RASHI 2 (2) : 23-28 (2017)

On a Procedure of Estimation of Data-Dispersion Matrix under Complex Spatial Structure Prevalent in Field Experiments

Satyabrata Pal

Honorary Visiting Professor, ISEC, Indian Statistical Institute, Kolkata-700108
Formerly, Dean, Post Graduate Studies and Professor, BCKV, Nadia, W.B., Pin-741252

Kajal Dihidar

Assistant Professor, SOSU, Indian Statistical Institute, Kolkata -700108

ABSTRACT

The estimation of the elements of the data-dispersion matrix poses a real problem when the spatial structure assumes a complex form. Motivated to find a solution to the above problem, a method of estimation of the data dispersion matrix is proposed. Though the method is illustrated taking a specific form of the data-dispersion matrix in a particular case, the method (illustrated in the Introduction section) is quite general and is applicable to any form (whatever be) of the data-dispersion matrix impersonating even complex spatial structures.

1. INTRODUCTION

This paper considers the problem of estimation of the data-dispersion matrix in case of known (unknown) complex spatial structure governing the data emanated from agricultural field experiments under controlled conditions. Heterogeneity is induced in data under field layouts when they become independent. About a century ago, Fisher realised the presence of correlation in data collected from contiguous plots. He has proposed the fundamental principles (randomisation, replication and blocking) of design of experiments in order that the data (generated after the application of the three principles) become amenable to simple analysis (by application of the analysis of variance technique, which he himself proposed). The role of the principle of randomisation is supposed to remove bias, so to say, induce homogeneity among plots within a block. It has been an observed fact over decades that the data remain correlated even after the application of the three fundamental principles. As the existence of serial correlation is a reality in data obtained from controlled field experiments, it is necessary to generate procedures of analysis when the correlation structure is of complex nature. Under complex correlation structure satisfied by the field experimental data (when such structures are known or when those are assumed to be satisfied with respect to the experimental data available on hand) the parameters (functions of the first order serial correlation) contained as elements of the data-dispersion matrix (known or assumed as mentioned above) becomes difficult to estimate by the application of the method of maximum likelihood as the estimating equations so obtained are not easily tractable. This paper proposes a procedure of estimation of the data-dispersion matrix under such situations delineated above by invoking the simulation technique.

Data are arranged in rows and columns (treatments are allotted on plots in rows which are actually terraces and different terraces represent the blocks (RBD case). In case when CRD (with equal replication) is used such terraces do represent replications keeping the randomisation structure of treatments imposed (in terms of their plot-wise positions) in the first terrace as fixed and considers the respective positions in the terraces along the down ward slopes generating, so to say, replications corresponding to the treatment-lay-out (as in the first terrace). Under the above spatial situations (terrace cultivation), it is natural to consider existence of correlation $\left(\rho_1^1 = \rho_1\right)$ in data emanating from contiguous plots within a terrace and also existence of correlation $\left(\rho_2^1 = \rho_2\right)$ from contiguous plots lying in different terraces. In the background of the above premise, we may consider that the observations, y_{ij}'s are generated from a multivariate normal (MN) distribution with some mean (depending on the set-up of the experiment, completely randomised design or randomised block design as referred to above) and Dispersion Matrix (DM), Σ. Considering the CRD model, $y_{ij} = \mu + \alpha_i + e_{ij}$, where e_{ij} 's ~ MN $(0, \Sigma)$, and y_{ij} 's ~ MN (β, Σ), β being the mean vector, for all i and j ("i" represents the treatment number and "j" represents replication number).

Superimposing the structure mentioned above, the elements of the Σ matrix are as follows:

V $(y_{ij}) = \sigma^2$, for all i and j; Covariance (y_{ij} and $y_{i+l,\, j+k}$) $= \sigma^2,\ \rho_1^l,\ \rho_2^k$; where Cov.(y_{ij} and $y_{i+1,\, j}$) $= \sigma^2,\ \rho_1^1$ and Cov.(y_{ij} and $y_{i,\, j+1}$) $= \sigma^2,\ \rho_2^1$ (1.1)

E-mail: spbbr18@gmail.com

In a CRD set up as mentioned above if we take 4 treatments and 3 replications per treatment (the first terrace determines the arrangement of treatments (after randomisation) and the terraces lying below the first terrace follow the same arrangement of treatments, thus the respective plots on the terraces (below the first terrace) can effectively be considered as replications with respect to the treatments considered in the first terrace, the structure of the Σ matrix is as follows :

In particular, 12 plot-positions along the first row (also in other rows) follow the order (of the observations) as mentioned in the arrangement just below (with respect to the SIGMA matrix as above) :

[Tr, Rep] : (1, 1) (1, 2) (1, 3) (2, 1) (2, 2) (2, 3) (3, 1) (3, 2) (3, 3) (4, 1) (4, 2) (4, 3)

Also, 12 plot-positions along the first column (also in other columns) follow the order (of the observations) as mentioned in the arrangement just below (with respect to the SIGMA matrix as above):

[Tr, Rep]: [(1, 1) (1, 2) (1, 3) (2, 1) (2, 2) (2, 3) (3, 1) (3, 2) (3, 3) (4, 1) (4, 2) (4, 3)] - Transposed (perpendicular arrangement).

The form of the theoretical SIGMA matrix (with entries represented in terms of?_(1 &) ?_2) is presented below (in particular, Covariance (y_{ij} and $y_{i+l,\, j+k}$) = σ^2, ρ_1^l, ρ_2^k). Indeed, cov ((i, j) and (l,k)) = $(\rho_1^{|l-i|}\rho_2^{|k-j|})\,\sigma^2$ (1.2)

is used to denote the elements of the SIGMA matrix as below.

$\Sigma = \sigma^2$

	(1, 1)	(1, 2)	(1, 3)	(2, 1)	(2, 2)	(2, 3)	(3, 1)	(3, 2)	(3, 3)	(4, 1)	(4, 2)	(4, 3)
(1, 1)												
(1, 2)					ρ_1							
(1, 3)									ρ_1^2			
(2, 1)												
(2, 2)												
(2, 3)												
(3, 1)			$\rho_1^2\rho_2^2$									
(3, 2)											ρ_1	
(3, 3)			ρ_1^2									
(4, 1)												
(4, 2)							$\rho_1\rho_2$					
(4, 3)												

cov ((i,j) and (l, k)) = $(\rho_1^{|l-i|}\rho_2^{|k-j|})\,\sigma^2$; cov ((1, 2) and (2, 2)) = $(\rho_1\rho_2^0)\sigma^2 = \sigma^2\rho_1$; cov ((1, 3) and (3, 3)) = $(\rho_1^2\rho_2^0)\sigma^2 = \sigma^2\rho_1^2$; cov ((3, 1) and (1, 3)) = $(\rho_1^2\rho_2^2)\sigma^2 = \sigma^2\rho_1^2\rho_2^2$; cov((3, 2) and (4, 2)) = $(\rho_1^1\rho_2^0)\sigma^2 = \sigma^2\rho_1$; cov ((3, 3) and (1, 3)) = $(\rho_1^2\rho_2^0)\sigma^2 = \sigma^2\rho_1^2$; cov ((4, 2) and (3, 1)) = $(\rho_1^1\rho_2^1)\sigma^2 = \sigma^2\rho_1\rho_2$

The application of the model ((1.1) in conjunction with (1.2)) can be found in Pal, S., and Basak, S. (2016), and Pal, et al (2016). In Pal et al. (2017), a variant of (1.2) has been used.

The method of estimation of the elements of the Σ matrix can be effected by considering a large number of simulations(with in each simulation period generating several samples of eij's from multivariate normal population with mean vector 0 and the above mentioned dispersion matrix Σ).

Illustrationson the theoretical SIGMA and estimated SIGMA are presented for some particular cases in Table 1 to Table 4.

The precision of the estimated SIGMA is determined by the criterion D* defined below:

Let $\Sigma = (\sigma_{ij})$ and $\hat{\Sigma} = (\widehat{\sigma_{ij}})$, then D = $\Sigma_{ij}|\sigma_{ij} - \widehat{\sigma_{ij}}|$ and D* = $\frac{1}{n}\Sigma_{ij}\left|\sigma_{ij} - \widehat{\sigma_{ij}}\right|$, n being the total number of elements in the SIGMA matrix, The values of the criteria, D and D* for different combinations of values of ρ_1 and ρ_2 are given in Table 5 (Section 2, titled, Results and Conclusion).

The two matrices, Σ and $\hat{\Sigma}$ are also compared on the basis of their eigen values (minimum and maximum) in order to detect possible discrepancies in the values of the two matrices Σ and $\hat{\Sigma}$ and such results are also presented in Table 5.

Table 1: Sigma matrix for rho 1 = 0.4 and rho 2 = 0.35 and v = 4, r = 3

	(1, 1)	(1, 2)	(1, 3)	(2, 1)	(2, 2)	(2, 3)	(3, 1)	(3, 2)	(3, 3)	(4, 1)	(4, 2)	(4, 3)
1	1	0.34	0.1156	0.4	0.136	0.04624	0.16	0.0544	0.018496	0.064	0.02176	0.007398
2	0.34	1	0.34	0.136	0.4	0.136	0.0544	0.16	0.0544	0.02176	0.064	0.02176
3	0.1156	0.34	1	0.04624	0.136	0.4	0.018496	0.0544	0.16	0.007398	0.02176	0.064
4	0.4	0.136	0.04624	1	0.34	0.1156	0.4	0.136	0.04624	0.16	0.0544	0.018496
5	0.136	0.4	0.136	0.34	1	0.34	0.136	0.4	0.136	0.0544	0.16	0.0544
6	0.04624	0.136	0.4	0.1156	0.34	1	0.04624	0.136	0.4	0.018496	0.0544	0.16
7	0.16	0.0544	0.018496	0.4	0.136	0.04624	1	0.34	0.1156	0.4	0.136	0.04624
8	0.0544	0.16	0.0544	0.136	0.4	0.136	0.34	1	0.34	0.136	0.4	0.136
9	0.018496	0.0544	0.16	0.04624	0.136	0.4	0.1156	0.34	1	0.04624	0.136	0.4
10	0.064	0.02176	0.007398	0.16	0.0544	0.018496	0.4	0.136	0.04624	1	0.34	0.1156
11	0.02176	0.064	0.02176	0.0544	0.16	0.0544	0.136	0.4	0.136	0.34	1	0.34
12	0.007398	0.02176	0.064	0.018496	0.0544	0.16	0.04624	0.136	0.4	0.1156	0.34	1

Table 2: Sigma_hat matrix for rho1 = 0.4 and rho2 = 0.35

	(1, 1)	(1, 2)	(1, 3)	(2, 1)	(2, 2)	(2, 3)	(3, 1)	(3, 2)	(3, 3)	(4, 1)	(4, 2)	(4, 3)
1	0.997518	0.340112	0.115919	0.398033	0.134678	0.052773	0.161433	0.059913	0.029168	0.061364	0.01928	0.00362
2	0.340112	1.009186	0.347815	0.13397	0.396958	0.138803	0.043591	0.159619	0.047521	0.021472	0.058579	0.011577
3	0.115919	0.347815	0.981837	0.035162	0.134688	0.393845	0.012753	0.05029	0.148206	0.011218	0.019539	0.055559
4	0.398033	0.13397	0.035162	0.989409	0.345481	0.117613	0.391138	0.147225	0.051475	0.156942	0.049485	0.018576
5	0.134678	0.396958	0.134688	0.345481	1.008009	0.34873	0.124849	0.403433	0.139045	0.052096	0.156055	0.052479
6	0.052773	0.138803	0.393845	0.117613	0.34873	0.995291	0.045611	0.144138	0.400836	0.014197	0.057205	0.156135
7	0.161433	0.043591	0.012753	0.391138	0.124849	0.045611	0.992496	0.323975	0.109136	0.409042	0.128134	0.036753
8	0.059913	0.159619	0.05029	0.147225	0.403433	0.144138	0.323975	0.993261	0.342057	0.140256	0.391232	0.136565
9	0.029168	0.047521	0.148206	0.051475	0.139045	0.400836	0.109136	0.342057	1.006889	0.048046	0.139465	0.409613
10	0.061364	0.021472	0.011218	0.156942	0.052096	0.014197	0.409042	0.140256	0.048046	1.016574	0.345732	0.11025
11	0.01928	0.058579	0.019539	0.049485	0.156055	0.057205	0.128134	0.391232	0.139465	0.345732	0.995206	0.344778
12	0.00362	0.011577	0.055559	0.018576	0.052479	0.156135	0.036753	0.136565	0.409613	0.11025	0.344778	1.013253

Table 3: Sigma matrix for rho1 = 0.45 and rho2 = 0.43

	(1, 1)	(1, 2)	(1, 3)	(2, 1)	(2, 2)	(2, 3)	(3, 1)	(3, 2)	(3, 3)	(4, 1)	(4, 2)	(4, 3)
1	1	0.43	0.1849	0.45	0.1935	0.083205	0.2025	0.087075	0.037442	0.091125	0.039184	0.016849
2	0.43	1	0.43	0.1935	0.45	0.1935	0.087075	0.2025	0.087075	0.039184	0.091125	0.039184
3	0.1849	0.43	1	0.083205	0.1935	0.45	0.037442	0.087075	0.2025	0.016849	0.039184	0.091125
4	0.45	0.1935	0.083205	1	0.43	0.1849	0.45	0.1935	0.083205	0.2025	0.087075	0.037442
5	0.1935	0.45	0.1935	0.43	1	0.43	0.1935	0.45	0.1935	0.087075	0.2025	0.087075
6	0.083205	0.1935	0.45	0.1849	0.43	1	0.083205	0.1935	0.45	0.037442	0.087075	0.2025
7	0.2025	0.087075	0.037442	0.45	0.1935	0.083205	1	0.43	0.1849	0.45	0.1935	0.083205
8	0.087075	0.2025	0.087075	0.1935	0.45	0.1935	0.43	1	0.43	0.1935	0.45	0.1935
9	0.037442	0.087075	0.2025	0.083205	0.1935	0.45	0.1849	0.43	1	0.083205	0.1935	0.45
10	0.091125	0.039184	0.016849	0.2025	0.087075	0.037442	0.45	0.1935	0.083205	1	0.43	0.1849
11	0.039184	0.091125	0.039184	0.087075	0.2025	0.087075	0.1935	0.45	0.1935	0.43	1	0.43
12	0.016849	0.039184	0.091125	0.037442	0.087075	0.2025	0.083205	0.1935	0.45	0.1849	0.43	1

Table 4: Sigma_hat Matrix for ρ_1

	(1, 1)	(1, 2)	(1, 3)	(2, 1)	(2, 2)	(2, 3)	(3, 1)	(3, 2)	(3, 3)	(4, 1)	(4, 2)	(4, 3)
1	1.007648	0.428615	0.187596	0.45096	0.188426	0.090594	0.202261	0.085946	0.048085	0.09254	0.03681	0.019832
2	0.428615	0.998226	0.42494	0.18898	0.447312	0.193619	0.087341	0.207547	0.087186	0.035953	0.083641	0.026677
3	0.187596	0.42494	0.978155	0.082697	0.194575	0.451154	0.036443	0.083156	0.198164	0.01585	0.027016	0.081485
4	0.45096	0.18898	0.082697	0.999333	0.429375	0.196279	0.440319	0.192954	0.090984	0.202053	0.074365	0.040266
5	0.188426	0.447312	0.194575	0.429375	0.998318	0.43617	0.188848	0.449846	0.198426	0.088835	0.198572	0.085135
6	0.090594	0.193619	0.451154	0.196279	0.43617	1.005539	0.080904	0.19407	0.453674	0.034103	0.080933	0.200153
7	0.202261	0.087341	0.036443	0.440319	0.188848	0.080904	0.979739	0.415659	0.17045	0.456333	0.190118	0.073039
8	0.085946	0.207547	0.083156	0.192954	0.449846	0.19407	0.415659	0.995678	0.428362	0.198871	0.44796	0.191232
9	0.048085	0.087186	0.198164	0.090984	0.198426	0.453674	0.17045	0.428362	1.002264	0.085545	0.197136	0.459938
10	0.09254	0.035953	0.01585	0.202053	0.088835	0.034103	0.456333	0.198871	0.085545	1.014177	0.435531	0.181725
11	0.03681	0.083641	0.027016	0.074365	0.198572	0.080933	0.190118	0.44796	0.197136	0.435531	1.002713	0.434792
12	0.019832	0.026677	0.081485	0.040266	0.085135	0.200153	0.073039	0.191232	0.459938	0.181725	0.434792	1.017151

RESULTS AND CONCLUSION

Table 5: Values of the D, D^* and other Coefficients

v	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
r	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
rho1	0.4	0.4	0.4	0.4	0.4	0.45	0.45	0.45	0.45	0.45	0.5	0.5	0.5	0.5	0.5	0.55	0.55	0.55	0.55	0.55
rho2	0.34	0.35	0.36	0.37	0.38	0.39	0.4	0.41	0.42	0.43	0.44	0.45	0.46	0.47	0.48	0.49	0.5	0.51	0.52	0.53
Det(sigma)	0.0779	0.0732	0.0686	0.0641	0.0598	0.0349	0.0323	0.0299	0.0276	0.0254	0.0134	0.0123	0.0112	0.0102	0.0092	0.0043	0.0039	0.0035	0.0031	0.0028
Det(sigma_hat)	0.0769	0.0722	0.0677	0.0633	0.0590	0.0344	0.0319	0.0295	0.0273	0.0251	0.0132	0.0121	0.0111	0.0101	0.0091	0.0043	0.0039	0.0035	0.0031	0.0028
Sum of element wise Absolute Difference (D)	0.7736	0.7429	0.6928	0.7166	0.7213	0.7595	0.7591	0.7195	0.7231	0.6789	0.7165	0.7032	0.7245	0.7008	0.6903	0.7223	0.6974	0.6995	0.6684	0.6817
Dstar = D / Total number of elements	0.0054	0.0052	0.0048	0.0050	0.0050	0.0053	0.0053	0.0050	0.0050	0.0047	0.0050	0.0049	0.0050	0.0049	0.0048	0.0050	0.0048	0.0049	0.0046	0.0047
min(eigen(sigma))	0.2724	0.2672	0.2620	0.2569	0.2518	0.2200	0.2155	0.2111	0.2067	0.2023	0.1751	0.1713	0.1675	0.1637	0.1600	0.1370	0.1338	0.1305	0.1274	0.1242
min(eigen (sigma_hat))	0.2747	0.2695	0.2642	0.2591	0.2539	0.2219	0.2174	0.2129	0.2084	0.2040	0.1767	0.1728	0.1690	0.1652	0.1614	0.1382	0.1349	0.1317	0.1285	0.1253
max(eigen(sigma))	2.8015	2.8340	2.8667	2.8997	2.9329	3.1796	3.2156	3.2518	3.2883	3.3251	3.6008	3.6407	3.6808	3.7212	3.7618	4.0697	4.1137	4.1581	4.2027	4.2476
max(eigen (sigma_hat))	2.7917	2.8241	2.8567	2.8895	2.9226	3.1683	3.2042	3.2403	3.2767	3.3133	3.5880	3.6277	3.6677	3.7079	3.7484	4.0551	4.0990	4.1431	4.1876	4.2323

A critical investigation in the values (Table 5) of the coefficients (D and D^*)reveals only miniscule differences in the values of the true and estimated matrices (also very small differences are found to exist in the values of the minimum and maximum eigen values of the Σ and $\hat{\Sigma}$ matrices). Thus this study concludes that matrix can be regarded as a valid estimate (based on the above criterion) in further application of analysis of variance technique in complex spatial situations.

REFERENCES

Pal, S., and Basak, S. (2016). "Optimum Plot Size–Robust Determination". Asian Academic Research Journal of Multi Disciplinary, **3(1)**: 278-90.

Pal, S., Basu, A., Mandal, G., Chakraborty, A. and Ghosh, A. (2016),"A state of the Art Review on Methods of Determination of Robust Optimum Plot Size". *International Journal of Science & Technology,* **6(4)**:1-10.

Pal, S., Dihidar, K., Mandal, G., Basak, S. Ghosh, A., Chakraborty, A.K.and Pal, S. (2017), " A Novel Approach in Robust Estimation of Optimum Sizeof Plots"International. *Journal of Agricultural and Statistical.Sciiences.* **13 (2)** : 423-29