Markov Chain Monte Carlo Methods

3. Statistical Concepts

K B Athreya, Mohan Delampady and T Krishnan

K B Athreya is a Professor at Cornell University. His research interests include mathematical analysis, probability theory and its application and statistics. He enjoys writing for *Resonance*. His spare time is spent listening to Indian classical music.

Mohan Delampady is at the Indian Statistical Institute, Bangalore. His research interests include robustness, nonparametric inference and computing in Bayesian statistics.

T Krishnan is now a full-time Technical Consultant to Systat Software Asia-Pacific Ltd., in Bangalore, where the technical work for the development of the statistical software Systat takes place. His research interests have been in statistical pattern recognition and biostatistics.

1. Introduction

In parts 1 and 2 of this series it was shown how Markov chain Monte Carlo (MCMC) methods can be employed to obtain satisfactory approximations for integrals that are not easy to evaluate analytically. Such integrals arise routinely in statistical problems. Some of the statistical concepts that are relevant for the application of MCMC methods and for understanding the examples to be discussed in Part 4 are explained in this part.

2. Inference for Multinomial Distribution

Recall the statistical inference problem for the binomial probability in a previous article (see [1]). If a statistical experiment involves n identical and independent trials, each of which can result in two types of outcomes (Yes or No, 1 or 0, Success or Failure, etc.) then the (random) number X of trials which result in, say, outcome of type 1 can be modelled as a binomial random variable with the probability distribution:

$$P(X = x \mid \mu) = \binom{n}{x} \mu^x (1 - \mu)^{n-x}; \quad x = 0, 1, \ldots, n.$$

where µ is the probability that any trial will result in outcome of type 1. The statistical problem in this case is to make inferences about µ from the data X.

How does one model the situation when there are more than two types of outcomes? This needs a generalization of the binomial distribution.

Example 1. In crosses between two kinds of maize, Lindstrom (cited in Snedecor and Cochran [2]) found

[1] Simple Monte Carlo, *Resonance*, Vol.8, No.4, p.17, 2003. The Markov Chain Case, *Resonance*, Vol.8, No.7, p.63, 2003.

four distinct types of plants in the second generation. The simple Mendelian model specifies the probabilities of these types as 9/16, 3/16, 3/16 and 1/16, respectively. In 1301 plants Lindstrom observed the number of these types to be $n_1 = 773$, $n_2 = 231$, $n_3 = 238$, and $n_4 = 59$. Are these observations compatible with the simple Mendelian model?

Example 2. A newly cast die (with dots 1–6 on the six different sides) is rolled n times, and the number of rolls leading to the different sides showing up are recorded. How does one check that the die is balanced and not loaded?

Example 3. Consider the following artificial problem. Take 2 coins, each having the same unknown probability p of coming up heads in any toss. Toss these two coins simultaneously n times. The possible outcomes for each trial are '2 Heads', '2 Tails', and 'One Head and One Tail'. What is the probability distribution of the number of occurrences of the different outcomes?

In all the examples above, one first needs to know the joint distribution of the vector N of the numbers of occurrences of the different outcomes. In Example 1, $N = (N_1; N_2; N_3; N_4)$, with $\sum_{i=1}^{4} N_i = 1301$. The given numbers $n_1; n_2; n_3; n_4$ are a realization of the random vector N. In Example 2, $N = (N_1; N_2; \ldots; N_6)$, where N_i is the number of rolls leading to side i being up. Here, $\sum_{i=1}^{6} N_i = n$. In Example 3, $N = (N_1; N_2; N_3)$, where suffix 1 corresponds to '2 Heads', suffix 2 to '2 Tails' and suffix 3 to 'One Head and One Tail'. Here $\sum_{i=1}^{3} N_i = n$.

Generalizing the binomial distribution to $k \geq 2$ types or categories to deal with questions like this leads to the notion of a multinomial distribution.

Keywords
Statistical inference, multinomial distribution, maximum likelihood estimation, Bayes theorem, Dirichlet distribution, posterior distribution, maximum a posteriori estimate, sufficiency, Rao-Blackwell Theorem, conditional expectation, Gibbs sampler, Poisson distribution.

Suppose a statistical experiment involves n identical and independent trials, each of which can result in $k \geq 2$ types of outcomes (type j, $j = 1, 2, \dots, k$). Let the probability that any trial will lead to outcome of type j be p_j, and the (random) number of trials (out of a total of n) which result in outcome of type j be denoted by N_j, $j = 1, 2, \dots, k$. Then the joint probability distribution of the vector $N = (N_1, N_2, \dots, N_k)$ is given by

$$P(N_1 = n_1, N_2 = n_2, \dots, N_k = n_k) =$$

$$\frac{n!}{\prod_{i=1}^{k} n_i!} \prod_{i=1}^{k} p_i^{n_i}\, f(n_1, n_2, \dots, n_k | p_1, p_2, \dots, p_k), \qquad (1)$$

for non-negative integers n_j such that $\sum_{j=1}^{k} n_j = n$.

To see this note that if n distinct balls are thrown one by one into k boxes, with probability p_i for landing in box i, then the number of ways in which n_1 balls fall in box 1, n_2 in box 2, $\dots$, n_k fall in box k is

$$\binom{n}{n_1}\binom{n - n_1}{n_2} \cdots \binom{n_k}{n_k} = \frac{n!}{\prod_{i=1}^{k} n_i!}$$

and each such way has probability $\prod_{i=1}^{k} p_i^{n_i}$.

The multinomial distribution reduces to the binomial distribution if $k = 2$. In Example 1, the number of cells is $k = 4$ and it is of interest to ‘test’ whether $(p_1, p_2, p_3, p_4) = (9/16, 3/16, 3/16, 1/16)$. On the other hand, in Example 2, one wants to see if all the 6 categories are equally likely, i.e., $p_j = 1/6$, $j = 1, 2, \dots, 6$. In Example 3, the three cell probabilities are, respectively, p^2, $(1 - p)^2$ and $2p(1 - p)$ which depend on a common parameter p.

Note that maximum likelihood estimation of the unknown probability vector $(p_1, p_2, \dots, p_k)$ is straightforward if these probabilities vary freely (subject, of course, to the constraint that they add up to 1). The likelihood function for the unknown parameter $p = (p_1, p_2, \ldots, p_k)$ from (1) above is

$$\ell(p) = \ell(p_1, p_2, \ldots, p_k) = \frac{n!}{\prod_{i=1}^{k} n_i!} \prod_{i=1}^{k} p_i^{n_i}.$$

Here one regards $\ell(p)$ as a function of the parameter p for given data $n = (n_1, n_2, \ldots, n_k)$.

The principle of maximum likelihood (enunciated by the great statistician R A Fisher, see [3]) says that for observed data n, choose that value of the parameter p that explains the data best, i.e., that maximises the likelihood $\ell(p)$. Since $\log(x)$ is a monotone increasing function on $(0, \infty)$, maximising $\ell(p)$ is equivalent to maximising $\log \ell(p)$. Now

$$\log \ell(p) = \log \ell(p_1, p_2, \ldots, p_k) = \text{constant} + \sum_{i=1}^{k} n_i \log(p_i).$$

Since p_i need to add up to 1, using a Lagrange multiplier the problem reduces to maximising

$$\sum_{i=1}^{k} n_i \log(p_i) + \lambda \left(\sum_{i=1}^{k} p_i - 1 \right).$$

Routine calculus involving setting the partial derivatives equal to zero and so on yields the maximum likelihood estimates to be

$$\hat{p}_j = \frac{n_j}{n}, \; j = 1, 2, \ldots, k,$$

Note that $\hat{p}_j$ is simply the observed relative frequency of outcome j.

As discussed in Delampady and Krishnan [1] for the binomial case, a Bayesian alternative to the maximum likelihood approach is possible in the multinomial case also. In the binomial case, there was only one parameter, i.e., θ. As a prior distribution for θ, the Beta (α, γ)

with probability density $c\theta^{\alpha-1}(1-\theta)^{\gamma-1}$; $\alpha, \gamma > 0$; c a normaising constant, was suggested. An interesting property of this prior distribution was also pointed out there. Since the likelihood function and the prior density have the same functional form in θ, upon applying the Bayes Theorem to compute the posterior density, the posterior distribution turns out to be another Beta distribution. Indeed, by Bayes Theorem the posterior density of θ given data x, namely $\pi(\theta|x)$ is proportional to $f(x|\theta)g(\theta)$, where $f(x|\theta)$ is the density of data x given θ and $g(\theta)$ is the prior density of the parameter θ. In the binomial context, $\pi(\theta|x)$ is proportional to $\theta^{\alpha+x-1}(1-\theta)^{\gamma+n-x-1}$. The parameter space there is the interval $[0, 1]$. The parameter space in the multinomial case is the simplex in k dimensions:

$$\left\{(p_1, p_2, \ldots, p_k) : p_i \geq 0, \sum_{i=1}^{k} p_i = 1\right\}. \qquad (2)$$

What is the appropriate generalization of the Beta prior now? It is called the Dirichlet (prior) distribution. A random vector $p = (p_1, p_2, \ldots, p_k)$ has the Dirichlet distribution with parameters $\alpha_1, \alpha_2, \ldots, \alpha_k$, each of which is positive, if the joint probability density function of p is

$$\pi(p_1, p_2, \ldots, p_k) = \frac{\Gamma\left(\sum_{i=1}^{k}\alpha_i\right)}{\prod_{i=1}^{k}\Gamma(\alpha_i)} \prod_{i=1}^{k} p_i^{\alpha_i-1}, \qquad (3)$$

for any $(p_1, p_2, \ldots, p_k)$ lying in the k-dimensional simplex (2). Here Γ is the complete Gamma function. Note now that exactly the same phenomenon as in the binomial case repeats here. In other words, combining (1) and (3) using the Bayes Theorem yields the posterior density of p given the data $n = (n_1, n_2, \ldots, n_k)$ as

$$\begin{aligned}&\pi(p_1, p_2, \ldots, p_k|n_1, n_2, \ldots, n_k)\\ &= \frac{f(n_1, n_2, \ldots, n_k|p_1, p_2, \ldots, p_k)\pi(p_1, p_2, \ldots, p_k)}{m(n_1, n_2, \ldots, n_k)}\end{aligned}$$

$$= c\prod_{i=1}^{k} p_i^{n_i+\alpha_i-1}; \qquad (4)$$

where c is a normalising constant and $m(n_1, n_2, \ldots, n_k)$ is the marginal probability of $n_1, n_2, \ldots, n_k$. Comparison with (3) yields

$$c = \frac{\prod_{i=1}^{k}\Gamma(n_i+\alpha_i)}{\Gamma(n+\sum_{i=1}^{k}\alpha_i)}.$$

To provide Bayesian estimates for the multinomial cell probabilities p_j, we could consider the maximum a posteriori estimate and the posterior mean. The computation involved in finding the former is the same as that in deriving the maximum likelihood estimate, and it is evident that this estimate for p_j is $(n_j+\alpha_j-1)/(n+\sum_{i=1}^{k}\alpha_i-k)$, $j = 1, 2, \ldots, k$. Finding the posterior mean is also very easy. It can be shown that this estimate for p_j turns out to be $(n_j+\alpha_j)/(n+\sum_{i=1}^{k}\alpha_i)$.

In some situations, such as the ones that frequently occur in genetics, the multinomial cell probabilities do not vary freely, but instead are functions of other unknown parameters θ. In such situations, neither the maximum likelihood estimation nor the Bayesian approach will be as simple, and MCMC methods will be found useful. This will be discussed later in Part 4. Next some questions related to inferences from MCMC samples will be addressed.

Sufficiency and Rao–Blackwell Theorem

Theory of statistics uses probability models to extract information from sample data. The first step in this direction is to identify data summaries which are relevant for inference and exclude those parts of data which do not contain any relevant information. For example, if we intend to estimate the average yield of mango per tree for a certain location using a random sample of trees from this location, the order in which the observations are collected is irrelevant, even though while recording the data this information may be included.

To make this concept precise, suppose one has a random sample of observations from a population with a certain probability distribution. Further, suppose that this probability distribution has probability density (or mass function) $f(x|\theta)$ where θ is the unknown parameter of interest. Any function of the sample is called a statistic. A statistic is sufficient for the parameter θ if the conditional distribution of the sample given the statistic does not involve the parameter θ. In other words, a sufficient statistic contains all the information in the sample which is relevant as far as inference on θ is concerned.

Example 4. Suppose $X_1, X_2, \ldots, X_k$ are i.i.d. Poisson random variables with mean θ. Here the sample mean $\bar{X} = \frac{1}{k}\sum_1^k X_i$ is sufficient for θ, i.e., it contains all the relevant information. To see this, first note that if X_1 and X_2 are independent Poisson random variables with means λ_1 and λ_2, then for any integer $n \geq 0$, $P(X_1 + X_2 = n) = \sum_{r=0}^{n} P(X_1 = r, X_2 = n - r) = \sum_{r=0}^{n} P(X_1 = r)P(X_2 = n - r)$ (by independence of X_1 and X_2)

$$= \sum_{r=0}^{n} e^{-\lambda_1}\frac{\lambda_1^{r}}{r!}e^{-\lambda_2}\frac{\lambda_2^{n-r}}{(n-r)!} = e^{-(\lambda_1+\lambda_2)}\frac{(\lambda_1+\lambda_2)^n}{n!}$$

(by the Binomial theorem). Thus $X_1 + X_2$ is Poisson $(\lambda_1 + \lambda_2)$ random variable. By induction, $T = \sum_{i=1}^{k} X_i$ is Poisson $(\lambda_1 + \lambda_2 + \ldots + \lambda_k)$. Therefore, for any sequence $(x_1, x_2, \ldots, x_k)$ of nonnegative integers and any nonnegative integer t, the conditional distribution of the data vector $(X_1, X_2, \ldots, X_k)$ given that $T = t$ satisfies:

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_{k-1} = x_{k-1}, X_k = x_k | T = t)$$

$$= \frac{P(X_1 = x_1; X_2 = x_2; :::; X_{k¡1} = x_{k¡1}; X_k = x_k; T = t)}{P(T = t)}$$

$$= \frac{P(X_1 = x_1; X_2 = x_2; :::; X_{k¡1} = x_{k¡1}; X_k = t ¡ x_k)}{P(T = t)}$$

$$= \frac{(\overset{k}{\underset{i=1}{Y}} \frac{e^{¡µ}µ^{x_i}}{x_i!})}{\frac{e^{¡kµ}(kµ)^t}{t!}}$$

$$= \frac{t!}{\overset{k}{\underset{1}{Y}} x_i!} \overset{k}{\underset{1}{Y}} (\frac{1}{k})^{x_i} \qquad (5)$$

if $\overset{k}{\underset{1}{X}} x_i = t$, 0 otherwise, yielding two results. First, the conditional distribution of $X_1; X_2; :::; X_k$ given their sum is the multinomial distribution having the probability mass function given in (5). Secondly, this conditional probability distribution does not involve µ, and hence T is sufficient for µ.

Example 5. Suppose $X_1; X_2; :::; X_n$ is a random sample from the Gaussian distribution with mean ¹ and variance $¾^2$, both of which are unknown. Then $µ = (¹; ¾^2)$ can be thought of as the parameter of interest. Intuition suggests that $T = (¹X; S^2)$, where $S^2 = P^n_{i=1}(X_i ¡ ¹X)^2$ must be sufficient for µ, as ¹X is the sample mean and $\frac{S^2}{n}$ is the sample variance. This is indeed true and for a proof consult Casella and Berger(pp.218-19)[4]. This implies that if the population is Gaussian, there is no need to retain any other part of the sample than ¹X and S^2.

One consequence of sufficiency is that it is enough to consider estimators of µ which are functions of a sufficient statistic. This can be made precise by the Rao{Blackwell Thorem [1].

Rao{Blackwell Theorem. Let $±(X_1; X_2; :::; X_n)$ be an estimator of µ with finite variance. Suppose that

[1] 'Rao' here is the famous C R Rao who has won many awards and distinctions for his contributions to statistical theory and methodology. He was with the Indian Statistical Institute for well over 40 years and was a teacher to many leading statisticians from India. David Blackwell is a well-known statistician from the University of California at Berkeley.

T is sufficient for θ, and let $\delta^*(T)$ defined by $\delta^*(t) = \mathbb{E}(\delta(X_1, X_2, \ldots, X_n)|T = t)$, be the conditional expectation of $\delta(X_1, X_2, \ldots, X_n)$ given $T = t$. Then

$$\mathbb{E}(\delta^*(T) - \theta)^2 \le \mathbb{E}(\delta(X_1, X_2, \ldots, X_n) - \theta)^2.$$

The inequality is strict unless $\delta = \delta^*$, or equivalently, δ is already a function of T.

Proof. By the property of iterated conditional expectation,

$$\mathbb{E}(\delta^*(T)) = \mathbb{E}\,[\mathbb{E}(\delta(X_1, X_2, \ldots, X_n)|T)] =$$

$$\mathbb{E}(\delta(X_1, X_2, \ldots, X_n)).$$

Therefore, to compare the mean square errors (MSE) of the two estimators, we need to compare their variances only. A standard result similar to the iterated conditional expectation (see Casella and Berger, pp.167-68)[4], says that

$$\begin{aligned} \mathrm{Var}(\delta(X_1, X_2, \ldots, X_n)) &= \mathrm{Var}\,[\mathbb{E}(\delta|T)] + \mathbb{E}\,[\mathrm{Var}(\delta|T)] \\ &= \mathrm{Var}(\delta^*) + \mathbb{E}\,[\mathrm{Var}(\delta|T)] \\ &> \mathrm{Var}(\delta^*); \end{aligned}$$

unless $\mathrm{Var}(\delta|T) = 0$, which is the case only if δ itself is a function of T.

What Rao–Blackwell theorem says is that any estimator can be improved upon by conditionally averaging over partitioning sets of the sample space where the value of the sufficient statistic T is kept fixed. In these sets the sample points do vary, but this variation has no relevance as far as θ is concerned. Note also that by the sufficiency of T, $\delta^*(T)$ is also a function of only the data and does not depend on θ. This method of improving an estimator $\delta(x)$ by taking its conditional expectation given t, i.e., using $\delta^*(T)$ is called Rao–Blackwellisation in the statistics literature.

Consider the following implication of Rao–Blackwell Theorem. In the context of Example 5, is there any reason why one should not choose the first observation X_1 as an estimate for μ, instead of using all the observations in some way? There certainly is. That $\mathbf{E}(X_1|\bar{X}) = \bar{X}$ is superior to X_1 follows from Rao–Blackwell Theorem. (Note that $\mathbf{E}(X_1|\bar{X}) = \mathbf{E}(X_i|\bar{X})$, for $i = 2, 3, \ldots, n$ since X_i are all identically distributed, and hence $\mathbf{E}(X_1|\bar{X}) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{E}(X_i|\bar{X}) = \mathbf{E}(\frac{1}{n}\sum_{i=1}^{n}X_i|\bar{X}) = \bar{X}$.)

The Rao–Blackwell Theorem involves two key steps: variance reduction by conditioning and conditioning by a sufficient statistic. The first step is based on the analysis of variance formula that says: For any two random variables S and T, the variance of S equals the sum of the variance of the conditional expectation of S given T and the expectation of the conditional variance S given T written, as noted earlier, as

$$\mathrm{Var}(S) = \mathrm{Var}(\mathbf{E}(S|T)) + \mathbf{E}(\mathrm{Var}(S|T)).$$

Thus one can reduce the variance of a random variable S by taking conditional expectation given some auxiliary information T. The second step exploits the fact that the conditional expectation of any statistics S is a function only of the data, i.e., it does not depend on any underlying parameter, given a sufficient statistic.

Let us see how Rao–Blackwellisation is useful in MCMC estimation. Consider the example where we generated random samples from the bivariate normal using Gibbs Sampling. Suppose we have a sample of n from the joint distribution of (X, Y) produced in this manner. Using this, how could we estimate such quantities as the mean μ_x of X or the marginal density $f(x)$ of X? Let us consider μ_x first. One would think that $\bar{x} = (1/n)\sum_{i=1}^{n}x_i$ is the best estimator here. However, we have some more information available here from $y_1, y_2, \ldots, y_n$ also. This can be seen from the fact that

$$\mu_x = \mathbf{E}(X) = \mathbf{E}\,[\mathbf{E}(X|Y)],$$

and the RHS of the above equation can be estimated by

$$(1/n)\sum_{i=1}^{n}\mathbb{E}(X|y_i) = \rho(1/n)\sum_{i=1}^{n} y_i;$$

since we know the form of the conditional expectation of X given $Y = y$. To show that this alternative estimator is superior to $\bar{x}$, we use the proof of Rao–Blackwell Theorem. As noted there,

$$\mathrm{Var}(X) = \mathrm{Var}[\mathbb{E}(X|Y)] + \mathbb{E}[\mathrm{Var}(X|Y)] \geq \mathrm{Var}[\mathbb{E}(X|Y)];$$

so that $\bar{x}$ has a larger variance than the new estimator $(1/n)\sum_{i=1}^{n}\mathbb{E}(X|y_i)$. We can use this improved Rao–Blackwellised estimator only in situations where we know the exact functional form of $\mathbb{E}(X|Y)$, as in this example. The same logic gives us an improved estimator for the marginal density $f(x)$ using $f(x|y_i)$. Since we know the form of the conditional density $f(x|y)$, we can use the estimator

$$\hat{f}(x) = (1/n)\sum_{i=1}^{n} f(x|y_i);$$

which in our case becomes

$$\hat{f}(x) = (1/n)\sum_{i=1}^{n} \phi(x; \rho y_i; 1 - \rho^2):$$

The application of the first step in the MCMC context is explained now:

Let $(X_j; Y_j); j = 1, 2, \ldots, N$ be the data generated by a single run of the Gibbs sampler algorithm (see Part 2) with a target distribution of a bivariate random vector $(X; Y)$. Let $h(X)$ be a function of the X component of $(X; Y)$ and let its mean value be μ. Suppose the goal is to estimate μ. A first estimate is the sample mean of the $h(X_j); j = 1, 2, \ldots, N$. From the MCMC theory, it can be shown that as $N \to \infty$, this estimate will converge to

μ in probability. The computation of variance of this estimator is not easy due to the (Markovian) dependence of the sequence $\{X_j; j = 1, 2, \ldots, N\}$. Now suppose we make n independent runs of Gibbs sampler and generate $(X_{ij}, Y_{ij}); j = 1, 2, \ldots, N; i = 1, 2, \ldots, n$. Now suppose that N is sufficiently large so that (X_{iN}, Y_{iN}) can be regarded as a sample from the limiting target distribution of the Gibbs sampling scheme. Thus $(X_{iN}, Y_{iN}); i = 1, 2, \ldots, n$ are i.i.d. and hence form a random sample from the target distribution. Then one can offer a second estimate of μ – the sample mean of $h(X_{iN}); i = 1, 2, \ldots, n$. This estimator ignores a good part of the MCMC data but has the advantage that the variables $h(X_{iN}); i = 1, 2, \ldots, n$ are independent and hence the variance of their mean is of order $\frac{1}{n}$. Now using the variance reduction idea outlined above and using the auxiliary information $Y_{iN}; i = 1, 2, \ldots, n$, one can improve this estimator as follows:

Let $k(y) = \mathbb{E}((h(X)|Y = y)$. Then for each i, $k(Y_{iN})$ has a smaller variance than $h(X_{iN})$ and hence the following third estimator, the sample mean of $k(Y_{iN}); i = 1, 2, \ldots, n$ has a smaller variance than the second one.

This is illustrated above for the Gaussian case where two special choices of the function $h(.)$ are considered: (1) $h(x) = x$; and (2) $h(x)$: the pdf of X evaluated at x. In Part 2 we illustrated the estimation of the marginal pdf of X by Gibbs sampler with 1000 independent runs each of length 1000. In Part 4, we provide another – a more realistic – example of the use of Gibbs sampling for Bayesian inference in the multinomial case with Dirichlet priors.

Suggested Reading

[1] Mohan Delampady and T Krishnan, Bayesian Statistics, *Resonance*, Vol.7 No.4, pp. 27-38, 2002.

[2] G Snedecor and W Cochran, *Statistical Methods*, 6th Edition, Iowa State University Press, Ames, Iowa, USA, 1967.

[3] T Krishnan, Fisher's contribution to Statistics, *Resonance*, Vol.2, No.9, pp.32-36, 1997. [Also reproduced in M Delampady, T Krishnan and S Ramasubramanian (Editors), *Echoes from Resonance: Probability and Statistics*, Bangalore, Indian Academy of Sciences Hyderabad: Universities Press, Ch. 24; pp.185-190, 2001]

[4] G Casella and R L Berger, *Statistical Inference*, Second Edition, Pacific Grove, CA, Duxbury, 2001.

Address for Correspondence
K B Athreya
School of ORIE
Rhodes Hall
Cornell University, Ithaca
New York 14853, USA.

Mohan Delampady
Indian Statistical Institute
8th Mile, Mysore Road
Bangalore 560 059, India.

T Krishnan
Systat Software Asia-Pacific Ltd.
Floor 5, 'C' Tower
Golden Enclave, Airport Road
Bangalore 560 017, India.