Note on Sample Mean in the Dependent Case

Note on Sample Mean in the Dependent Case

by

Jōgi HENNA*

We often estimate the population mean by the sample mean. However, in the case of dependent samples, it is not necessarily adequate to increase the sample size. In this note, we study the sample mean in the dependent case.

1. Introduction

Let $\{X_i\}$ be a set of uni-variate random variables, and $E(X_i)=\theta$, $\mathrm{Var}(X_i) \leqq M$ for some M for each element X_i of $\{X_i\}$. Let the joint distribution of any finite elements of $\{X_i\}$ be the normal distribution. It is well known that the distribution which has the maximun entropy for given mean and variance and the range $(-\infty, \infty)$ for the variable is the normal distribution, and that the entropy is $\log \sqrt{2\pi e}\,\sigma$, where σ^2 is the variance of the distribution [3]. Accordingly, we may consider the variance when we study the entropy of the normal distribution.

Let $X_1, X_2, \ldots\ldots, X_k$ be elements of $\{X_i\}$ and $\bar{X}_k = (1/k)\sum_{i=1}^{k} X_i$. In Section 2, we study the interrelation between the conditional entropy H of X_{k+1}, for given $X_1, X_2, \ldots\ldots, X_k$, and the precision (variance) of $\bar{X}_{k+1}$. In Section 3, we compare the independent sample size n with the dimension k when each sample has k meaurements.

2. On maximum entropy

If $\bar{X}_{k+1}$, as an estimator of θ, is better than $\bar{X}_k$, then we have $Var(\bar{X}_{k+1}) < Var(\bar{X}_k)$. We rewrite the inequality as follows:

$$(1)\quad \sum_{j=1}^{k} \sigma_{k+1\,j} + \frac{1}{2}\sigma_{k+1}^2 < \frac{2k+1}{2k^2}\sum_{j=1}^{k}\sigma_{ij}$$

where σ_{ij} is the covariance between X_i and X_j, and $\sigma_{ii}=\sigma_i^2$ is the variance of X_i. In particular, if $\sigma_i^2=\sigma^2$ ($i=1, 2, \ldots\ldots, k+1$), we have

$$(2)\quad \sum_{j=1}^{k}\rho_{k+1\,j} < \frac{2k+1}{2k^2}\sum_{i,j=1}^{k}\rho_{ij} - \frac{1}{2}$$

where ρ_{ij} is the correlation coefficient between X_i and X_j. Now, let us determine which, $X_{k+1}^{(1)}$ or $X_{k+1}^{(2)}$, we should use as the $k+1$-th measurement. For convenience sake, put

$$(3)\quad \beta_1 = \sum_{j=1}^{k}\sigma_{k+1\,j}^{(1)} + \frac{1}{2}\sigma_{k+1}^{(1)2}$$

$$(4)\quad \beta_2 = \sum_{j=1}^{k}\sigma_{k+1}^{(2)} + \frac{1}{2}\sigma_{k+1}^{(2)2}$$

where $\sigma_{k+1\,j}^{(i)}$ ($i=1, 2$) is the covariance between $X_{k+1}^{(i)}$ and X_j, and $\sigma_{k+1}^{(i)2}$ ($i=1,2$)

Recieved: April 30, 1975

* Department of Mathematics

is the variance of $X_{k+1}^{(i)}$. It can be easily seen that if $\beta_1=\beta_2$, $X_{k+1}^{(1)}$ and $X_{k+1}^{(2)}$ give the same precision for $\bar{X}_{k+1}$, and that if $\beta_1<\beta_2$, $X_{k+1}^{(1)}$ gives better precision for $\bar{X}_{k+1}$ than $X_{k+1}^{(2)}$. So, if $\beta_1<\beta_2$, we may use $X_{k+1}^{(1)}$ as the $k+1$-th measurement.

Let X_1, X_2, ..., X_k ($X_i \varepsilon \{X_i\}$) be given and let

$$S=\left\{ X : Var\,(X)=\sigma^2,\ \sum_{j=1}^{k}\sigma_{k+1\,j}=\beta,\ X \varepsilon \{X_i\} \right\},$$

where β is constant. Let X_{k+1} be an element of S, then any element of S gives the same precision for $\bar{X}_{k+1}$. Let

$$(5)\quad \Sigma_{21}=(\sigma_{k+1\,1},\ \sigma_{k+1\,2},\ \ldots\ldots,\ \sigma_{k+1\,k})$$

$$=(\sigma\sigma_1\ \rho_{k+1\,1},\ \sigma\sigma_2\ \rho_{k+1\,2},\ \ldots\ldots,\ \sigma\sigma_k\ \rho_{k+1\,k}),$$

$$\Sigma_{11}=\begin{pmatrix}\sigma_1^2, & \sigma_{12}, & \ldots\ldots, & \sigma_{1k}\\ \sigma_{21}, & \sigma_2^2, & \ldots\ldots, & \sigma_{2k}\\ & \ldots\ldots\ldots & & \\ \sigma_{k1}, & \sigma_{k2}, & \ldots\ldots, & \sigma_k^2\end{pmatrix}$$

Then the conditional variance of X_{k+1} for given X_1, X_2, ..., X_k is

$$\sigma^{*2}\equiv Var\,(X_{k+1}\,|\,X_1,\ X_2,\ \ldots\ldots,\ X_k)=\sigma^2-\Sigma_{21}\,\Sigma_{11}^{-1}\,\Sigma_{21}',$$

where Σ_{21}' is the transpose of Σ_{21}.

Lemma 1.

$$(7)\quad \max_{X\varepsilon S}\ \sigma^{*2}=\sigma^2\,(1-\beta^2\ \Sigma_{21}^{*}\ \Sigma_{11}^{-1}\ \Sigma_{21}^{*\prime})$$

where

$$\Sigma_{21}^{*}=\left(\frac{\sum_{j=1}^{k}\sigma_{1j}}{\sigma\sum_{i,j=1}^{k}\sigma_{ij}},\ \frac{\sum_{j=1}^{k}\sigma_{2j}}{\sigma\sum_{i,j=1}^{k}\sigma_{ij}},\ \ldots\ldots,\ \frac{\sum_{j=1}^{k}\sigma_{kj}}{\sigma\sum_{i,j=1}^{k}\sigma_{ij}}\right)$$

Proof.

$$(9)\quad \sigma^{*2}=\sigma^2-\sigma^2\,(\sigma_1\,\rho_{k+1\,1},\ \ldots\ldots,\ \sigma_k\,\rho_{k+1\,k})\begin{pmatrix}\sigma^{11}, & \ldots\ldots, & \sigma^{1k}\\ & \vdots & \\ \sigma^{k1}, & \ldots\ldots, & \sigma^{kk}\end{pmatrix}\begin{pmatrix}\sigma_1\,\rho_{k+1,\,1}\\ \vdots\\ \sigma_k\,\rho_{k+1,\,k}\end{pmatrix}$$

where σ^{ij} is the $(i,\ j)$ element of Σ_{11}^{-1}. Let λ be the Lagrange-multiplier and

$$(10)\quad f=\sigma^{*2}+\lambda\,(\sigma_1\,\rho_{k+1\,1}+\ \ldots\ldots\ +\sigma_k\,\rho_{k+1\,k})$$

then we have

$$(11)\quad \frac{\partial f}{\partial\rho_{k+1\,j}}=-2\,\sigma^2\,\sigma_j\,(\sigma^{j1},\ \ldots\ldots,\ \sigma^{jk})\begin{pmatrix}\sigma_1\,\rho_{k+1\,1}\\ \vdots\\ \sigma_k\,\rho_{k+1\,k}\end{pmatrix}+\lambda\sigma_j \qquad (j=1,\ 2,\ \ldots\ldots,\ k)$$

If $\partial f / \partial \rho_{k+1\ j} = 0$ $(j = 1, 2, \ldots\ldots, k)$, we have

$$(12)\quad \begin{pmatrix} \sigma^{11}, & \sigma^{12}, & \ldots\ldots, & \sigma^{1k} \\ \sigma^{21}, & \sigma^{22}, & \ldots\ldots, & \sigma^{2k} \\ & \vdots & & \\ \sigma^{k1}, & \sigma^{k2}, & \ldots\ldots, & \sigma^{kk} \end{pmatrix} \begin{pmatrix} \sigma_1 & \rho_{k+1\ 1} \\ \sigma^2 & \rho_{k+1\ 2} \\ & \vdots \\ \sigma_k & \rho_{k+1\ k} \end{pmatrix} = \frac{\lambda}{2\,\sigma^2} \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$$

Multiply (12) by Σ_{11} from the left hand side, then

$$(13)\quad \begin{pmatrix} \sigma_1 & \rho_{k+1\ 1} \\ \sigma_2 & \rho_{k+1\ 2} \\ & \vdots \\ \sigma_k & \rho_{k+1\ k} \end{pmatrix} = \frac{\lambda}{2\,\sigma^2} \begin{pmatrix} \sum_{j=1}^{k} \sigma_{1j} \\ \vdots \\ \sum_{j=1}^{k} \sigma_{kj} \end{pmatrix}$$

Accordiglgly,

$$(14)\quad \sigma_1\ \rho_{k+1\ 1} : \sigma_2\ \rho_{k+1\ 2} : \cdots\cdots : \sigma_k\ \rho_{k+1\ k}$$
$$= \sum_{j=1}^{k} \sigma_{1j} : \sum_{j=1}^{k} \rho_{2j} : \ \cdots\cdots\ : \sum_{j=1}^{k} \sigma_{kj}$$

Since $\sum_{j=1}^{k} \sigma_{k+1\ j} = \beta$, we have

$$(15)\quad \sigma_i\ \rho_{k+1\ i} = \frac{\beta \sum_{j=1}^{k} \sigma_{ij}}{\sigma \sum_{i,j=1}^{k} \sigma_{ij}} \qquad (i = 1, 2, \ldots\ldots, k)$$

This completes the proof.

As an immediate consequence of Lemma 1, we have

Corollary 1. *For given* $X_1, X_2, \ldots\ldots, X_k$, *the maximum of the conditional entropy* H *of* X_{k+1} *in* S *increases as* $|\beta|$ *decreases.*

Next we study the interrelation between the precision of $\overline{X}_{k+1}$ and the conditional entropy H of X_{k+1}. Let

$$S_1 = \left\{ X_{k+1}^{(1)} : Var\,(X_{k+1}^{(1)}) = \sigma^2,\ \sum_{j=1}^{k} \sigma_{k+1\ j}^{(1)} = \beta_1,\ X_{k+1}^{(1)} \varepsilon \{X_i\} \right\},$$

$$S_2 = \left\{ X_{k+1}^{(2)} : Var\,(X_{k+1}^{(2)}) = \sigma^2,\ \sum_{j=1}^{k} \sigma_{k+1\ j}^{(2)} = \beta_2,\ X_{k+1}^{(2)} \varepsilon \{X_i\} \right\}$$

for given $X_1, X_2, \ldots\ldots, X_k$, where $\sigma_{k+1\ j}^{(i)}$ $(i=1, 2)$ is the covariance between $X_{k+1}^{(i)}$ and X_j. In this case, any element of S_1 gives the same precision for $\overline{X}_{k+1}$, and any element of S_2 gives the same precision for $\overline{X}_{k+1}$. If $\beta_1 < \beta_2$, an element of S_1 gives better precision for $\overline{X}_{k+1}$ than that of an element of S_2. If $(X_1, X_2, \ldots\ldots, X_k, X_{k+1}^{(1)})$ and $(X_1, X_2, \ldots\ldots, X_k, X_{k+1}^{(2)})$, respectively, are distributed according to the $k+1$-variate normal distributions then the next lemma follows from Lemma 1.

Lemma 2. *If* $\beta_1, \beta_2 > 0$, *the necessary and sufficient condition for* $\max_{S_1} H > \max_{S_2} H$ *is* $\beta_1 < \beta_2$.

If $\beta_1, \beta_2 < 0$, *the necessary and sufficient condition for* $\max_{S_1} H > \max_{S_2} H$ *is* $\beta_1 > \beta_2$.

3. Comparison of independent sample size n with dimension k

Let each element of a population Ω has k different measurements $X_1, X_2, \ldots, X_k$, Suppose each measurement has an equal mean and an equal variance, i. e. , $E(X_i) = \theta$, Var $(X_i) = \sigma^2$ ($i = 1, 2, \ldots\ldots, k$). Let the correlation coefficient between X_i and X_j be ρ (>0) for arbitrary i and j ($i \neq j$). Let $(X_1^{(1)}, X_2^{(1)}, \ldots\ldots, X_k^{(1)})$, $(X_1^{(2)}, X_2^{(2)}, \ldots\ldots, X_k^{(2)})$, ..., $(X_1^{(n)}, X_2^{(n)}, \ldots\ldots, X_k^{(n)})$ be the independent random samples from Ω. Let us define ${}_k\overline{X}_n = (1/nk) \sum_{i=1}^{k} \sum_{\alpha=1}^{n} X_i^{(\alpha)}$ as an estimator of θ.

Now, we comare the independent sample size n with the dimension k. We consider two cases.

Case I : We use only the i-th measurements on n independent samples.

Case II : We use k different measurements on one sample.

In Case I, the precision of the estimator is $Var\ ({}_1\overline{X}_n) = (1/n)\ \sigma^2$.

In Case II, the precision of the estimator is

(16) $$Var\ ({}_k\overline{X}_1) = \left(\frac{1-\rho}{k} + \rho\right) \sigma^2\ (>\rho\,\sigma^2).$$

Accordigly, if $\rho < (1/n)$ (i.e., $\rho\sigma^2 < \sigma^2/n$), the precision of the estimator in Case II, for large k, can attain to that in Case I. The number of measurements on each sample is

(17) $$k^* = \left[\frac{(1-\rho)n}{1-n\rho}\right] + 1$$

where [·] is the Gaussian symbol, namely, $[x]$ is the maximum integer which is not larger than x. However, if $\rho \geqq (1/n)$, the precision of the estimator in Case II can not attain to that in Case I how large k may be. In this case, the precision in Case II can attain to that for at most $[1/\rho]$ independent samples how large k may be. On the othe other hand, suppose we estimate θ by k different measurements on one sample, and suppose the precision of the estimator (Case II) corresponds to that for n^* independent samples, then we have

(18) $$n^* = \left[\frac{k}{1-\rho+k\rho}\right] \qquad (< [1/\rho]).$$

Now, we consider next two cases.

Case I′ : We estimate θ by k_1 different measurements on each of n_1 independent samples.

Cace II′ : We estimate θ by k different measurements on each of n independent samples.

In Case I′, the precision of the estimator is

(19) $$Var\ ({}_{k_1}\overline{X}_{n_1}) = \frac{1}{n_1}\left(\frac{1-\rho}{k_1} + \rho\right) \sigma^2.$$

In Case II′, the precision of the estimator is

(20) $Var\,({}_kX_n) = \frac{1}{n}\left(\frac{1-\rho}{k}+\rho\right)\sigma^2.$

Now, let n ($<n_1$) be given in Case II′. How many measurements should we take from each sample to attain the same precision as in Case I′? If $\rho < n/(n_1 k_1 + n - n k_1)$, the precision of the estimator in Case II′ can attain to that in Case I′ for large k. In this case, the number of measurements on each sample is

(21) $K^* = \left[\frac{n_1 k_1 (1-\rho)}{n(1-\rho+k_1\rho) - n_1 k_1 \rho}\right] + 1.$

If we put $n = k_1 = 1$, then we have the previous result (17). However, if $\rho \geqq n/(n_1 k_1 + n - n k_1)$, how large k may be, the precision of the estimator in Case II′ can not attain to that in Case I′.

If the number of the measurements on each sample is given in Case II′, how many independent samples, at least, are necessary for the precision of the estimator to attain to that in Case I′ ? In this case, the number of independent samples is

(22) $N^* = \left[\frac{(1-\rho+k\rho)k_1}{(1-\rho+k_1\rho)k}\, n_1\right] + 1$

Let k and n be given in Case II′, and let k_1 be given in Case I′, and suppose the precision of the estimator in Case II′ corresponds to that for N_1^* independent samples. Then we have

(23) $N_1^* = \left[\frac{(1-\rho+k_1\rho)k}{(1-\rho+k\rho)k_1}\, n\right] \qquad \left(< \left[\frac{1-\rho+k_1\rho}{k_1\rho}\, n\right]\right)$

Accordingly, how large k may be, the precision of the estimator in Case II′ can attain to the precision of the estimator (Case I′) for at most $[(1-\rho+k_1\rho)n/k_1\rho]$ independent samples. Let $n = k_1 = 1$, then we have the result (18). Let k and n be given in Case II′, and let n_1 ($>n$) be given in Case I′, and suppose the precision of the estimator in Case II′ corresponds to that for k_1^* different measurements. Then we have

(24) $k_1^* = \left[\frac{n k (1-\rho)}{n_1(1-\rho+k\rho) - n k \rho}\right] + 1 \qquad \left(< \left[\frac{n(1-\rho)}{(n_1-n)\rho}\right] + 1\right).$

Accordingly, how large k may be, the precision of the estimator in Case II′ can attain to the precision of the estimator for at most $[n(1-\rho)/(n_1-n)\rho]$ different measurements on each sample in Case I′.

The tables of N_1^* for $k_1 = 1$, $n = 1, 2, 3$.

$\rho = 0.1$

n \ k	1	2	3	4	5	6	7	8	9	10	.....	∞
1	1	1	2	3	3	4	4	4	5	5	......	10
2	2	3	5	6	7	8	8	9	10	10	.....	20
3	3	5	7	9	10	12	12	14	15	15	.....	30

$\rho = 0.2$

n \ k	1	2	3	4	5	6	7	8	9	10	.....	∞
1	1	1	2	2	2	3	3	3	3	3	.....	5
2	2	3	4	5	5	6	6	6	6	7	.....	10
3	3	4	6	7	8	9	9	9	10	10	.....	15

$\rho = 0.3$

n \ k	1	2	3	4	5	6	7	8	9	10	.....	∞
1	1	1	1	2	2	2	2	2	2	2	.....	3
2	2	3	3	4	4	4	5	5	5	5	.....	6
3	3	4	5	6	6	7	7	7	8	8	.....	10

$\rho = 0.4$

n \ k	1	2	3	4	5	6	7	8	9	10	.....	∞
1	1	1	1	1	1	2	2	2	2	2	.....	2
2	2	2	3	3	3	4	4	4	4	4	.....	5
3	3	4	4	5	5	6	6	6	6	6	.....	7

$\rho = 0.5$

n \ k	1	2	3	4	5	6	7	8	9	10	.....	∞
1	1	1	1	1	1	1	1	1	1	1	.....	2
2	2	2	3	3	3	3	3	3	3	3	.....	4
3	3	3	4	4	4	5	5	5	5	5	.....	6

References

1. Anderson T. W., *An Introduction to Multivariate Statistical Analisis,* New York : John Wiley and Sons, Inc., 1958.
2. Kullback S., *Information Theory and Statatistics,* New York : John Wiley and Sons., 1959.
3. Rao C. R., *Linear Statistical Inference and Its Applications,* New York : John Wiley and Sons, Inc., 1965.