Math-Net.Ru

Общероссийский математический портал

T. Sugiyama, Distributions and test of each characteristic root in principal component analysis, *Теория вероятн. и ее примен.*, 2010, том 55, выпуск 3, 607–608

DOI: <https://doi.org/10.4213/tvp4257>

Использование Общероссийского математического портала Math-Net.Ru подразумевает, что вы прочитали и согласны с пользовательским соглашением
<http://www.mathnet.ru/rus/agreement>

Параметры загрузки:

IP: 54.162.85.209

26 апреля 2023 г., 13:23:52

the switching options of regulation and deregulation for pollution, the switching dynamics of environmental policy is described by an alternating renewal process. It is shown that the double Laplace transform of transition density of a pollutant level can be obtained by a novel application of renewal theory. The expected level of overall pollutants is then calculated numerically and found to exhibit either a Λ -shaped or an N -shaped pattern in time. Our results present a simple explanation for the environmental Kuznets curve within a real options framework.

Sugiyama T. (Chuo University, Tokyo; *e-mail*: takakazu@math.chuo-u.ac.jp). **Distributions and test of each characteristic root in principal component analysis.** Principal components analysis is a data dimension reduction procedure that transforms the data representing a number of correlated variables into a new set of uncorrelated variables called principal components, and is to find the linear combinations with large variance. Let $\ell_1, \dots, \ell_p$ ($\ell_1 > \dots > \ell_p$) be the characteristic roots of the Wishart matrix $\mathbf{A} = n\mathbf{S}$, where $\mathbf{S}$ is the sample covariance matrix. Let $\lambda_1, \dots, \lambda_p$ be the characteristic roots of $\mathbf{\Sigma}$ ($\lambda_1 > \dots > \lambda_p > 0$). We apply principal component analysis for the covariance matrix $\mathbf{S}$, namely based on $\mathbf{S}$, we calculate the principal components and their variances. Each principal component accounts for a decreasing proportion of the entire variance inherent in the data; thus, the first principal component represents the largest variation (variance $n^{-1}\ell_1$), the second principal component represents the second largest variation (variance $n^{-1}\ell_2$), and so on. So, the distributions of each variance (characteristic root) are important.

The exact distribution of the largest characteristic root ℓ_1 is given in [1], [2]:

$$F(\ell_1 < nx) = \frac{|\mathbf{\Sigma}|^{n/2} \Gamma_p((p+1)/2)}{2^{np/2} \Gamma_p((n+p+1)/2)} \exp\left(-\frac{nx}{2} \text{tr } \mathbf{\Sigma}\right) \cdot {}_1F_1\left(\frac{p+1}{2}; \frac{n+p+1}{2}; \frac{nx}{2} \mathbf{\Sigma}^{-1}\right).$$

Now, the numerical computation of above distribution is possible until $p = 4$, $n = 50$. Sugiyama [3] also obtained the approximation $z_{1\alpha}$ of the upper 100α percentile point of the variance ℓ_1 , solving the following equation:

$$\prod_{i=1}^p \gamma\left(\frac{n-i+1}{2}; \frac{nz_{1\alpha}}{2\lambda_i}\right) / \Gamma\left(\frac{n-i+1}{2}\right) = \alpha,$$

where $\gamma(m/2; x/2)/\Gamma(m/2)$ is the characteristic distribution function of the chi-square distribution with m degrees of freedom. The above approximation is not so precise. So, Konishi and Sugiyama [4] obtained that $\ell_1^{1/3}$ is approximately distributed according to the normal distribution $\mathcal{N}(\lambda_1^{1/3}, 2\lambda_1^{2/3}/9)$.

We do not know the good mathematical results for the distributions of the second largest characteristic root ℓ_2 , the third ℓ_3 , and so on. Though we have these asymptotic distributions, it is not so simple. We wish to know the simple approximation of the distributions. When $\lambda_1 \gg \lambda_2 \gg \lambda_3 \gg \dots$ we propose the following χ^2 approximations:

$$\begin{aligned} \Pr(\ell_2 < z) &= \left[\gamma\left(\frac{n-1}{2}; \frac{nz}{2\lambda_2}\right) / \Gamma\left(\frac{n-1}{2}\right) \right], \\ \Pr(\ell_3 < z) &= \left[\gamma\left(\frac{n-2}{2}; \frac{nz}{2\lambda_3}\right) / \Gamma\left(\frac{n-2}{2}\right) \right], \quad \text{and so on.} \end{aligned} \tag{1}$$

To investigate the goodness of approximation, we consider the case $\lambda_1 = 100$, $\lambda_2 = 64$, $\lambda_3 = 25$, $\lambda_4 = \dots = \lambda_p = 1$, for the characteristic roots of $\mathbf{\Sigma}$, when $p = 10$.

Table 1. The probability based on the χ^2 approximation (1)

n	Second largest ℓ_2		Third largest ℓ_3	
	exact prob. 0.90	exact prob. 0.95	exact prob. 0.90	exact prob. 0.95
100	0.8847	0.9381	0.8973	0.9528
200	0.8939	0.9500	0.9047	0.9530

We see that approximations (1) are valid for $n = 200$. Obtaining the upper or lower percentile points of each characteristic root ℓ_j , we may consider the test of the null hypothesis $H_0: \lambda_j = \lambda_{j0}$.

REFERENCES

1. *Sugiyama T.* On the distribution of the largest latent root of the covariance matrix. — Ann. Math. Statist., 1967, v. 38, p. 1148–1151.
2. *Sugiyama T.* Distribution of the largest latent root and the smallest latent root of the generalized B statistic and F statistic in multivariate analysis. — Ann. Math. Statist., 1967, v. 38, p. 1152–1159.
3. *Sugiyama T.* Approximation for the distribution of the largest latent root of a Wishart matrix. — Austral. J. Statist., 1972, v. 14, p. 17–24.
4. *Konishi S., Sugiyama T.* Improved approximations to distributions of the largest and the smallest latent roots of a Wishart matrix. — Ann. Inst. Statist. Math., 1981, v. 33, № 1, p. 27–33.

Wakaki H. [*joint work with Shimizu H.*] (Graduate School of Science, Hiroshima University; *e-mail*: wakaki@math.sci.hiroshima-u.ac.jp). **An asymptotic expansion of the local power of a general test for testing a general covariance structure.** Let S be a $p \times p$ random matrix having a Wishart distribution $W_p(n, n^{-1}\Sigma)$. For testing a general covariance structure $\Sigma = \Sigma(\xi)$, we consider a class of test statistics $T_h = n \inf \rho_h(S, \Sigma(\xi))$, where $\rho_h(\Sigma_1, \Sigma_2) = \sum_{j=1}^p h(\lambda_j)$ is a divergence measure from Σ_1 to Σ_2 , λ_j 's are the eigenvalues of $\Sigma_1 \Sigma_2^{-1}$, and h is a given C^4 -class function on $(0, \infty)$ such that $h(1) = 0$, $h_1 = 0$, $h_2 = 1$, and $h(\lambda) > 0$ for any $\lambda \neq 1$, here h_r denotes the r -th derivative of h at $\lambda = 1$. It is assumed that all the elements of $\Sigma(\xi)$ are known C^4 -class functions on Ξ , an open subset of $\mathbf{R}^q$, and the Jacobian matrix of $\Sigma(\xi)$ is of full rank.

Wakaki, Eguchi, and Fujikoshi [1] derived an asymptotic expansion of the null distribution of T_h up to the order n^{-1} . Using the general asymptotic formula, they gave a condition for T_h to have a Bartlett adjustment factor.

We consider a sequence of local alternatives: $\Sigma = \Sigma_0 + n^{-1/2} \Sigma_0^{1/2} \Delta \Sigma_0^{1/2}$, where $\Sigma_0 = \Sigma(\xi_0)$, Δ is an arbitrarily fixed symmetric matrix such that $I_p + \Delta$ is positive definite. We derive an asymptotic expansion of the distribution function of T_h up to the order $n^{-1/2}$ under the local alternatives. Using the asymptotic formula, an asymptotic expansion of the power function of T_h is derived. Let β_h and β_{LR} be the power function of T_h and the likelihood ratio test, respectively. Then we find that

$$\begin{aligned} \beta_h = \beta_{LR} + \frac{1}{\sqrt{n}} \left(1 + \frac{h_3}{2} \right) & \left[\left\{ (p+1) \operatorname{tr}[D_\Delta] - \sum_{a,b=1}^q g^{ab} \operatorname{tr}[J_a J_b D_\Delta] \right\} g_{r+4}(\chi_\alpha^2; \tau) \right. \\ & \left. + \frac{1}{3} \operatorname{tr}[D_\Delta^3] g_{r+6}(\chi_\alpha^2; \tau) \right] + O(n^{-1}), \end{aligned}$$

where χ_α^2 is the upper α point of chi-square distribution with degree of freedom $r = p(p+1)/2 - q$, $g_s(x; \tau)$ is the noncentral chi-square density function with s degree of