Listy Biometryczne – Biometrical Letters
Vol. 41 (2004), No. 2, 71-76

A remark on the comparison of mean estimates in bivariate normal sample with monotone pattern of missing data

Joanna Tarasińska

Institute of Applied Mathematics, University of Agriculture,
Akademicka 13, 20-934 Lublin

SUMMARY

The distribution of standardised ML estimate of mean in bivariate normal sample with monotone pattern of missing data is considered and compared with the distribution of standardised complete case estimate. The plots of pdf's and calculations of cdf's are given for chosen small sample sizes.

KEY WORDS: bivariate normal distribution, missing data, maximum likelihood estimate.

1. Introduction

Let us assume that the joint distribution of random variables (X, Y) is normal with unknown mean $\boldsymbol{\mu} = [\mu_1, \mu_2]'$ and variance-covariance matrix

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{bmatrix}.$$

Suppose we have paired observations $\{(x_i, y_i), i = 1 \ldots m\}$ on both variables X and Y and k additional observations $\{x_i, i = m+1, \ldots, m+k\}$ on X alone. It means we have "monotone pattern of missing data" (Little and Rubin, 1987). Let us also assume that "missing data" $\{y_i, i = m+1, \ldots, m+k\}$ are missing completely at random (Rubin, 1976). Maximum likelihood estimate (MLE) of μ_2 is of the form

$$\widehat{\mu_2} = \bar{Y_0} + \hat{\beta} \cdot (\bar{X} - \bar{X_0}),$$

where $\bar{Y_0} = \frac{1}{m}\sum_{i=1}^{m} y_i$, $\bar{X_0} = \frac{1}{m}\sum_{i=1}^{m} x_i$, $\bar{X} = \frac{1}{m+k}\sum_{i=1}^{m+k} x_i$, $\hat{\beta} = \frac{\sum_{i=1}^{m}(x_i - \bar{X_0})(y_i - \bar{Y_0})}{\sum_{i=1}^{m}(x_i - \bar{X_0})^2}$. The "complete case" estimate of μ_2, ignoring additional observations on X, is $\widetilde{\mu_2} = \bar{Y_0}$.

It is known (Morrison, 1971) that $\widehat{\mu_2}$ is an unbiased estimate of μ_2 with variance

$$Var(\widehat{\mu_2}) = \frac{\sigma_2^2}{m}[1 + \frac{k}{(m+k)(m-3)}(1-\rho^2(m-2))].$$

So, assuming $m > 3$, ML estimate has got less variance than "complete case" estimate iff $\rho^2 > \frac{1}{m-2}$. Little (1976) investigated the distribution of $\frac{\widehat{\mu_2}-\mu_2}{\sqrt{asyvar(\widehat{\mu_2}-\mu_2)}}$ in order to establish the approximated procedures for confidence intervals and testing hypotheses concerning μ_2.

Now we are interested in comparison of distributions of standardised MLE and "complete case" estimates, it means $\frac{\widehat{\mu_2}-\mu_2}{\sigma_2}$ and $\frac{\bar{Y_0}-\mu_2}{\sigma_2}$. Our special interest is in small sample size.

2. Results

Because $\frac{\sigma_1\sqrt{m-1}(\hat{\beta}-\beta)}{\sigma_2\sqrt{1-\rho^2}} \sim t_{m-1}$ (Cramer,1966), we have $\hat{\beta}\sigma_1 = \sigma_2\rho + \sigma_2\sqrt{\frac{1-\rho^2}{m-1}} \cdot t$, where $t \sim t_{m-1}$ (Student's t distribution with $m-1$ df). So, $W = \frac{\widehat{\mu_2}-\mu_2}{\sigma_2}$ can be written as

$$W = \frac{\bar{Y_0}-\mu_2}{\sigma_2} + \frac{k}{m+k}(\rho + \sqrt{\frac{1-\rho^2}{m-1}} \cdot t) \cdot \frac{\bar{X_1}-\bar{X_0}}{\sigma_1},$$

where $\bar{X_1} = \frac{1}{k}\sum_{i=m+1}^{m+k} x_i$. Thus W can be expressed as

$$W = V + (\rho + \sqrt{\frac{1-\rho^2}{m-1}} \cdot t)Z,$$

where (V,Z) and t are independently distributed,

$$t \sim t_{m-1}$$

$$\begin{bmatrix} V \\ Z \end{bmatrix} \sim \mathcal{N}(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \frac{1}{m} & -\rho\frac{k}{m(k+m)} \\ -\rho\frac{k}{m(k+m)} & \frac{k}{m(k+m)} \end{bmatrix}).$$

The probability density function (pdf) of W can be found by integrating $f(w) = \int_{-\infty}^{\infty} g(w \mid t)h(t)dt$, where $h(t)$ is pdf of t_{m-1} and $g(w \mid t)$ is pdf of normal distribution with mean 0 and variance $\frac{1}{m} + \frac{k}{m(m+k)}(t^2\frac{1-\rho^2}{m-1} - \rho^2)$.

The pdf function $f(w)$ for chosen values of m, k, ρ was calculated by means of MAPLE V Release program. Plots are given in the figures 1 and 2 (for $w > 0$ only because $f(w)$ is symmetric at zero).

The cumulative distribution function (cdf) $F(w)$ of random variable W is given by double integral.

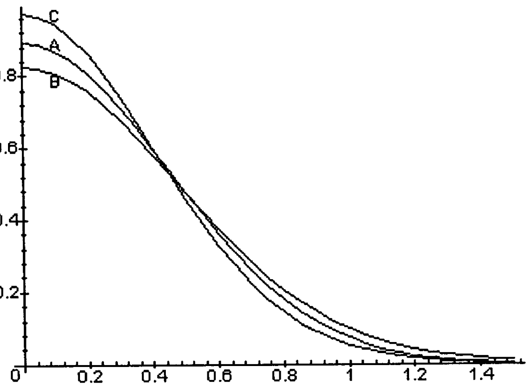

Fig. 1. The pdf's of **A:** $\frac{\bar{Y}_0-\mu_2}{\sigma_2}$; **B:** $\frac{\widehat{\mu_2}-\mu_2}{\sigma_2}, \rho^2 = 0$; **C:** $\frac{\widehat{\mu_2}-\mu_2}{\sigma_2}, \rho^2 = 0.5$. m=k=5.

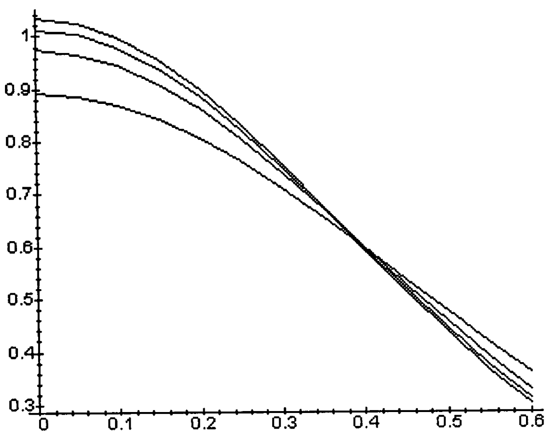

Fig. 2. The pdf's of $\frac{\widehat{\mu_2}-\mu_2}{\sigma_2}, \rho^2 = 0.5$ for $m = 5$ and (from bottom to top), $k = 0, 5, 10, 15$.

A good practical measure of the precision of the estimate $\widehat{\mu_2}$ can be its relative error $\mid \frac{\widehat{\mu_2}-\mu_2}{\mu_2} \mid$. Exactly speaking, the probability that the relative error does not exceed an admissible value ϵ can be a good measure of accuracy.

For a certain ϵ we have $Pr(\mid \frac{\widehat{\mu_2}-\mu_2}{\mu_2} \mid < \epsilon) = 2F(\frac{\epsilon}{|\nu_2|}) - 1$, where ν_2 is the coefficient of variation $\frac{\sigma_2}{\mu_2}$. For the "complete case" estimate of μ_2 we have $Pr(\mid \frac{\bar{Y}_0-\mu_2}{\mu_2} \mid < \epsilon) = 2\Phi(\frac{\epsilon\sqrt{m}}{|\nu_2|}) - 1$, where $\Phi(\cdot)$ is cdf of standard normal distribution. So, $\widehat{\mu_2}$ is superior to $\bar{Y}_0$ under relative error as the criterion, when $F(\frac{\epsilon}{|\nu_2|}) > \Phi(\frac{\epsilon\sqrt{m}}{|\nu_2|})$. For sufficiently large values of ρ^2, estimate $\widehat{\mu_2}$ becomes better than $\bar{Y_0}$. However, it is impossible to give an analytic formula for such a "critical value" of ρ^2 as it is very complicated function of $\frac{\epsilon}{|\nu_2|}, m$ and k.

Table 1 gives the values of $Pr(|\frac{\widehat{\mu_2}-\mu_2}{\mu_2}|<\epsilon)$ for $m=k=5$, $\frac{\epsilon}{|\nu_2|}=0.1(0.1)1$, $\rho^2=0(0.1)0.9$ compared with $Pr(|\frac{\bar{Y_0}-\mu_2}{\mu_2}|<\epsilon)$ in the last column. Tables 2 and 3 give the same for $m=5, k=10$ and $m=k=10$, respectively.

Table 1. $Pr(|\frac{\widehat{\mu_2}-\mu_2}{\mu_2}|<\epsilon)$ for $m=k=5$

$\frac{\epsilon}{\vert\nu_2\vert}$ \ ρ^2	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	$2\Phi(\frac{\epsilon\sqrt{m}}{\vert\nu_2\vert})-1$
0.1	.164	.169	.174	.180	.186	.193	.201	.210	.220	.233	.177
0.2	.321	.330	.340	.350	.362	.374	.389	.405	.424	.446	.345
0.3	.465	.477	.490	.504	.519	.535	.554	.575	.598	.625	.498
0.4	.591	.605	.619	.635	.651	.670	.690	.712	.736	.763	.629
0.5	.697	.711	.725	.741	.758	.775	.794	.815	.837	.860	.736
0.6	.782	.795	.809	.823	.838	.854	.870	.887	.905	.923	.820
0.7	.848	.859	.871	.883	.896	.909	.922	.935	.948	.961	.882
0.8	.897	.906	.916	.926	.935	.945	.955	.964	.973	.981	.926
0.9	.931	.939	.946	.954	.961	.968	.975	.981	.987	.992	.956
1	.955	.961	.967	.972	.977	.982	.986	.990	.994	.997	.975

Table 2. $Pr(|\frac{\widehat{\mu_2}-\mu_2}{\mu_2}|<\epsilon)$ for $m=5, k=10$

$\frac{\epsilon}{\vert\nu_2\vert}$ \ ρ^2	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	$2\Phi(\frac{\epsilon\sqrt{m}}{\vert\nu_2\vert})-1$
0.1	.161	.167	.174	.182	.190	.200	.212	.227	.245	.268	.177
0.2	.316	.327	.340	.354	.369	.388	.409	.435	.466	.506	.345
0.3	.458	.473	.489	.508	.529	.552	.579	.611	.649	.695	.498
0.4	.582	.600	.618	.639	.662	.687	.716	.748	.785	.828	.629
0.5	.687	.705	.724	.745	.767	.791	.818	.846	.878	.912	.736
0.6	.773	.789	.807	.826	.846	.867	.889	.912	.935	.958	.820
0.7	.839	.854	.869	.885	.901	.918	.935	.952	.968	.982	.882
0.8	.888	.901	.913	.926	.939	.951	.963	.974	.985	.993	.926
0.9	.924	.934	.944	.954	.963	.972	.980	.987	.993	.997	.956
1	.949	.957	.964	.971	.978	.984	.989	.993	.997	.999	.975

It is obvious that for sufficiently large values of ρ^2, increasing the number of additional observations for variable Y (increasing k) improves the estimate $\widehat{\mu_2}$. However there exists "the limit" for $Pr(|\frac{\widehat{\mu_2}-\mu_2}{\mu_2}|<\epsilon)$. These limiting values can be calculated by taking $k=\infty$ in cumulative distribution function $F(\cdot)$. Such values of $Pr(|\frac{\widehat{\mu_2}-\mu_2}{\mu_2}|<\epsilon)$ are given for $m=5$ and $m=10$ in tables 4 and 5, respectively.

Table 3. $Pr(\mid \frac{\widehat{\mu_2}-\mu_2}{\mu_2} \mid < \epsilon)$ for $m = k = 10$

$\frac{\epsilon}{\lvert\nu_2\rvert}$ \ ρ^2	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	$2\Phi(\frac{\epsilon\sqrt{m}}{\lvert\nu_2\rvert}) - 1$
0.1	.241	.247	.254	.262	.270	.279	.289	.301	.314	.328	.248
0.2	.460	.471	.483	.496	.510	.525	.542	.560	.581	.603	.473
0.3	.642	.655	.669	.684	.699	.716	.734	.753	.774	.796	.657
0.4	.779	.792	.805	.818	.832	.847	.862	.877	.893	.910	.794
0.5	.874	.884	.894	.905	.915	.926	.936	.946	.956	.966	.886
0.6	.933	.940	.947	.954	.961	.968	.974	.979	.984	.989	.942
0.7	.967	.972	.976	.980	.984	.987	.990	.993	.995	.997	.973
0.8	.985	.988	.990	.992	.994	.996	.997	.998	.999	.999	.989
0.9	.994	.995	.996	.997	.998	.999	.999	.999	1	1	.996
1	.998	.998	.999	.999	.999	1	1	1	1	1	.998

Table 4. $Pr(\mid \frac{\widehat{\mu_2}-\mu_2}{\mu_2} \mid < \epsilon)$ for $m = 5$, $k \to \infty$

$\frac{\epsilon}{\lvert\nu_2\rvert}$ \ ρ^2	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	.157	.165	.175	.187	.201	.220	.245	.281	.341	.466
0.2	.307	.322	.341	.362	.389	.422	.466	.527	.619	.780
0.3	.445	.466	.490	.519	.553	.594	.646	.714	.805	.926
0.4	.567	.591	.619	.650	.686	.729	.780	.840	.909	.976
0.5	.671	.696	.723	.754	.789	.827	.870	.915	.960	.991
0.6	.576	.780	.805	.833	.862	.893	.926	.956	.982	.996
0.7	.823	.844	.866	.889	.912	.936	.958	.977	.991	.998
0.8	.874	.891	.909	.927	.945	.962	.976	.987	.995	.999
0.9	.911	.926	.940	.953	.966	.977	.986	.993	.997	.999
1	.938	.949	.960	.969	.978	.985	.991	.995	.998	1

Table 5. $Pr(\mid \frac{\widehat{\mu_2}-\mu_2}{\mu_2} \mid < \epsilon)$ for $m = 10$, $k \to \infty$

$\frac{\epsilon}{\lvert\nu_2\rvert}$ \ ρ^2	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	.235	.247	.262	.279	.301	.328	.364	.415	.496	.655
0.2	.450	.471	.496	.525	.559	.602	.655	.724	.817	.939
0.3	.630	.655	.683	.715	.752	.794	.842	.896	.952	.994
0.4	.767	.791	.817	.845	.875	.907	.939	.968	.991	1
0.5	.863	.882	.903	.923	.944	.963	.980	.992	.998	1
0.6	.924	.939	.952	.965	.977	.987	.994	.998	1	1
0.7	.961	.970	.978	.986	.991	.996	.998	1	1	1
0.8	.981	.986	.991	.994	.997	.999	1	1	1	1
0.9	.991	.994	.996	.998	.999	1	1	1	1	1
1	.996	.997	.998	.999	1	1	1	1	1	1

At the end it should be admitted that presented results are rather theoretical. They can be of practical importance only if we have some knowledge, especially about correlation between X and Y.

3. Example

Consider the situation that we have only five complete pairs of observations ($m = 5$). Let us assume we know that $|\nu_2| > 0.5$ and $\rho^2 < 0.5$. We can not expect that $Pr(| \frac{\widehat{\mu_2} - \mu_2}{\mu_2} | < 0.1)$ is greater than 0.422 no matter how many additional observations for Y is taken (see Table 4).

References

Cramer H. (1966). *Mathematical Methods of Statistics.* Princ. Univ. Press.

Little R.J.A. (1976). Inference about mean from incomplete multivariate data. *Biometrika* **63**, 594-604.

Little R.J.A., Rubin D.B (1987). *Statistical Analysis with Missing Data.* J. Wiley & Sons, New York.

Morrison D.F. (1971). Expectations and variances of maximum likelihood estimates of the multivariate normal distribution parameters with missing data. *J. Am. Stat. Assoc.* **66**, 602-604.

Rubin D.B. (1976). Inference and missing data. *Biometrika* **63**, 581-592.

Received 10 November 2004

O porównaniu estymatorów średnich w dwuwymiarowej próbie normalnej z monotonicznym układem brakujących obserwacji

Streszczenie

W pracy porównano rozkład standaryzowanego estymatora największej wiarygodności średniej w dwuwymiarowej próbie z monotonicznym układem brakujących obserwacji z rozkładem estymatora dla próby kompletnej. Podano wykresy funkcji gęstości prawdopodobieństwa oraz obliczenia wartości dystrybuanty rozkładów dla wybranych małych liczebności prób.

Słowa kluczowe: dwuwymiarowy rozkład normalny, brakujące dane, estymator największej wiarygodności.