Sankhyā : The Indian Journal of Statistics
2004, Volume 66, Part 2, pp 234-242

A Note on the Problem of Heaps

Antonio Lijoi and Igor Prünster
Università degli Studi di Pavia, Italia

Abstract

The model for the so-called "heaps" problem as set in Kingman (1975) is considered and an explicit expression for evaluating the expectation of the mean search time of a demanded item in equilibrium is provided. Particular attention is devoted to the γ-stable case and Kingman's results are recovered in the limit.

AMS (2000) *subject classification.* Primary 60G57, 60G51; secondary 60G52.
Keywords and phrases. Heaps problem, Lévy process, normalized random measure, random discrete distribution, stable subordinator.

1 Introduction

Models that apply to the so-called "heaps" problem were introduced and studied in Hendricks (1972), Burville and Kingman (1973), Burville (1974) and Kingman (1975). The problem of heaps can be described as follows. Suppose there is a finite collection of items, $I_1, \ldots, I_N$, which are literally or figuratively stored in a heap. Such items might be, for instance, books on a shelf, pieces of information stored in a computer or papers on a desk. Each item is assigned a measure of popularity which coincides with the probability the item is demanded by a potential customer or user. Let p_i denote the popularity of the i-th item, so that $p_i \geq 0$ for each i and $\sum_{i=1}^{N} p_i = 1$. From time to time an item is requested: it is searched through the heap starting from the top. Then, after being used, it is returned to the top of the heap. Moreover, successive requests are independent and the popularities are allowed to be random.

The present paper draws inspiration from Kingman (1975), where the p_i's are defined by means of subordinators, *i.e.* Lévy processes with non negative increments. More recently, Donnelly (1991) has given a stimulating treatment of the problem by obtaining an alternative derivation of Kingman's limiting results together with some extensions that can be applied to the Poisson-Dirichlet case.

Letting μ denote the mean search time for a typical item demanded, in Burville and Kingman (1973) it is shown that, in statistical equilibrium,

$$\mu = \sum_{i \neq j} \frac{p_i \, p_j}{p_i + p_j} \, .$$

In Kingman (1975) it is first supposed that the distribution of the random vector $(p_1, \ldots, p_N)$ is a symmetric $(N-1)$-variate Dirichlet distribution whose density function, on the simplex $\Delta_{N-1} := \{(p_1, \ldots, p_{N-1}) : p_i \geq 0, \sum_{i=1}^{N-1} p_i \leq 1\}$, is given by

$$\frac{\Gamma(N\alpha)}{[\Gamma(\alpha)]^N} \, [p_1 \cdots p_{N-1}(1 - p_1 - \cdots - p_{N-1})]^{\alpha - 1}$$

for some $\alpha > 0$. With this position, one easily computes the expected mean search time $\mathrm{E}(\mu)$ and, by taking the limits $N \to +\infty$ and $\alpha \to 0$ in such a way that $N\alpha \to \lambda > 0$, Kingman (1975) proves that $\mathrm{E}(\mu) \to \lambda$. In this case, the ranked vector of the first n random probabilities converges in distribution to the n-th marginal of the so-called Poisson-Dirichlet distribution.

A second example one might consider arises when the p_i's are defined by normalizing a γ-stable subordinator. Unlike the previous case, the joint distribution of the random vector $(p_1, \ldots, p_N)$ is not generally available, for any $N \geq 1$. Indeed, in Kingman (1975), just a limiting form of $\mathrm{E}(\mu)$, as $N \to +\infty$, is evaluated, whereas no expression is obtained for $\mathrm{E}(\mu)$, with N finite. The latter issue would still be open for any subordinator used in the place of the γ-stable one. Here, we fill in this gap by providing an expression for $\mathrm{E}(\mu)$ which is valid for any subordinator one might use for defining the p_i's. In particular, in the γ-stable case, the expression for $\mathrm{E}(\mu)$ turns out to be quite simple and one can easily recover the limiting result stated by Kingman (1975).

The paper is structured as follows. In Section 2, normalized random measures with independent increments are employed in order to assign the random probabilities p_i. An expression for $\mathrm{E}(\mu)$ is then provided and it is explicitly evaluated in the stable case. The result is then compared with the known limiting form. In Section 3 some numerical examples are carried out for illustrative purposes.

2 Main Result

Let $\xi = \{\xi_t : t \geq 0\}$ be a subordinator and let ν be the corresponding Lévy measure. For an exhaustive account on the theory of subordinators, the reader can refer to, e.g., Bertoin (1996) and Sato (1999). If $\nu(\mathbb{R}^+) = +\infty$, ξ is known to be an *infinite activity* process implying ξ to be (almost surely) positive. Introduce a finite and non null measure α on $\mathbb{R}$ according to which the time change $t = A(x) = \alpha((-\infty, x])$ is carried out. This operation yields the reparametrized process $\xi^A = \{\xi_{A(x)} : x \in \mathbb{R}\}$, which has still independent, but generally not stationary, increments. Moreover, by virtue of the Lévy-Khintchine representation one has

$$\mathrm{E}\left[\mathrm{e}^{-u\,\xi_{A(x)}}\right] = \mathrm{e}^{-A(x)\,\psi(u)}$$

where $\psi(u) = \int_{(0,+\infty)} (1 - \mathrm{e}^{-uv})\,\nu(\mathrm{d}v)$ is also designated as the *characteristic exponent* of ξ. Since ξ^A is (almost surely) finite, we are in a position to consider

$$\tilde{F}(x) = \frac{\xi_{A(x)}}{\xi_a} \qquad \text{for every } x \in \mathbb{R}$$

as a random probability distribution function on $\mathbb{R}$, where $a := \alpha(\mathbb{R})$. The corresponding random probability measure is denoted by $\tilde{P}$ and it will be referred to as *normalized random measure with independent increments*. Such measures represent a subclass of a general family of random probability measures introduced and studied in a Bayesian nonparametric setting by Regazzini et al. (2003).

Take $J_1, \ldots, J_N$ to be a partition of the support of α and set the popularities of the N different items as $p_i := \xi_{\alpha_i}/\xi_a$ where $\alpha_i = \alpha(J_i)$, for $i = 1, \ldots, N$. For notational simplicity, set $\psi^{(k)}(u+v) = (\partial^k/\partial u^k)\psi(u+v)$ and introduce the following quantity

$$\mathcal{I}(\alpha_i, \alpha_j) = a \int_{(0,+\infty)} \mathrm{e}^{-a\psi(u)} \int_{(0,+\infty)} \psi^{(2)}(u+v)\mathrm{e}^{-(\alpha_i+\alpha_j)(\psi(u+v)-\psi(u))}\mathrm{d}v\,\mathrm{d}u.$$

A simple formula for evaluating $\mathrm{E}(\mu)$ is now provided.

Proposition 1. *Let the popularities in the "heap" problem be derived from a normalized random measure with independent increments. Then*

$$\mathrm{E}(\mu) = \sum_{i \neq j} \frac{\bar{\alpha}_i \bar{\alpha}_j}{\bar{\alpha}_i + \bar{\alpha}_j} \left[1 + \mathcal{I}(\alpha_i, \alpha_j)\right],$$

where $\bar{\alpha}_i = \alpha_i/a$, for any $i = 1, \ldots, N$.

Proof. By the definition of the p_i's one has

$$\mathrm{E}(\mu) = \sum_{i \neq j} \mathrm{E}\left[\frac{\xi_{\alpha_i}\xi_{\alpha_j}}{\xi_a\ (\xi_{\alpha_i} + \xi_{\alpha_j})}\right] = \sum_{i \neq j} S_{i,j}.$$

Each summand, $S_{i,j}$, is then computed by exploiting the independence of the increments of ξ^A as follows

$$S_{i,j} = \int_{(0,\infty)^2} \mathrm{E}\left[\xi_{\alpha_i}\xi_{\alpha_j}\mathrm{e}^{-u\xi_a - v(\xi_{\alpha_i}+\xi_{\alpha_j})}\right] \,\mathrm{d}u\,\mathrm{d}v$$

$$= \int_{(0,\infty)^2} \mathrm{E}\left[\mathrm{e}^{-u(\xi_a-\xi_{\alpha_i}-\xi_{\alpha_j})}\right]\mathrm{E}\left[\xi_{\alpha_i}\mathrm{e}^{-(u+v)\xi_{\alpha_i}}\right]\mathrm{E}\left[\xi_{\alpha_j}\mathrm{e}^{-(u+v)\xi_{\alpha_j}}\right]\mathrm{d}u\,\mathrm{d}v$$

$$= \int_{(0,\infty)^2} \mathrm{e}^{-(a-\alpha_i-\alpha_j)\psi(u)}\,\mathrm{E}\left[-\frac{\partial}{\partial u}\,\mathrm{e}^{-(u+v)\xi_{\alpha_i}}\right]\mathrm{E}\left[-\frac{\partial}{\partial u}\,\mathrm{e}^{-(u+v)\xi_{\alpha_j}}\right]\mathrm{d}u\,\mathrm{d}v$$

If one interchanges derivatives with the integrals above, one has

$$S_{i,j} = \alpha_i\alpha_j \int_{(0,+\infty)} \mathrm{e}^{-(a-\alpha_i-\alpha_j)\psi(u)} \int_{(0,+\infty)} [\psi^{(1)}(u+v)]^2 \mathrm{e}^{-(\alpha_i+\alpha_j)\psi(u+v)}\mathrm{d}v\,\mathrm{d}u$$

$$= \frac{\alpha_i\alpha_j}{\alpha_i+\alpha_j}\int_{(0,+\infty)} \mathrm{e}^{-(a-\alpha_i-\alpha_j)\psi(u)}\left[\psi^{(1)}(u)\mathrm{e}^{-(\alpha_i+\alpha_j)\psi(u)} + \int_{(0,+\infty)} \psi^{(2)}(u+v)\mathrm{e}^{-(\alpha_i+\alpha_j)\psi(u+v)}\mathrm{d}v\right]\mathrm{d}u\ .$$

Since $\nu(\mathbb{R}^+) = +\infty$, it is straightforward to show that

$$\int_{(0,+\infty)} \psi^{(1)}(u)\ \mathrm{e}^{-a\psi(u)}\mathrm{d}u = \frac{1}{a}\ ,$$

thus, leading to write

$$S_{i,j} = \frac{\alpha_i\alpha_j}{\alpha_i+\alpha_j}\left[\frac{1}{a} + \int_{(0,+\infty)} \mathrm{e}^{-a\psi(u)} \int_{(0,+\infty)} \psi^{(2)}(u+v)\mathrm{e}^{-(\alpha_i+\alpha_j)(\psi(u+v)-\psi(u))}\mathrm{d}v\mathrm{d}u\right]$$

$$= \frac{\bar{\alpha}_i\bar{\alpha}_j}{\bar{\alpha}_i+\bar{\alpha}_j}\left[1+\mathcal{I}(\alpha_i,\alpha_j)\right]. \qquad \square$$

The merit of such a formula relies upon the fact that it resembles both symmetric and asymmetric case, as in Kingman's terminology, and this holds true for any possible choice of the subordinator ξ. Moreover, note that, since $\mathcal{I}(\alpha_i,\alpha_j) < 0$, one has $\mathrm{E}(\mu) < \sum_{i\neq j}\bar{\alpha}_i\bar{\alpha}_j/(\bar{\alpha}_i+\bar{\alpha}_j)$. The weights $\bar{\alpha}_i$ admit an interesting interpretation that can be deduced, e.g., from Pitman (2003). Indeed, $\bar{\alpha}_i = E(p_i)$ for any i, thus mimicking what happens for the Dirichlet process.

A notable example we focus on is strictly related to a limiting result proved in Kingman (1975). We consider the γ-stable subordinator having Lévy measure

$$\nu(\mathrm{d}v) = v^{-\gamma-1}\,\mathrm{d}v$$

for any γ in $(0,1)$. In this case, $\mathcal{I}(\alpha_i, \alpha_j)$ can be determined explicitly.

PROPOSITION 2. *If ξ is a γ-stable subordinator, then*

$$\mathrm{E}(\mu) = \sum_{i \neq j} \frac{\bar{\alpha}_i \bar{\alpha}_j}{\bar{\alpha}_i + \bar{\alpha}_j} \left[1 - (1-\gamma)\, {}_2F_1\left(1, 1; 1 + \frac{1}{\gamma}; 1 - \bar{\alpha}_i - \bar{\alpha}_j\right)\right]$$

where ${}_2F_1(a, b; c; x)$ is the Gauss hypergeometric function. Moreover, if $\bar{\alpha}_i = 1/N$, for each $i = 1, \ldots, N$, then

$$\lim_{N \to +\infty} \mathrm{E}(\mu) = \begin{cases} \gamma/(1-2\gamma) & \gamma < 1/2 \\ \infty & \gamma \geq 1/2 \end{cases}$$

PROOF. If ξ is the γ-stable subordinator, then $\psi(u) = \Gamma(1-\gamma)u^\gamma/\gamma$ and, by resorting to the transformations $(u+v)^\gamma = x$ and $u^\gamma = y$, one has

$$\begin{aligned} &\mathcal{I}(\alpha_i \alpha_j) \\ &= -a\frac{\Gamma(2-\gamma)}{\gamma^2} \int_{(0,+\infty)} \mathrm{e}^{-(a-\alpha_i-\alpha_j)\frac{\Gamma(1-\gamma)}{\gamma}y}\, y^{\frac{1}{\gamma}-1} \int_{(y,+\infty)} x^{-\frac{1}{\gamma}} \mathrm{e}^{-(\alpha_i+\alpha_j)\frac{\Gamma(1-\gamma)}{\gamma}x}\,\mathrm{d}x\,\mathrm{d}y. \end{aligned}$$

Application of (3.381.6) and of (7.621.3) in Gradshteyn and Ryzhik (2000) and some algebra lead to the following expression for $\mathcal{I}(\alpha_i, \alpha_j)$

$$-(1-\gamma)\, {}_2F_1\left(1, 1; 1 + \frac{1}{\gamma}; 1 - \bar{\alpha}_i - \bar{\alpha}_j\right).$$

Finally, with reference to Kingman's limiting result, it can be easily achieved by virtue of the asymptotic properties of the Gauss hypergeometric functions. Indeed in this case, *i.e.* $\alpha_i = 1/N$ for $i = 1, \ldots, N$,

$$\mathrm{E}(\mu) = \frac{N-1}{2}\left[1 - (1-\gamma)\, {}_2F_1\left(1, 1; 1 + \frac{1}{\gamma}; 1 - \frac{2}{N}\right)\right]$$

and the so-called duplication formula for the Gauss hypergeometric function (see, *e.g.*, Corollary 2.3.3 in Andrews et al., 1999) yields

$$\begin{aligned} {}_2F_1\left(1,1;1+\frac{1}{\gamma};1-\frac{2}{N}\right) &= \frac{\Gamma\left(1+\frac{1}{\gamma}\right)\Gamma\left(\frac{1}{\gamma}-1\right)}{\Gamma\left(\frac{1}{\gamma}\right)^2}\, {}_2F_1\left(1,1;2-\frac{1}{\gamma};\frac{2}{N}\right) \\ &\quad + \left(\frac{2}{N}\right)^{\frac{1}{\gamma}-1} \Gamma\left(1+\frac{1}{\gamma}\right)\Gamma\left(1-\frac{1}{\gamma}\right) {}_2F_1\left(\frac{1}{\gamma},\frac{1}{\gamma};\frac{1}{\gamma};\frac{2}{N}\right). \end{aligned}$$

From the previous expression it can be deduced that if $\gamma < 1/2$,

$$ {}_2F_1\left(1,1;1+\frac{1}{\gamma};1-\frac{2}{N}\right) = \frac{1}{1-\gamma}\left(1-\frac{\gamma}{1-2\gamma}\frac{2}{N}\right) + o\left(\frac{1}{N}\right) $$

as $N \to +\infty$. Hence, one has $\mathrm{E}(\mu) \to \gamma/(1-2\gamma)$. On the other hand, if $\gamma \geq 1/2$, then

$$ {}_2F_1\left(1,1;1+\frac{1}{\gamma};1-\frac{2}{N}\right) = \frac{1}{1-\gamma}+\left(\frac{2}{N}\right)^{\frac{1}{\gamma}-1}\Gamma\left(1+\frac{1}{\gamma}\right)\Gamma\left(1-\frac{1}{\gamma}\right)+o\left(\frac{1}{N^{\frac{1}{\gamma}-1}}\right) $$

as $N \to +\infty$ and $\mathrm{E}(\mu)$ diverges. □

The result contained in Proposition 2 has the advantage of providing a simple expression for $\mathrm{E}(\mu)$. Indeed, it can be easily evaluated since many mathematical packages allow for computation of hypergeometric functions. Moreover, with reference to the particular symmetric case, it provides some hint on the rate at which $\mathrm{E}(\mu)$ converges, as N tends to $+\infty$, to the limit $\gamma/(1-2\gamma)$.

As a final remark for this section, it is to be observed that results analogous to the one contained in Proposition 2 can be achieved for any other normalized random measure used in the place of the γ-stable. The key point in the procedure relies on the computation of $\mathcal{I}(\alpha_i, \alpha_j)$, for any i, j. This can be carried out either analytically or numerically.

3 Numerical Illustration

For illustrative purposes, here we provide a comprehensive treatment of the "heaps" problem investigated in Kingman (1975). In the framework of Proposition 2, with $\bar{\alpha}_i = 1/N$ for any $i = 1, \ldots, N$, $\mathrm{E}(\mu)$ is computed for different values of $\gamma \in (0,1)$ and of the number of items N in the heap. The exact values of $\mathrm{E}(\mu)$ are then compared with the limit one obtains as the number of items N goes to $+\infty$.

TABLE 1. EXPECTATIONS OF μ CORRESPONDING TO DIFFERENT VALUES OF N AND OF THE PARAMETER OF THE STABLE SUBORDINATOR

	$N=10$	$N=50$	$N=100$	$N=200$	$N=+\infty$
$\gamma=0.2$	0.257	0.315	0.324	0.328	0.333
$\gamma=0.4$	0.758	1.278	1.451	1.590	2
$\gamma=0.6$	1.615	4.292	6.061	8.327	$+\infty$
$\gamma=0.8$	2.868	11.438	19.854	34.048	$+\infty$

Note that the closer γ to 0.5 and the slower is the convergence of $\mathrm{E}(\mu)$ to its limiting value: this situation is apparent when comparing behaviours at $\gamma = 0.2$ and at $\gamma = 0.4$.

Finally, a description of the distribution of μ is provided by exploiting a numerical algorithm for simulating Lévy processes as set forth by Wolpert and Ickstadt (1998). When applied to the γ-stable subordinator, such an algorithm, which is also known as the *inverse Lévy measure algorithm*, has just one free parameter coinciding with the number of jumps of the trajectory of the process one simulates. A sensible criterion for tuning it may be based on matching the empirical mean of μ with the exact value $\mathrm{E}(\mu)$, given above. This has led us to fix such a number of jumps equal to 500. For each of the estimates illustrated in the following figures, 1000 trajectories of the stable subordinator have been simulated. Below one is provided with kernel density estimates of the distributions of μ corresponding to different values of γ, with $N = 50$.

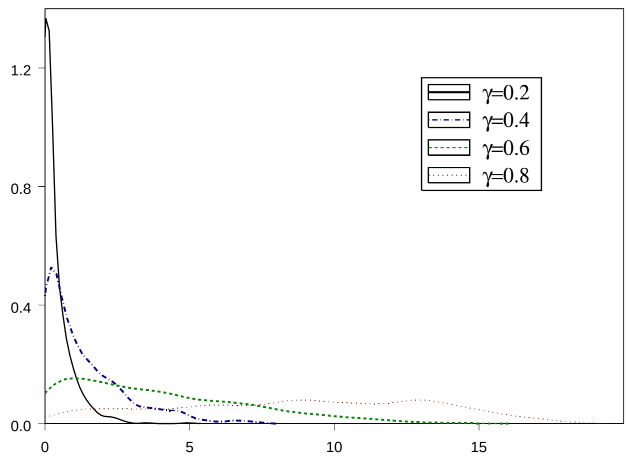

FIGURE 1. KERNEL DENSITY ESTIMATES OF THE DISTRIBUTION OF μ, WITH $N = 50$.

As far as the asymptotic behaviour is concerned, kernel density estimates are depicted in the figures below for $\gamma = 0.4,\ 0.6$ as N varies. Both kernel density estimates are such that as N increases the mass in the tails tends to increase. This phenomenon is much more apparent with $\gamma = 0.6$, in which case the estimate becomes flatter thus being consistent with divergence of the expected value of μ.

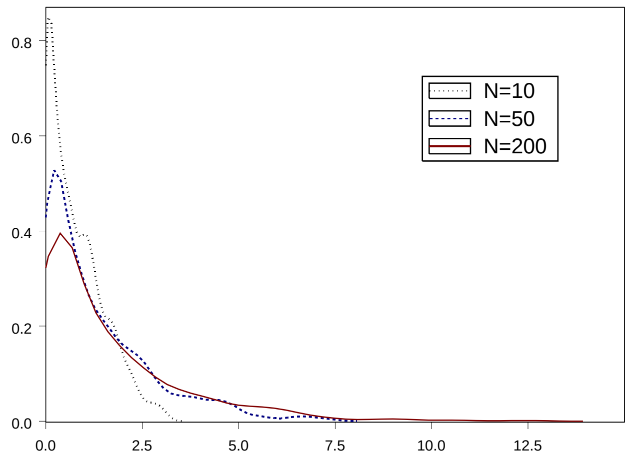

FIGURE 2. KERNEL DENSITY ESTIMATES OF THE DISTRIBUTION OF μ FOR $\gamma = 0.4$ WITH INCREASING N.

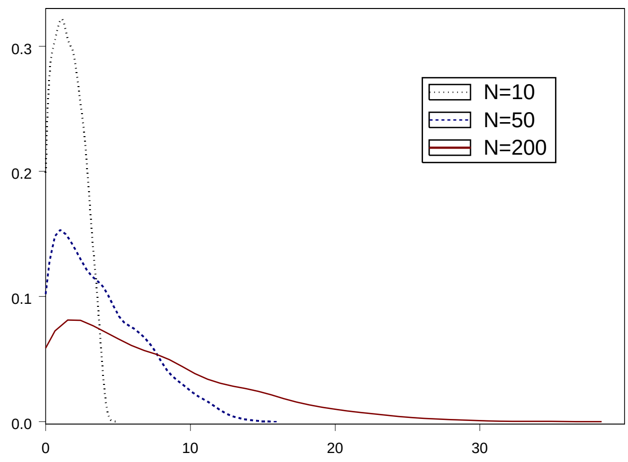

FIGURE 3. KERNEL DENSITY ESTIMATES OF THE DISTRIBUTION OF μ FOR $\gamma = 0.6$ WITH INCREASING N.

Acknowledgments. The authors were partially supported by the Italian Ministry of University and Research (MIUR), research project "Bayesian nonparametric methods and their applications".

References

ANDREWS, G.E., ASKEY, R. and ROY, R. (1999). *Special Functions.* Cambridge University Press, Cambridge.

BERTOIN, J. (1996). *Lévy processes.* Cambridge University Press, Cambridge.

BURVILLE, P.J. (1974). HEAPS: a concept in optimization. *J. Inst. Math. Appl.* **13**, 263-278.

BURVILLE, P.J. and KINGMAN, J.F.C. (1973). On a model for storage and search. *J. Appl. Probab.* **10**, 697-701.

DONNELLY, P. (1991). The heaps process, libraries, and size-biased permutations. *J. Appl. Probab.* **28**, 321-335.

GRADSHTEYN, I.S. and RYZHIK, J.M. (2000). *Table of Integrals, Series, and Products.* 6th. Ed. Academic Press, New York.

HENDRICKS, W.J. (1972). The stationary distribution of an interesting Markov chain. *J. Appl. Probab.* **9**, 231-233.

KINGMAN, J.F.C. (1975). Random discrete distributions. *J. Roy. Statist. Soc. Ser. B*, **37**, 1-22.

PITMAN, J. (2003). Poisson-Kingman partitions. In *Science and Statistics: A Festschrift for Terry Speed*, D.R. Goldstein, ed., Lecture Notes and Monograph Series **40**, 1-35. Institute of Mathematical Statistics, Hayward.

REGAZZINI, E., LIJOI, A. and PRÜNSTER, I. (2003). Distributional results for means of random measures with independent increments. *Ann. Statist.* **31**, 560-585.

SATO, K. (1999). *Lévy Processes and Infinitely Divisible Distributions.* Cambridge University Press, Cambridge.

WOLPERT, R. and ICKSTADT, K. (1998). Poisson/gamma random field models for spatial statistics. *Biometrika* **85**, 251-267.

ANTONIO LIJOI AND IGOR PRÜNSTER
DIPARTIMENTO DI ECONOMIA POLITICA E METODI QUANTITATIVI
UNIVERSITA' DEGLI STUDI DI PAVIA
VIA SAN FELICE, 5, 27100 PAVIA, ITALY
E-mail: lijoi@unipv.it
igor.pruenster@unipv.it

Paper received: March 2003; revised February 2004.