SOME ALTERNATIVE ESTIMATORS FOR A POPULATION MEAN

Donald T. Searls, Westat Research Analysts, Inc.

The problem considered in this paper is one that most samplers encounter early in their work with sample data. The problem can be illustrated by the question that a sampler frequently asks himself, particularly if he is working with relatively small samples. The question is, "What do I do with large or extreme observations in the sample?" The sampler first attempts to answer this question by a careful review of the data to see if an outlier has somehow appeared or if in fact the offending observation or observations are actually true observations. This paper is concerned only with the latter case where the use of outlier theory would be both unrealistic and unstatistical.

Suppose the data has been collected in order to obtain an estimate of the mean (μ) of the sampled distribution. The types of distributions generally encountered in practical sampling situations can be characterized by the following:

a. Unimodality
b. Positive skewness or symmetry
c. Non-negative values.

Unless the precise form of the distribution is known exactly, $\bar{y}$ will probably be used as an estimate of the distribution mean. An occasional sample will contain one or more observations from the right tail of the distribution due to the sampling process. When this occurs, and the sample size is small, $\bar{y}$ will probably exceed μ by a considerable amount. In this situation the client being a practical man is frequently quick to point out the fact that the one or more large observations are unduly influencing the estimate of the mean. The argument that the procedure is unbiased falls on deaf ears. The client is not interested in what happens in the long run--he wants an estimate as close to μ as possible for this particular case. In fact he is apt to regard any difference between $\bar{y}$ and μ as a bias. By this reasoning any sample estimate is biased unless it coincides with the parameter value.

This is an interesting viewpoint since it points up the fact that merely because these "biases" tend to average out in the long run does not really imply any particular merit for the estimator unless the estimates obtained are accumulated or manipulated in some fashion.

The above leads to the using of mean-squared-error as a criterion for comparing estimators rather than using unbiasedness and variance. There are not always unique minimum mean-squared-error estimators nor is it always possible to determine the MMSE estimator but this does not preclude its use as a criterion for comparing alternative estimators.

Let us examine the estimator generally employed both by statisticians and others in situations of this type. More often than not the large observations are ignored (generally there is only one) and the sample mean is derived from the remaining observations. The question immediately arises: Is this practice tenable with sound statistical theory? Surprisingly, the answer is a qualified yes. And in fact it appears that under proper conditions, if the observation was not ignored, it should have been.

Part of the qualification is that there is some predetermined point t such that if an observation is larger than this point it will be ignored. This point should be explicitly stated in editing instructions before the sample is drawn.

This paper will not provide pat answers for the handling of large observations but it will develop results to indicate that some of the procedures being employed should not necessarily be condemned and that rough guidelines can be derived for future use. Three procedures, in addition to the above, will be considered.

The first estimator formalizes as follows:

$$\bar{y}_1 = \frac{\sum_{j=1}^{r} y_j}{r} \quad , \qquad (y_j < t) \quad (r > 0)$$

$$= t \ . \qquad (r = 0)$$

The case where r = 0 would be rare but is included for completeness.

For a given $r \geq 1$, $\bar{y}_1$ can be regarded as a simple random sample of size r from the truncated portion of the distribution function and as such it provides an unbiased estimate of the mean (μ_t) of the truncated distribution. Also r is distributed as the binomial with parameters n and F(t) or p. Let q = 1 - p.

$$E(\bar{y}_1) = (1 - q^n)\, \mu_t + q^n t \ ,$$

and

$$MSE(\bar{y}_1) = (1 - q^n)\, [\sigma_t^2\, E(1/r) + (\mu - \mu_t)^2] + q^n\, (t - \mu)^2 \ .$$

When t is equal to the upper limit (b), $MSE(\bar{y}_1) = \sigma^2/n$. If $\frac{d}{dt}[MSE(\bar{y}_1)]$ is positive as t approaches b, then $MSE(\bar{y}_1)$ must be less than σ^2/n for some region of t. This turns out to be true.

$$\frac{d}{dt}[MSE(\bar{y}_1)] = \frac{f(t)}{p}\Big\langle (1-q^n)\Big[\sigma_t^2 + E[\tfrac{1}{r}][(t-\mu_t)^2 - \sigma_t^2] - 2(\mu-\mu_t)(t-\mu_t)\Big] - np(1-q^{n-1})\,\sigma_t^2\,E[\tfrac{1}{r} \mid n-1,p] - npq^{n-1}(t-\mu_t)(t+\mu_t-2\mu)\Big\rangle + 2q^n(t-\mu) \quad .$$

If $b = +\infty$ the term $(t-\mu_t)^2 E(\frac{1}{r})$ will dominate as t approaches b. Thus the first derivative is positive as t approaches b.

A more detailed proof is presented in [1], along with proofs for cases where $b \neq +\infty$.

A second estimator that is sometimes used is one where the large observations are ignored but sample size is kept constant by replacing deleted observations.

$$\bar{y}_2 = \frac{\sum_{j=1}^{r} y_j}{r} \qquad \begin{matrix} (y_j < t) \\ (r \text{ fixed}) \end{matrix}$$

For any r, $\bar{y}_2$ behaves as the sample mean of the truncated distribution.

$$E(\bar{y}_2) = \mu_t \quad ,$$

and

$$MSE(\bar{y}_2) = \sigma_t^2/r + (\mu-\mu_t)^2 \quad .$$

The proof for $\bar{y}_2$ is similar to that for $\bar{y}_1$ and can be found in [1].

A third estimator can be formed by replacing all large observations with the value of the cutoff point t.

$$\bar{y}_3 = \frac{\sum_{j=1}^{r} y_j + (n-r)t}{n} \qquad \begin{matrix} (0 \le r \le n) \\ (y_j \le t) \end{matrix}$$

where r is distributed as the binomial with parameters n and p.

$$E(\bar{y}_3) = p\,\mu_t + qt \quad ,$$

and

$$MSE(\bar{y}_3) = \frac{p}{n}[\sigma_t^2 + q(t-\mu_t)^2] + q^2(\mu_{t'}-t)^2,$$

where $\mu_{t'}$ is the mean of the right truncated portion of the distribution.

$$\frac{d}{dt}[MSE(\bar{y}_3)] = 2q[\frac{p}{n}(t-\mu_t) - q(\mu_{t'}-t)] \quad .$$

The derivative is positive as t approaches b since the second term in brackets is approaching zero while the first term is approaching a positive constant.

Figure 1 and table 1 demonstrate the gain achieved for the exponential distribution by the use of $\bar{y}_3$ for various values of t/μ. Table 2 presents specified characteristics for $\bar{y}_3$ when optimum values of t are used.

If a near optimum value of t were used for $\bar{y}_3$ over half (56%) of the samples of size five would have one or more observations exceeding t. As the sample size increases this proportion approaches 1. Correspondingly, the expected number of observations exceeding t increases from .75 for samples of size five up to approximately 4 for samples of size 500. The optimum point for t varies from about double the true mean up to five times the true mean over this range.

The preceding results raise the question of whether or not conditions can be found such that one or more of the large observations can be arbitrarily discarded.

Consider the case where only the maximum sample observation (y_m) is discarded.

$$\bar{y}_4 = \frac{\sum_{j=1}^{n} y_j - y_m}{n-1}$$

$$E(\bar{y}_4) = \frac{1}{n-1}(n\mu - \mu_m) \qquad \text{where}$$

$$E(y_m) = \mu_m \quad .$$

$$MSE(\bar{y}_4) = \frac{1}{(n-1)^2}[n\sigma^2 + \sigma_m^2 + (\mu_m-\mu)^2 - 2n\,Cov(\bar{y}, y_m)] \quad .$$

$$MSE(\bar{y}_4) \le \sigma^2/n \qquad \text{if}$$

$$Cov(\bar{y}, y_m) \ge \frac{1}{2n}[(\frac{2n-1}{n})\sigma^2 + (\mu_m-\mu)^2 + \sigma_m^2].$$

For samples of size n = 2 equality holds for the exponential distribution and the strict inequality holds for the pareto. Table 3 presents results for the pareto distribution.

In conclusion it can be stated that the deletion of large observations from a sample may result in the use of an estimator with a smaller

mean-squared-error than the sample mean. Perhaps an even more important implication from the above results however is the possibility that even more dramatic gains can be achieved in the estimation of σ^2.

REFERENCES

[1] Searls, D. T., "On the Large Observation Problem," Institute of Statistics Mimeo Series No. 332, North Carolina State College, Raleigh, N. C.

ACKNOWLEDGEMENTS

This paper presents a portion of the research conducted for a Ph.D. thesis written under the direction of Dr. A. L. Finkner. Acknowledgement for helpful suggestions are also due Walter Hendricks, Dr. R. L. Anderson, Dr. W. L. Smith, and Dr. C. Proctor.

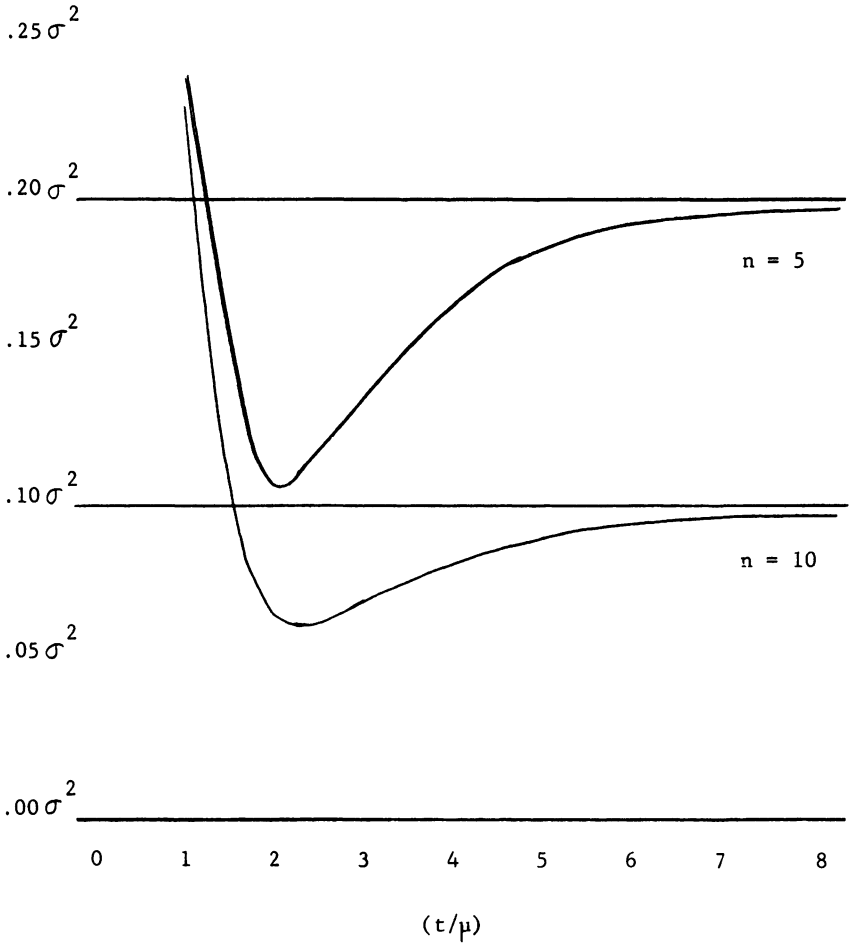

Figure 1. Comparison of $MSE(\bar{y}_3)$ and σ^2/n for the exponential distribution

Table 1. Relative efficiencies (%) of $\bar{y}_3$ for samples from the exponential distribution.

Values of t/μ	Sample Size 5	10	50	100	500
1	83.4	44.1	9.2	4.6	.9
2	183.1	153.3	66.7	39.1	9.1
3	140.5	138.0	120.6	104.2	49.9
4	117.0	116.7	114.9	112.7	97.7
5	107.2	107.2	106.9	106.7	104.7
6	103.1	103.1	103.0	103.0	102.7
7	101.3	101.3	101.3	101.3	101.3
8	100.5	100.5	100.5	100.5	100.5
9	100.2	100.2	100.2	100.2	100.2
10	100.1	100.1	100.1	100.1	100.1

Table 2. Exponential distribution; characteristics for $\bar{y}_3$ when optimum values for t are used.

Sample size	Values of t/μ	Exp. no. $>t_o$	Exp. % of samples with one or more $>t_o$	Relative eff. (%)
5	1.9	.75	55.5	184.9
10	2.2	1.11	69.1	156.7
50	3.2	2.04	87.5	121.3
100	3.6	2.73	93.7	113.7
500	4.9	3.72	96.3	104.7

$$f(y) = \frac{1}{\mu} e^{-y/\mu} \qquad (0 < y < +\infty)$$

Table 3. Pareto distribution; $\bar{y}_4$ with $n = 2$.

Value of α	Value of $r^2(\bar{y}, y_m)$	Relative Efficiency (%)
3	.964	250.0
4	.947	197.5
5	.938	150.0
6	.931	137.5
10	.918	118.8

$$f(y) = \alpha\, y^{-(\alpha+1)} \qquad (\alpha > 2)$$
$$(1 < y < +\infty)$$