SMALL SAMPLE RESULTS FOR A DISCRIMINANT FUNCTION ESTIMATED FROM A MIXTURE OF TWO INVERSE GAUSSIAN DISTRIBUTIONS

R.K. Amoh, BSc (Econ), MSc, PhD
Department of Mathematics, U.S.T., Kumasi

ABSTRACT

Estimation of a discriminant function on the basis of a small sample size from a mixture of two inverse Gaussian distributions is considered. Its performance is investigated by a series of simulation experiments. The relative efficiency of the mixture and classified discrimination procedures are evaluated from the simulation results and compared with available asymptotic relative efficiency results.

Keywords: *Discriminant Function; Estimation from Mixtures, Mixture of Inverse Gaussian Populations; Relative Efficiency.*

INTRODUCTION

The inverse Gaussian distribution as the first passage time distribution of a Brownian motion process was first derived by Schrödinger in [14]. Later, Wald [16] also derived it as a limiting form for the distribution of sample size in a sequential probability ratio test. The distribution is known variously as Wald's distribution, Gaussian first passage time distribution and the first passage time distribution of Brownian motion with positive drift.

From the viewpoint of statistics this distribution is relatively new and might be more appropriately called Tweedie's distribution since it had remained almost unnoticed until Tweedie [15] investigated its basic characteristics. He established some important statistical properties and drew analogies between it and the normal distribution. Tweedie proposed the name inverse Gaussian because of the inverse relationship between its cumulant generating function and that of the normal distribution.

The inverse Gaussian distribution has been found useful in applications connected to various fields, including the physical and biological sciences, economics and management. It is a strong candidate for modelling data that have been generated as first passage times of Wiener or Brownian process or data from unknown positively skewed long-tailed distribution. Because of this, it is a competitor to distributions like the gamma, log-normal, negative exponential and Weibull for modelling.

In Amoh [1] the inverse Gaussian is fitted to survival data from Kalbfleisch and Prentice [10] and failure data from Proschan [13]. Further examples of the use of the inverse Gaussian distribution can be found in Folks and Chhikara [7] and Chhikara and Folks [6].

Mixtures of life distributions occur when two different causes of failure are present each with the same parametric form of life distribution. Tweedie in the discussion of the review paper by Folks and Chhikara [7] gave account of observed data that could be fitted quite well by a mixture of two inverse Gaussian distributions.

In this paper we shall investigate the performance of a discriminant function estimated from a mixture of two inverse Gaussian distributions when sample size is small. Studies in this area have been undertaken by O'Neill [12] and Ganesalingam and McLachlan [8]. In all these studies the underlying populations are assumed to be normal. Ganesalingam and McLachlan [9] found from simulation experiments that the mixture discriminant function performed satisfactorily although the maximum likelihood estimates of the parameters estimated from the mixed samples were poor. O'Neill [12] and Ganesalingam and McLachlan [8] present only asymptotic results.

Apart from Amoh [3] and Amoh and Kocherlakota [4] we are not aware of any literature relating to estimation of a discriminant function from mixtures of populations which are not normal. Because of the similarity of the inverse Gaussian to the other positively skewed distributions we hope that the results will apply in a long measure to these alternative distributions.

CLASSIFICATION RULES

Suppose we have two distinct inverse Gaussian populations Π_1 and Π_2 with means μ_1 and μ_2 respectively, and a common known shape parameter λ. An observation X belonging to Π_i $(i = 1,2)$ has the probability density function

$$f_i(x) = \left(\frac{1}{2\pi x^3}\right)^{1/2} \exp\{-\lambda(x-\mu_i)^2/2\mu_i^2 x\},$$

$$x > 0,\ \mu_i > 0,\ \lambda > 0, \qquad (2.0)$$

We say $$X \sim IG(\mu_i, \lambda) \qquad (2.1)$$

R.K. Amoh

When all the parameters of the populations are known, we can construct an optimal discriminant function $L_o(x)$, based on the known parameters, for assigning a new observation X, of unknown origin, to Π_1 or Π_2. According as the optimal discriminant function

$$L_o(x) = ax + b \qquad (2.2)$$

is less or greater than zero, we assign the new observation X to Π_1 or Π_2 respectively; where

$$a = \lambda(1/\mu_1^2 - 1/\mu_2^2)/2,$$
$$b = -2a\ (1/\mu_1 + 1/\mu_2) + \log(q/p)$$

and p is the prior probability of the observation coming from with Π_1, $q = 1 - p$. We refer to this procedure as the optimal discrimination procedure. Details of this are provided in Amoh [3].

As it often happens, the means of the populations may not be known. In this case, if separate samples are available from each of the populations, we can obtain a classified discriminant function $L_c(x)$ based on estimates of the means from the classified samples. If there are n_i initial observations available from Π_i, with $n_1 + n_2$, the resulting classified discriminant function is

$$L_c(x) = \hat{a}x + b \qquad (2.3)$$

Where

$$\bar{a} = \lambda(1/\mu_1^2 - 1/\bar{\mu}_2^2)\}/2,$$
$$\bar{b} = -2\bar{a}\ (1/\mu_1 + 1/\bar{\mu}_2) + \log(\bar{q}/\bar{p})$$

with μ_i, μ_2 as maximum likelihood estimates of the means computed from the initial classified samples and $p = n_i/n$

However, if we have a situation where only a sample from a mixture of Π_1 and Π_2 is available, our interest is to construct a mixture discriminant function, $L_m(x)$ based on the mixed sample and investigate its performance relative to the optimal and classified discriminant functions. Suppose the initial unclassified observations are a sample of size n from a mixture distribution

$$f(x) = pf_1(x) + qf_2(x) \qquad (2.4)$$

where $f_i(x)$ is given by (2.1). The mixture discriminant function, L_m, is obtained by replacing the unknown parameters in (2.2) by their maximum likelihood estimates. Amoh [1] gives details of iterative procedures for obtaining maximum likelihood estimates of the parameters from a sample from a mixture of two inverse Gaussian distributions.

A brief description of the iterative procedure is as follows: For the model (2.4) the maximum likelihood estimates of **a** and **b** from the **n** unclassified observations $x_1, x_2, \ldots, x_n$ are obtained by maximizing the likelihood

$$L = \prod_{j=1}^{n} f(x_j)$$

The maximum likelihood estimates,

$$\hat{a} = \lambda(1/\mu_1^2 - 1/\mu_2^2)/2 \quad \text{and}$$

$$b = -2\hat{a}/(1/\mu_1 + 1/\mu_2) + \log(q/p)$$

where

$$p = 1 - q = \sum_{j=1}^{n} w_{1j}/n,\ \mu_i = \sum_{j=1}^{n} w_{1j}x_j / \sum_{j=1}^{n} w_{1j},\ (i=1,2)$$

and $w_{1j} = 1 - w_{2j} = [1 + \exp(\hat{a}x_j + b)]^{-1}$. We shall refer to this as the mixture discrimination procedure and Lm(x) is written as

$$L_m(x) = \hat{a}x + b \qquad (2.5)$$

PROBABILITIES OF MISCLASSIFICATION

Let e_{ij} (i = 1,2; j = o, c, m) be the probability of misclassifying an observation from Π_i by the discriminant function $L_j(x)$; with e_j as the corresponding total probability of misclassification. For $\mu_1 > \mu_2$ we have

$$e_{ij} = F$$
$$e_{2j} \ldots\ldots\ldots\ldots \text{ etc.} \qquad (3.1)$$

where α_j is given by

$$\frac{-b}{\hat{a}},\ -\frac{b}{\hat{a}},\ -\frac{b}{\hat{a}} \qquad \text{for } j = o, c, m$$

respectively and $F(\alpha_j, \mu_i, \lambda)$ is the cumulative distribution function (cdf) of the inverse Gaussian distribution with parameters μ_i and λ. The cdf corresponding to (1.1) is given by

$$F(x, \mu, \lambda) = \phi(\lambda/x)^{1/2}(x/\mu - 1) +$$

$$e^{2\lambda/\mu}\phi(-(\lambda/x)^{1/2}(x/\mu + 1)),\ x > 0 \qquad (3.2)$$

where $\phi(.)$ is the cdf of the standard normal.

SIMULATION EXPERIMENTS AND RESULTS

A series of simulation experiments were performed to investigate the performance of Lm relative to Lc and Lo for small samples. Twelve combinations of the parameters were taken:

$$\lambda = 4.0,\ 8.0,\ 16.0;\quad \mu_1 = 2.0,\ 6.0;$$
$$\mu_2 = 1.0 \text{ and } p = 0.25,\ 0.50.$$

For each combination of parameters, classified and mixture samples of size n = 40 were generated from the mixture distribution. For the parameter combinations

$\lambda=4.0$; $\mu_1=2.0, 6.0$; and $p=0.50$, further samples of size n = 100 were taken.

The following procedure was used for generating the samples: An observation, **u**, is generated from U(0,1). If $\mathbf{u} \leq p$, then μ_1 is used for generating the inverse Gaussian variate by the procedure described by Michael et al [11]; this observation belongs to the first component which is $IG(\mu_1, \lambda)$.
If $\mathbf{u} > \mathbf{p}$, then μ_2 is used to generate the inverse Gaussian variate and this is assumed to be a sample from the second component which is $IG(\mu_2, \lambda)$. This procedure is continued **n** times resulting in $\mathbf{n}_1$ observations identifiable from the first component and $\mathbf{n}_2$ from the second component.

The maximum likelihood estimates of **p**, $\boldsymbol{\mu}_1$ and μ_2 are obtained from the mixed sample by iterative procedure. The iteration is terminated by the following criterion: Define $\delta a = [a^{v+1} - a^v]$ and $\delta b = [b^{v+1} - b^v]$, where $(v + 1)$ is the number of iterations. If $v < 1000$ and $\delta a + \delta b < 10^{-5}$, then $p = p^{(v+1)}$ and $\mu_1 = \mu_1^{(v+1)}$; otherwise $p = p^{(1000)}$ and $\mu_i = \mu_i^{(1000)}$. A series of trials showed that the starting points did not have much effect on the final values of the estimates. When λ is large, all the trials yielded convergence in less than 1000 iterations. For example, when $\lambda = 16.0$, $\mu_1 = 2.0$, $\mu_2 = 1.0$ and $p = 0.25$, with starting values, $a_o = b_o = 2.0$, the average number of iterations necessary for convergence per trial, was 64.5. When λ and $|\mu_1 - \mu_2|$ are small, it takes more iterations to achieve convergence but only a few trials fail to converge in 1000 iterations. For $\lambda = 4.0$, $\mu_1 = 2.0$, $\mu_2 = 1.0$, $p = 0.25$ and starting values $a_o = b_o = 2.0$, the average number of iterations needed to obtain convergence per trial was 207.1 but only 3 percent of them fail to converge in 1000 iterations and for half of these $\delta a + \delta b$ was less than 3×10^{-4}. Changing the terminating criterion to $\delta a + b < 5 \times 10^{-4}$ does not affect the results significantly.
With less than 2 percent of the trials failing to converge in 1000 iterations, in cases where the distributions are poorly separated, does not present any problems with the simulation. To reduce the number of iterations, the actual parametric values of **a** and **b** were used as starting points.

The estimates of the parameters were obtained for classified samples, where the population of origin of each observation in each sample was known with certainty. The maximum likelihood estimates of the parameters from these samples, given by

$$\mu_i = x_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij},\ i=1,\ 2;\ p = \frac{n_i}{n}$$

were computed.

Using the parameters and the calculated estimates, the individual and total conditional probabilities of misclassification, as defined in (3.1), were evaluated for the completely classified and mixture discrimination procedures for each sample generated. These were averaged over 50 repetitions for each combination of parameters, which appeared enough for each combination of the parameters considered. Increasing the number of repetitions to 100 reduced the variances of the estimates by about 25 percent as expected but did not affect the estimates. In fact when λ is large, 30 repetitions were found to be quite adequate. The sample means of the individual and total conditional probabilities of misclassification are denoted by e_{ij} and e_j ($i = 1,2$; $j = m, c$) respectively. The corresponding optimal probabilities of misclassification, $\mathbf{e}_{10}$, $\mathbf{e}_{20}$ and $\mathbf{e}_0$ were also evaluated for each combination of parameters.

Table 1 shows the individual probabilities of misclassification for the three discrimination procedures for n = 40 and n = 100. The standard deviations for the conditional probabilities of misclassification are shown in parentheses.

We find that generally e_{1j}($j = m,c$) are closer to the corresponding optimal values than e_{2j}. Considering these conditional probabilities of misclassification as estimates of the optimal probabilities of misclassification, we observe that for small values of λ and $d = |\mu_1 - \mu_2|$ the estimates e_{1j} are poor with e_{2j} consistently exceeding e_{20}. This is not surprising since when λ and d are both small the components of the mixture population are not well-separated and hence it is very difficult to discriminate between them. When λ and d are large, e_{1j} are quite good estimates but although the estimates e_{2j} improve, they are still not good estimates of e_{20}. The variances associated with e_{ij}, are quite large and every e_{ij} lies within one standard deviation of the corresponding optimal value e_{io}. As expected, for every parameter combination, the standard deviation of e_{ic} is smaller than that of e_{im} since more

TABLE 1: Individual Probabilities of Misclassification (μ_2 = 1.0)

Parameters				Classification Procedures					
				Mixtures		Completely Classified		Optimal	
n	λ	p	μ_1	e_{1m}	e_{2m}	e_{1c}	e_{2c}	e_{10}	e_{20}
40	4.0	0.25	2.0	0.5026 (0.3113)	0.2603 (0.3461)	0.6485 (0.0898)	0.0459 (0.0338)	0.6458	0.0394
			6.0	0.3279 (0.2540)	0.1418 (0.2923)	0.3242 (0.1039)	0.0325 (0.2580)	0.3338	0.0242
	8.0		2.0	0.4032 (0.2540)	0.2034 (0.2929)	0.4939 (0.1039)	0.0415 (0.2580)	0.4615	0.0436
			6.0	0.1389 (0.0396)	0.0268 (0.0330)	0.1433 (0.0318)	0.0192 (0.0132)	0.1458	0.0142
	16.0		2.0	0.2568 (0.1301)	0.0668 (0.0746)	0.2636 (0.0557)	0.0407 (0.0209)	0.2660	0.0349
			6.0	0.0331 (0.0109)	0.0061 (0.0046)	0.0343 (.0099)	0.0051 (0.0032)	0.0338	0.0040
40	4.0		2.0	0.3144 (0.2743)	0.4328 (0.3452)	0.3675 (0.0959)	0.2403 (0.1073)	0.3881	0.2027
			6.0	0.2284 (0.1160)	0.1510 (0.1299)	0.3675 (0.0470)	0.2403 (0.0555)	0.3216	0.0872
	8.0		2.0	0.2952 (0.2189)	0.2524 (0.2450)	0.2727 (0.0642)	0.1707 (0.0641)	0.2735	0.1585
			6.0	0.1033 (0.0367)	0.0554 (0.0437)	0.0969 (0.0204)	0.0519 (0.0285)	0.1008	0.0413
	16.0		2.0	0.1617 (0.0898)	0.1485 (0.1351)	0.1506 (0.0387)	0.1171 (0.0464)	0.1586	0.0991
			6.0	0.0231 (0.0077)	0.0142 (0.0100)	0.0224 (0.0057)	0.0136 (0.0080)	0.0234	0.0102
100	4.0	0.50	2.0	0.4316 (0.2809)	0.2904 (0.3073)	0.3789 (0.0697)	0.2201 (0.0696)	0.3881	0.2027
			6.0	0.2295 (0.0598)	0.1060 (0.0632)	0.2255 (0.0299)	0.0979 (0.0318)	0.2316	0.0872

information is known in the former case. The ratios of corresponding standard deviations vary from 1.1 to 3.5

Table 2 shows the total probabilities of misclassification, with the standard deviations of e_m and e_c shown in parentheses. Also shown are the standardized biases. The first entry in each cell under $B(e_j)$ is the value of the absolute bias from e_o standardized by the standard deviation of e_j and the second is the value of the ratio of the absolute bias to e_o. B is the value of the ratio of the bias of e_m from e_c.

From Table 2 we see that the total conditional probalilities of misclassification, as estimates of the optimal probabilities, are poor when λ and d are small. The standard deviations are smaller than those for individual probabilities but they are sill quite large. When λ and d are large the estimates are quite good and e_c does consistently better than e_m.

TABLE 2: Total Probabilities of Misclassification and Percentage biases ($\mu_2 = 1.0$)

Parameters				Classification Procedure					Relative Bias to
				Mixtures	Completely Classified		Optimal		Completely Classified
n	p	λ	μ_1	e_m	e_c	e_o	$B(e_m)$	$B(e_c)$	B
40	0.25	4.0	2.0	0.3209 (0.1886)	0.1911 (0.0068)	0.1911	62.87 68.01	82.32 2.93	63.22
			6.0	0.1133 (0.0169)	0.1546 (0.0054)	0.1016	69.23 11.52	70.56 3.74	7.49
		8.0	2.0	0.2534 (0.1698)	0.1546 (0.0137)	0.1481	62.01 71.10	47.72 4.39	63.91
			6.0	0.0548 (0.0176)	0.0502 (0.0045)	0.0471	44.11 16.35	70.32 6.58	9.16
		16.0	2.0	0.1166 (0.0358)	0.0965 (0.0057)	0.0927	• 66.56 25.78	65.30 4.10	20.83
			6.0	0.0129 (0.0017)	0.0124 (0.0010)	0.0114	84.12 13.16	97.50 8.77	4.03
40	0.50	4.0	2.0	0.3736 (0.0687)	0.3039 (0.0114)	0.2945	113.86 26.47	74.99 2.81	22.94
			6.0	0.1897 (0.0336)	0.1653 (0.0082)	0.1594	89.96 19.01	72.12 3.70	14.76
	8.0		2.0	0.2738 (0.0789)	0.2217 (0.0061)	0.2160	73.21 26.76	93.51 2.64	23.50
			6.0	0.0793 (0.0101)	0.0744 (0.0059)	0.0711	81.62 11.53	56.46 4.64	6.59
		16.0	2.0	0.1551 (0.0399)	0.1339 (0.0071)	0.1289	65.66 20.33	70.00 3.64	15.83
			6.0	0.0186 (0.0026)	0.0180 (0.0020)	0.0168	69.73 10.71	58.90 7.14	3.33
100	0.50	4.0	2.0	0.3610 (0.0630)	0.2995 (.0060)	0.2954	104.18 22.21	69.55 1.39	20.53
			6.0	0.1678 (0.0121)	0.1617 0 0038)	0.1594	69.66 5.27	33.70 1.44	3.77

From the last column we see that the mixture discrimination procedure relative to the classified performs poorly for d = 1.0 and especially for small λ. However, as both d and λ increase the performance improves. Generally, the performance when p = 0.50 is superior to the performance when p = 0.25.

When the sample size is increased from n = 40 to n = 100, all the estimates for the two combinations of parameters considered improve. Even for small values of d, the estimates of e_{im} move towards their optimum values. The standard deviations are smaller as expected and all the biases are considerably reduced. The performance of the mixture discrimination procedure relative to the completely classified procedure, as measured by total probabilities, is quite good.

RELATIVE EFFICIENCY RESULTS

We compare the efficiency of $L_m(x)$ relative to $L_c(x)$ for small samples. Estimating the expected total conditional probabilities of misclassification by e_j, the sample mean of the total conditional probabilities of misclassification, we define the efficiency of $L_m(x)$ relative to $L_c(x)$ based on simulation experiments as

$$\varepsilon_S = (e_c - e_o)/(e_m - e_o) \qquad (5.1)$$

Values for selected parameters are shown in Table 3. We also compare in Table 3 the corresponding asymptotic relative efficiency values. These latter values are quoted from Amoh [1]. The asymptotic relative efficiency ε_A is defined as the ratio

$$\varepsilon_A = \{E(e_c) - e_o\}/\{E(e_m - e_o)\} \qquad (5.2)$$

where $E(e_i)$ are evaluated up to and including the terms of the first order in its asymptotic expansion. This

measure of asymptotic relative efficiency was used by O'Neill [12] and Ganesalingam and McLachlan [8].

TABLE 3: Simulation and asymptotic relative efficiencies of mixture and classified discrimination

Parameters					Simulation	Asymptotic
n	p	λ	μ_1	μ_2	ε_S	ε_A
40	0.25	4.0	2.0	1.0	0.0430	0.0373
			6.0		0.3251	0.3006
		8.0	2.0		0.0620	0.0866
			6.0		0.4030	0.5332
		16.0	2.0		0.1573	0.2157
			6.0		0.6692	0.8155
	0.50	4.0	2.0		0.1092	0.0223
			6.0		0.1965	0.2193
		8.0	2.0		0.0984	0.0654
			6.0		0.4032	0.4462
		16.0	2.0		0.1907	0.1877
			6.0		0.6385	0.7631
100	0.50	4.0	2.0		0.0631	0.0223
			6.0		0.2767	0.2193

We find from Table 3 that the relative efficiencies are more sensitive to changes in d rather than λ. The asymptotic relative efficiency gives quite a reliable guide as to what happens when the sample size is small.

REFERENCES

1. **Amoh, R.K.:** *Classification Procedures Associated with the Inverse Gaussian Distribution. Unpublished Ph.D. Dissertation, University of Manitoba. Winnipeg, Canada. 1983.*

2. **Amoh, R.K.:** *Estimation of Parameters in Mixtures of Inverse Gaussian Distributions. Commun. Statist. Theor. Meth. 13, 1031-1043. 1984.*

3. **Amoh, R.K.:** *Asymptotic Results for a Linear Distriminant Function Estimated from a Mixture of Inverse Gaussian Populations. J.Statit. Planning and Inference 14, 233-243, 1986.*

4. **Amoh, R.K. and Kocherlakota, K.:** *Errors of Misclassification Associated with the Inverse Gaussian Distribution. Commun. Statit. Theor. Meth. 15, 589-612;, 1986*

5. **Amoh, R.K. and Kocherlakota:** *Updating Discrimi nant Functions Estimated from Inverse Gaussian Popula tions;Commun. Statist. Simul. Comp. 20, 619-637, 1991.*

6. **Chhikara, R.S. and Folks, J.L.:** *The Inverse Gaussian Distribution Theory, Methodology and Applications. Marcel Dekker, Inc. New York, 1989.*

7. **Folks J.L. and Chhikara, R.S.:** *The Inverse Gaussian Distribution and its Statistical Application - A review (with discussion). J.Roy. Statist. Soc. B 40, 263-289, 1978.*

8. **Ganesalingam, S. and McLachlan, G.J.:** *The Efficiency of a Linear Discriminant Function based on Unclassified Initial Samples. Biometrika 65, 658-662, 1978.*

9. **Ganesalingam, S. and McLachlan, G.J.:** *Small Sample Results for a Linear Discriminant Function Estimated from a Mixture of Normal Populations. J. Statist. Computation, Simul. 9, 151-158, 1979.*

10. **Kalbfleisch, J.D. and Prentice, R.L.:** *The Statistical Analysis of Failure Time Data. John Wiley and Sons, New York, 1980.*

11. **Michael, J.R., Schucany, W.R. and Hass, R.W.** *Generating Random Variables using Transformations with Multiple Roots. Amer. Statists. 30, 88-89, 1976.*

12. **O'Neill, T.J.** *Normal Discrimination with Unclassified Data. J. Amer. Statist. Assoc. 73, 821-826, 1978.*

13. **Proschan, F.:** *Theoretical Explanation of Observed Decreasing Failure Rate. Technometrics, 5, 3 375-383. 1963.*

14. **Schrodinger, E.:** *Zur Theorie dor Fall-und Steigversuche on Teilchen mit Browncher Bewegung. Physikalische Zeitschrift 16 191-204, 1915.*

15. **Tweedie, M.C.K.:** *Statistical Properties of Inverse Gaussian Distribution, I Ann Meth. Statist. 28, 362-377. 1957*

16. **Wald, A.:** *Sequential Analysis. John Wiley, New York, 1947.*

A NOTE ON COMBINING UNBIASED ESTIMATORS

R.K. Amoh, PhD
Dept. of Mathematics, University of Science & Technology, Kumasi, Ghana

Graybill and Deal [1], in the paper with the same title as above, have proved the following result. If $X_i \sim N(\mu, \sigma^2/n_i)$, i = 1, 2 and s_1^2 and s_2^2 are independent $X^{2\ 2}$ with m_1 and m_2 degrees of freedom respectively then the estimator $\mu = (n_1 s_2^2 + n_2 s_1^2\ n_2^2 s_1^2 x_2)/(n_1 s_2^2 + n_2 s_1^2)$ is an unbiased estimator of whose variance is smaller than Min $\sigma^2/n_1, \sigma^2/n_2)$ provided m_1 and m_2 are both greater than 9. The theorem also further states that if either of m_1 or m_2 is ≤ 9 then the above result does not hold. The purpose of this present note is to show that the last statement is not true and thus sharpen the bounds on m_1, m_2 for which the result holds.

Following the proof of Graybill and Deal it is clear that the values of (m_1, m_2), for which μ would have smaller variance than that of either x_1 or x_2, are those which satisfy

$$\left.\begin{array}{l} Q(m_1, m_2) = m_1 m_2 - 8m_1\ 2m_2 \geq 0 \\ Q(m_1, m_2) = m_1 m_2 - 8m_1\ 2m_2 \geq 0 \end{array}\right\} \quad (1)$$

Since the two expressions are symmetric in (m_1, m_2) we will work with $A(m_1, m_2)$ only. We also assume that both m_1 and m_2 are at least five, this condition being necessary for existence of certain integrals involved in the proof.

Consider a pair like $m_1 = 9$ and $m_2 = 20$ say, then it is clear that (m_1, m_2) satisfy (1) and therefore there are additional values of (m_1, m_2) for which the result holds. In order to get the additional values of (m_1, m_2) for which (1) is true we proceed as follows. We note that

$$Q(m_1, m_2) = m_1 (m_2 - 8) - 2m_2,$$

Therefore for any m_1 and $m_2 \leq 8$ the condition (1) is not satisfied. If $m_2 = 9$, then for any $m_1 < 18$, $Q(m_1, m_2) < 0$. Similarly for any m_2 and $m_1 \geq 8$, $Q(m_2, m_1)$ 0 and for $m_1 = 9$ and $m_2 < 18$, $Q(m_2, m_1) < 0$. However, if $m_2 = 9$ $m_1 \leq 18$ if $m_1 = 9$ and $m_2 \leq 18$, both $Q(m_1, m_2)$ and $Q(m_2, m_1)$ are positive and (1) is satisfied. Thus, to the range $m_1 \geq 10$, $m_2 \geq 10$ obtained by Graybill and Deal we can add $m_1 = m_2 \geq 18$ or m_2 and $m_1 \geq 18$

"The extension of the limits will be found most useful when it is very difficult or expensive to obtain samples from one population whilst member of the other population exist in abundance."

REFERENCE

1. **Graybill, F.A. and Deal, R.B.** *Combining Unbiased Estimators. Biometrics. Vol. 15, pp 543-550. 1959.*