Tatra Mt. Math. Publ. **26** (2003), 391–397

Tatra Mountains
Mathematical Publications

ON CERTAIN CONSEQUENCES OF AVERAGING OBSERVATIONS DERIVED FROM NORMAL POPULATION

Joanna Tarasińska

ABSTRACT. It is assumed that we have a large number of observations from normal population which can be used in classical statistical inference about the mean of the population (confidence interval, testing hypothesis). Sometimes the investigator initially takes means of a fixed number of observations and makes statistical inference based on this "sample of means". It is a good solution when he is not sure that the population is normal. But if *it is normal*, this averaging procedure is superfluous and causes loss of information.

The aim of this paper is to check how much the quality of inference based on "the sample of means" is worse in comparison to the quality of inference based on "the raw sample". It turns out that the length of confidence intervals and the power of tests are rather insensitive to averaging.

1. Introduction

The central limit theorem speaks about advantages of averaging observations when we want to make statistical inference about the mean of the distribution. The classical confidence intervals and significance tests for the mean based on a t-distribution are valid under the assumption that the sample is taken from the normal population. If it is not, we can make inference based not on the raw sample but on the sample of means (i.e., the sample given by averaging the fixed number of elements of the raw sample).

On the other hand, if the distribution in a population *is normal*, this averaging procedure is superfluous and causes a loss of information contained in the raw sample. In consequence the statistical inference about the mean is less efficient: the precision of the estimator is lower, the confidence intervals are longer and the power of the tests decreases. The question is: how much worse is the statistical inference. In other words, how much the quality of the inference based on normal assumptions is sensitive to averaging.

2000 Mathematics Subject Classification: 62F03, 62F25.
Keywords: sample, normal distribution, confidence interval, power of test.

Let us assume that we have $n = rk$ independent observations of the normal random variable X:

$$X_{ij} \sim \mathcal{N}(\mu, \sigma^2) \qquad i = 1, \ldots, r\,; \quad j = 1, \ldots, k\,. \tag{1}$$

Let us assume that before starting the classical inference about the mean we have averaged the observations X_{ij} taking the arithmetic mean of each k elements. So we have the model

$$\bar{X}_{i\cdot} \sim \mathcal{N}\left(\mu, \frac{1}{k}\sigma^2\right) \qquad i = 1, \ldots, r\,; \tag{2}$$

where $\bar{X}_{i\cdot} = \frac{1}{k}\sum\limits_{j=1}^{k} X_{ij}$.

Let us compare the quality of the standard statistical inference about μ in models (1) and (2). The confidence intervals and the hypothesis testing in model (1) are based on the pivot

$$t = \frac{\bar{X} - \mu}{\sqrt{\frac{\widehat{\sigma^2}}{rk}}} \sim t_{rk-1}\,,$$

where $\bar{X} = \frac{1}{rk}\sum\limits_{i=1}^{r}\sum\limits_{j=1}^{k} X_{ij}$, $\widehat{\sigma^2} = \frac{1}{rk-1}\sum\limits_{i=1}^{r}\sum\limits_{j=1}^{k}(X_{ij} - \bar{X})^2$ and t_{rk-1} denotes a t-distribution with $rk - 1$ degrees of freedom.

In model (2) the inference is based on the t_{r-1} distributed pivot

$$t = \frac{\bar{X} - \mu}{\sqrt{\frac{\widetilde{\sigma^2}}{rk}}} \sim t_{r-1}\,,$$

where $\widetilde{\sigma^2} = \frac{k}{r-1}\sum\limits_{i=1}^{r}(\bar{X}_{i\cdot} - \bar{X})^2$.

2. The comparison of confidence intervals

Let us compare the length of $(1-\alpha)100\ \%$ confidence intervals for μ in models (1) and (2). The Table 1 gives the probability distributions of the lengths of confidence intervals, their expected values and variances. $t_{p,\nu}$ is 100p% percentage point of a t-distribution with ν degrees of freedom, χ_ν is a random variable of a χ distribution with ν degrees of freedom, $w(\nu) = \frac{\Gamma\left(\frac{\nu}{2}\right)}{\sqrt{\frac{\nu(\nu-1)}{2}}\Gamma\left(\frac{\nu-1}{2}\right)}$, $h(\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)\Gamma\left(\frac{\nu-1}{2}\right) - \Gamma^2\left(\frac{\nu}{2}\right)}{\Gamma^2\left(\frac{\nu-1}{2}\right)}$, $\Gamma(\cdot)$ is the complete Gamma function.

Table 1

	in model (1)	in model (2)
the length	$L = 2t_{1-\frac{\alpha}{2},rk-1}\sqrt{\frac{\widehat{\sigma^2}}{rk}}$	$\bar{L} = 2t_{1-\frac{\alpha}{2},r-1}\sqrt{\frac{\widehat{\sigma^2}}{rk}}$
the distribution of length	$L \sim \frac{2}{\sqrt{rk(rk-1)}} t_{1-\frac{\alpha}{2},rk-1} \cdot \sigma \cdot \chi_{rk-1}$	$\bar{L} \sim \frac{2}{\sqrt{rk(r-1)}} t_{1-\frac{\alpha}{2},r-1} \cdot \sigma \cdot \chi_{r-1}$
the expected value of length	$E(L) = 2t_{1-\frac{\alpha}{2},rk-1} \cdot \sigma \cdot w(rk)$	$E(\bar{L}) = 2t_{1-\frac{\alpha}{2},r-1} \cdot \frac{\sigma}{\sqrt{k}} \cdot w(r)$
the variance of length	$V(L) = \frac{8}{rk(rk-1)} h(rk)\sigma^2 t^2_{1-\frac{\alpha}{2},rk-1}$	$V(\bar{L}) = \frac{8}{rk(r-1)} h(r)\sigma^2 t^2_{1-\frac{\alpha}{2},r-1}$

The Figure 1 presents the probability density functions of L (in model (1)) and $\bar{L}$ (in model (2)) in the case of $n = 100$, $k = 10$, $\sigma = 1$.

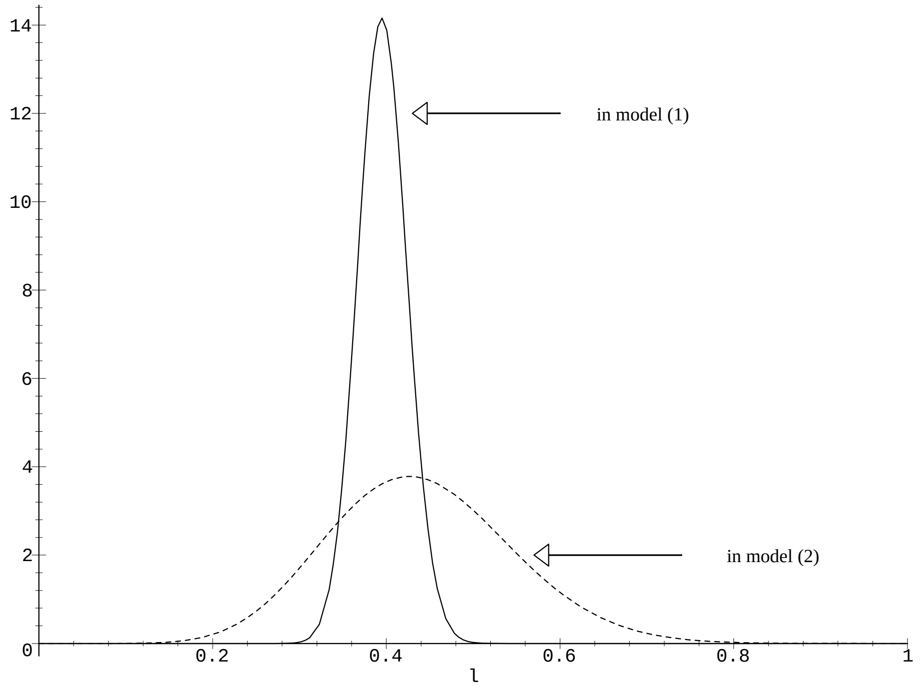

Figure 1

The Table 2 contains the comparison of the expected values and the variances of L and $\bar{L}$ in the case of $n = 100$, and different numbers of averaged observations in model (2), i.e., $k = 2, 5, 10, 20$. The upper values in the cells are for 90 % confidence intervals, the mean for 95 % and the lower one for 99 % c.i.

TABLE 2

k	2	5	10	20
$\frac{E(\bar{L})}{E(L)}$	1.00	1.03	1.08	1.21
	1.01	1.04	1.11	1.32
	1.02	1.08	1.21	1.65
$\frac{V(\bar{L})}{V(L)}$	2.05	5.59	13.05	38.10
	2.07	5.73	13.92	45.25
	2.10	6.11	16.40	71.02

The following conclusions can be drawn from the Table 2:

- the expected value of the length of confidence interval hardly increases after averaging;
- the variance of the length of confidence interval increases very much after averaging, so the length of confidence intervals becomes much less stable;
- the greater the confidence level is, the greater the differences between the lengths of confidence intervals are.

The lengths L and $\bar{L}$ can be also compared from another point of view, namely, the probability $\Pr(L < p\bar{L})$, where $p \in [0,1]$. With $p = 1$ it means the probability of lengthening the confidence interval if it is based on the previously averaged observations.

Let us notice that the following sequence of equivalencies is true

$$L < \bar{L}p \iff L^2 < \bar{L}^2 p^2 \iff \frac{\sum_{i=1}^{r}\sum_{j=1}^{k}(X_{ij}-\bar{X})^2}{k\sum_{i=1}^{r}(\bar{X}_{i\cdot}-\bar{X})^2} < p^2 \cdot \frac{t^2_{1-\frac{\alpha}{2},r-1}}{t^2_{1-\frac{\alpha}{2},rk-1}} \cdot \frac{rk-1}{r-1}$$

$$\iff \frac{\sum_{i=1}^{r}\sum_{j=1}^{k}(X_{ij}-\bar{X}_{i\cdot})^2}{k\sum_{i=1}^{r}(\bar{X}_{i\cdot}-\bar{X})^2} < p^2 \cdot \frac{t^2_{1-\frac{\alpha}{2},r-1}}{t^2_{1-\frac{\alpha}{2},rk-1}} \cdot \frac{rk-1}{r-1} - 1$$

$$\iff \frac{\chi^2_{r(k-1)}}{\chi^2_{r-1}} < p^2 \cdot \frac{t^2_{1-\frac{\alpha}{2},r-1}}{t^2_{1-\frac{\alpha}{2},rk-1}} \cdot \frac{rk-1}{r(k-1)} - \frac{r-1}{r(k-1)},$$

where $\chi^2_{r(k-1)}$ and χ^2_{r-1} are independent χ^2 variables with $r(k-1)$ and $r-1$ degrees of freedom, respectively. Thus

$$\Pr(L < p\bar{L}) = G\left(\frac{t^2_{1-\frac{\alpha}{2},r-1}}{t^2_{1-\frac{\alpha}{2},rk-1}} \cdot \frac{rk-1}{r(k-1)} \cdot p^2 - \frac{r-1}{r(k-1)}\right),$$

where G is the cumulative distribution function of F distribution with $r(k-1)$ and $r-1$ degrees of freedom.

The Table 3 contains the probabilities $\Pr(L < p\bar{L})$ for $n = 100$, $k = 2, 5, 10$ and $p = 0.5(0.1)1.0$. All the values are computed by the means of Maple V Release program.

TABLE 3. $\Pr(L < p\bar{L})$, $n = 100$.

k	p	90 % c.i.	95 % c.i.	99 % c.i.
$k=2$	0.5	0	0	0
	0.6	0	0	0
	0.7	0	0	0
	0.8	0	0	0
	0.9	0.070	0.077	0.097
	1.0	0.553	0.569	0.608
$k=5$	0.5	0	0	0
	0.6	0	0	0
	0.7	0.004	0.006	0.013
	0.8	0.074	0.090	0.014
	0.9	0.298	0.331	0.416
	1.0	0.578	0.610	0.686
$k=10$	0.5	0	0	0.003
	0.6	0.010	0.017	0.051
	0.7	0.080	0.109	0.210
	0.8	0.238	0.289	0.428
	0.9	0.435	0.491	0.625
	1.0	0.612	0.660	0.768

The following conclusions can be drawn from the Table 3:

- averaging of a small number of observations hardly increases the probability of lengthening the confidence interval (see $k = 2$, $p = 1$ in Table 3);
- the greater the confidence level of the interval is, the greater the probability of its lengthening after averaging of a few observations becomes.

3. The comparison of the power of test for μ

Let us consider the hypotheses

$$H_0 : \mu = \mu_0 \quad \text{versus} \quad H_1 : \mu \neq \mu_0 .$$

The test function in the model (1) is $t = \frac{\bar{X}-\mu_0}{\sqrt{\frac{\widehat{\sigma^2}}{n}}}$ and H_0 is rejected when $|t| > t_{1-\frac{\alpha}{2},n-1}$ (α is the significance level). If H_1 is true, the test function has got a

$t'_{n-1,\delta}$ distribution, i.e., noncentral t distribution with $n-1$ degrees of freedom and noncentrality parameter $\delta = \frac{(\mu-\mu_0)\sqrt{n}}{\sigma}$ [1]. So the power of the test is

$$\beta_n(\mu) = 1 + G_{n-1,\delta}\left(t_{\frac{\alpha}{2},n-1}\right) - G_{n-1,\delta}\left(t_{1-\frac{\alpha}{2},n-1}\right),$$

where $G_{n-1,\delta}$ is the cumulative distribution function of $t'_{n-1,\delta}$.
The test function in the model (2) is $t = \frac{\bar{X}-\mu_0}{\sqrt{\frac{\widetilde{\sigma^2}}{n}}}$, and the power of the test is

$$\bar{\beta}_n(\mu) = 1 + G_{r-1,\delta}\left(t_{\frac{\alpha}{2},r-1}\right) - G_{r-1,\delta}\left(t_{1-\frac{\alpha}{2},r-1}\right),$$

where $G_{r-1,\delta}$ is the cumulative distribution function of $t'_{r-1,\delta}$ (the same noncentrality parameter!). $G_{\nu,\delta}(t)$ can be calculated from the following formula [2]:

$$G_{\nu,\delta}(t) = \frac{\sqrt{2\pi}}{2^{\frac{\nu-2}{2}}\Gamma(\frac{\nu}{2})} \int_0^\infty u^{\nu-1}\phi(u)\,\Phi\left(\frac{tu}{\sqrt{\nu}} - \delta\right)du,$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the p.d.f. and the c.d.f. of the standard normal distribution, respectively. The Figure 2 presents the quotient $\frac{\bar{\beta}_n(\mu)}{\beta_n(\mu)}$ for $n = 100$, $k = 10$, $\alpha = 0.05$. It can be seen that even with such a large number of averaged observations ($k = 10$) there is no much loss of the test power (at its worst 15 %).

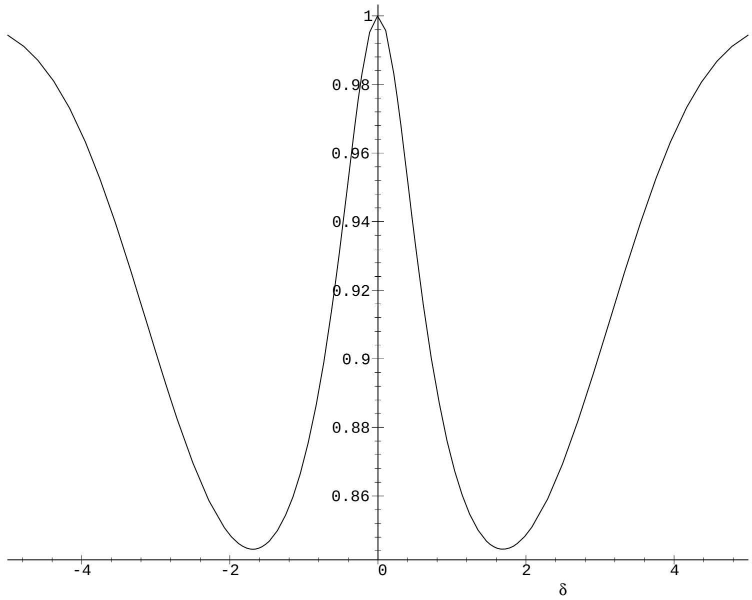

FIGURE 2

REFERENCES

[1] PATEL, J. K.—KAPADIA, C. H.—OWEN, D. B.: *Handbook of Statistical Distributions*, Marcel Dekker, Inc., New York, 1976.

[2] OWEN, D. B.: *A survey of properties and applications of the noncentral t-distribution*, Technometrics **10** (1968), 445–478.

Received February 27, 2002

Department of Applied Mathematics
University of Agriculture
Akademicka 13
PL–20-934 Lublin
POLAND

E-mail: johata@ursus.ar.lublin.pl