_1 \neq \theta_2$. Also, if $\theta_1 = \theta_2$, the probability that the estimators are real does not approach 1 as n approaches infinity, although the imaginary parts converge to zero in probability. It is assumed that $\theta_1 \neq \theta_2$. If it is not known whether $\theta_1 > \theta_2$ or $\theta_1 < \theta_2$, then it is also not known which pair of estimators is consistent, that is, which pair may be expected to be close to the true value when the sample is large. This admittedly is a real shortcoming of the method.

The asymptotic variance of θ_1^* may be found by the use of a formula given by Cramér ([1], p. 354, (27.7.3)). In the notation of the present paper, this formula is

$$\operatorname{var} \theta_1^* = \mu_2(m_1') \left(\frac{\partial \theta_1^*}{\partial m_1'}\right)^2 + 2\mu_{11}(m_1', m_2') \frac{\partial \theta_1^*}{\partial m_1'} \frac{\partial \theta_1^*}{\partial m_2'} + \mu_2(m_2') \left(\frac{\partial \theta_1^*}{\partial m_2'}\right)^2; \tag{13}$$

here $\mu_2(m_1')$ and $\mu_2(m_1')$ are the variances of m_1' and m_2' respectively, $\mu_{11}(m_1', m_2')$ is the covariance of these two moments, and the partial derivatives are to be evaluated at the point

$$m_1' = p\theta_1 + q\theta_2, \qquad m_2' = 2(p\theta_1^2 + q\theta_2^2). \tag{14}$$

The values of the coefficients of the partial derivatives in (13) may be obtained by using formulas given by Kendall [7, p. 206]. It is found that

$$\mu_2(m_1') = n^{-1}[(2p - p^2)\theta_1^4 - 2pq\theta_1^2\theta_2^2 + (2q - q^2)\theta_2^4], \tag{15}$$

$$\mu_{11}(m_1', m_2') = 2n^{-1}[(3p - p^2)\theta_1^3 - pq\theta_1^2\theta_2 - pq\theta_1\theta_2^2 + (3q - q^2)\theta_2^3], \tag{16}$$

$$\mu_2(m_2') = 4n^{-1}[(6p - p^2)\theta_1^4 - 2pq\theta_1^2\theta_2^2 + (6q - q^2)\theta_2^4]. \tag{17}$$

If $\theta_1 > \theta_2$ (in which case the upper sign holds in (12)) the partial derivatives needed are

$$\frac{\partial\theta_1^*}{\partial m_1'} = 1 - \frac{2^{\frac{1}{2}}q^{\frac{1}{2}}m_1'}{p^{\frac{1}{2}}(m_2' - 2m_1'^2)^{\frac{1}{2}}}, \tag{18}$$

$$\frac{\partial\theta_1^*}{\partial m_2'} = \frac{q^{\frac{1}{2}}}{2^{\frac{3}{2}}p^{\frac{1}{2}}(m_2' - 2m_1'^2)^{\frac{1}{2}}}. \tag{19}$$

At the point (14), these derivatives have the values

$$\frac{\partial\theta_1^*}{\partial m_1'} = \frac{-\theta_2}{p(\theta_1 - \theta_2)}, \qquad \frac{\partial\theta_1^*}{\partial m_2'} = \frac{1}{4p(\theta_1 - \theta_2)}. \tag{20}$$

If $\theta_1 < \theta_2$ (in which case the lower sign holds in (12)), the signs of the fractions on the right-hand sides of (18) and (19) are changed, and we again obtain equations (20).

Substituting (15), (16), (17), (20) in (13) and simplifying give [1, p. 366] for the variance of the asymptotic distribution of θ_1^*,

$$\frac{1}{4np^2(\theta_1 - \theta_2)^2}[p(6 - p)\theta_1^4 - 4p(3 - p)\theta_1^3\theta_2 + 2p(5 - 3p)\theta_1^2\theta_2^2 - 4p(1 - p)\theta_1\theta_2^3 + (1 - p^2)\theta_2^4]. \tag{21}$$

The variance of the asymptotic distribution of θ_2^* may be obtained by replacing p by q and interchanging θ_1 and θ_2 in (21).

It is the personal opinion of the author that data should not be assumed to have come from a mixed exponential distribution until it has been determined that they have not come from a simple exponential distribution $(1/\theta) \exp(-x/\theta)$. That is, the parameter θ of this distribution should be estimated, following which a chi-square test should be made to see whether the data conform to this distribution. If the hypothesis that they came from a simple exponential is rejected, a mixed exponential population may be assumed.

Of course the chi-square test may give the wrong conclusion, in which case it would be impossible to find, by the method under discussion, an estimate of θ. Even if the population is mixed and θ_1 and θ_2 are nearly equal, it might be difficult to obtain valid estimates of them. Unless further research reveals some way of remedying the shortcomings of the estimators they are not recommended for practical purposes.

Results for other types of mixed populations (Poisson, positive and negative binomial, and Weibull) have been obtained and will be reported later. However, the estimators seem to be subject to the same deficiencies as the estimators treated in this paper.

REFERENCES

[1] Harald Cramér, *Mathematical Methods of Statistics*, Princeton University Press, 1946.

[2] D. J. Davis, "An analysis of some failure data," *J. Amer. Stat. Assn.*, Vol. 47 (1952), pp. 113–150.

[3] Benjamin Epstein, "Stochastic models for length of life," *Proceedings of the Statistical Techniques in Missile Evaluation Symposium* held at Virginia Polytechnic Institute, 5–8 August, 1958, (Boyd Harshbarger, Editor), pp. 69–84.

[4] Benjamin Epstein and Milton Sobel, "Life testing," *J. Amer. Stat. Assn.*, Vol. 48 (1953), pp. 486–502.

[5] E. J. Gumbel, "La dissection d'une répartition," *Annales de l'Université de Lyon*, Series 3, Section A, Fascicule 2, (1940), pp. 39–51.

[6] A. Hald, *Statistical Theory with Engineering Applications*, John Wiley and Sons, New York, 1952.

[7] Maurice G. Kendall, *The Advanced Theory of Statistics*, Charles Griffin and Co., London, Vol. 1, 1948.

[8] William Mendenhall and R. J. Hader, "Estimation of parameters of mixed exponentially distributed failure time distributions from censored life test data," *Biometrika*, Vol. 45 (1958), pp. 504–520.

[9] Karl Pearson, "Contributions to the mathematical theory of evolution," *Philos. Trans. Roy. Soc. London*, Vol. 185A (1894), pp. 71–110.

[10] Sidney Weiner, "Samples from mixed-exponential populations," Mimeographed paper, ARINC Research Corporation, Washington, D. C. Research under Contract NObsr-64508.