PROBABILISTIC ANALYSIS OF A GENERALIZED BIN PACKING PROBLEM AND APPLICATIONS

AWI FEDERGRUEN and GARRETT VAN RYZIN
Columbia University, New York, New York

(Received September 1992; revisions received December 1992, April 1994, January 1995, June 1995; accepted July 1995)

We give a unified probabilistic analysis for a general class of bin packing problems by directly analyzing corresponding mathematical programs. In this general class of packing problems, objects are described by a given number of attribute values. (Some attributes may be discrete; others may be continuous.) Bins are sets of objects, and the collection of feasible bins is merely required to satisfy some general consistency properties. We characterize the asymptotic optimal value as the value of an easily specified linear program whose size is independent of the number of objects to be packed, or as the limit of a sequence of such linear program values. We also provide bounds for the rate of convergence of the average cost to its asymptotic value. The analysis suggests an (a.s.) asymptotically ϵ-optimal heuristic that runs in linear time. The heuristics can be designed to be asymptotically optimal while still running in polynomial time. We also show that in several important cases, the algorithm has both polynomially fast convergence and polynomial running time. This heuristic consists of solving a linear program and rounding its solution up to the nearest integer vector. We show how our results can be used to analyze a general vehicle routing model with capacity and time window constraints.

Many discrete planning problems can be formulated as, or are closely related to, bin packing problems in which a set of objects, defined by a finite number of attributes, needs to be packed into a minimal number of feasible bins. See Coffman et al. (1984, 1988) for surveys. In the simplest version, n objects with sizes $\{w_1, \ldots, w_n\}$, $0 < w_i \leq 1$, $i = 1, \ldots, n$ need to be packed in a minimum number of bins of unit size. Even this simplest version is NP-complete, see Karp (1972). Hence, attempts have been made to provide probabilistic analyses of the solution value of heuristic algorithms as well as the optimal solution value itself. Using this probabilistic approach, Rhee and Talagrand (1987) have established remarkably general results that characterize the convergence of the minimum cost value for a large class of combinatorial problems.

However, the actual heuristics that one is able to analyze probabilistically are typically less general, often requiring stylized problem formulations and restrictive probabilistic assumptions, see, e.g., Frederickson (1980), Knodel (1981), Lueker (1982), Karp (1982), and the recent book of Coffman and Lueker (1991). These heuristics typically exploit geometrical properties of the model or symmetry properties of the underlying distributions. It would thus appear that analytical tractability is achieved at the expense of model realism.

In contrast, the mathematical programming approaches used in practice employ a unified set of tools that can incorporate a diversity of easily adaptable constraints and do not require up-front knowledge of any specific statistical patterns among the model parameters. However, it has often been difficult to give rigorous performance guarantees for suboptimal solution approaches to math programming formulations and instead numerical testing was relied on for assessing solution quality.

Our objective is to demonstrate that these characteristics can be combined; that is, we propose heuristics based on general and versatile math programming approaches, which are computationally efficient and have provable performance guarantees under very general probabilistic assumptions. Our approach is to discretize attributes to reduce the dimensionality of the original problem. A feasible solution for the resulting reduced integer program is constructed by solving its linear programming relaxation and rounding the solution up to the nearest integer vector. Probabilistic analysis is then used to characterize the quality of the resulting solution.

Elements of this approach have been used before. Coffman and Lueker (1991, §2.6) suggest discretization and analysis of bin packing problems as integer programs as an interesting analysis technique. Karmarkar and Karp (1982) describe an efficient ϵ-approximation scheme for one-dimensional bin packing problems based on "grouping" objects into one of a small number of types. Similarly, Courcoubetis and Weber (1986) use this discrete model in analyzing a bin packing system in which each type of object arrives according to a renewal process. Our contribution is to use probabilistic analysis to characterize the solution quality and complexity of combined discretization and integer programming approaches for a wide class of packing problems.

Specifically, we obtain the following results: We characterize the asymptotic optimal value as the value of an easily specified linear program whose size is independent of the number of objects, or as the limit of a sequence of such linear program values. For discrete distributions of the attribute values, we obtain the exact limiting distribution of the minimum cost value. We also obtain upper bounds

Subject classifications: Analysis of algorithms: probabilistic. Transportation: vehicle routing. Production/scheduling: bin packing/cutting stock.
Area of review: Optimization.

0030-364X/97/4504-0596 $05.00

for the rate of convergence of the average cost to its asymptotic value. Our heuristic can be designed to be (a.s.) ϵ-optimal while running in linear time, or (a.s.) asymptotically optimal while still running in polynomial time. In important special cases (e.g., classical bin packing, vector packing and a packing problem arising from a time window VRP), the heuristic has both low complexity and polynomially fast convergence.

Rhee and Talagrand (1989a and 1989b) showed that the asymptotic optimal solution value of a one-dimensional bin packing problem is characterized by the expected value of a transformation, $u(\cdot)$, of the object sizes which is *dual feasible*, i.e., any bin which is feasible under the original object sizes remains feasible under the transformed sizes. We establish the existence of a dual feasible transformation function for our general d-dimensional problem using duality properties of the underlying linear programs.

We also apply all the above results to a general vehicle routing model with capacity and time window constraints. A planning period of T units of time is available during which all on-site service must start and end. A customer is characterized by four attributes (demand, service time, earliest and latest delivery time) in addition to its location. This model generalizes all previous vehicle routing models for which probabilistic analyses have been given. See Solomon (1987) and Solomon and Desrosiers (1988) for recent survey articles and Kolen et al. (1987) for a recent attractive branch-and-bound method. Bramel et al. (1993) recently provided a probabilistic analysis for a more restricted version of the problem in which the planning period of T time units is divided into P equally long intervals, such that each customer's earliest and latest delivery times are given by the end points of one of these P intervals, the service times are uniformly distributed on the interval $(0, T/p)$ and no capacity considerations prevail. Bramel and Simchi-Levi (1997) independently address the general model we analyze, characterize its asymptotic solution value by analyzing a randomized algorithm, and develop an exponential time, asymptotically optimal heuristic.

The plan of the paper is as follows: In Section 1 we specify the Generalized Bin Packing Problem (GBPP), discuss examples and review known results for this class of models. In Section 2 we analyze the discrete GBPP in which the distribution of the attribute values is discrete. In Section 3 we address continuous attribute distributions. Our approach is to approximate the continuous distribution by a sequence of progressively finer discretizations, thus allowing us to employ the integer programming analysis of Section 2. Our results on the GBPP are used, in Section 4, to analyze the above general routing model with capacity and time window constraints. In Section 5 we briefly discuss efficient implementations using column generation techniques. These can be used in conjunction with the ellipsoid method (Grötschel et al. 1981, 1988) to show that the complexity of the entire heuristic is often a low order polynomial, while maintaining asymptotic optimality.

1. THE GENERALIZED BIN PACKING PROBLEM (GBPP)

1.1. Problem Definition and Examples

We define a generalized version of the classical bin packing problem as follows. Objects are described by a d-dimensional vector $w = (w^1, \ldots, w^d)$ of bounded attributes, without loss of generality (w.l.o.g.), shifted and scaled so that $w^j \in [0, 1], j = 1, \ldots, d$. Let Ω denote the collection of all *feasible bins*, i.e., sets of objects B that can be packed using one bin. We assume that this collection satisfies the following properties:

P1. *Every set of objects can be packed in some collection of feasible bins.*

P2. *If $B \in \Omega$, then for every $w \in B$, $B - \{w\} \in \Omega$.*

P3. *If the vector of attributes w is restricted to a finite set, then the collection of feasible bins in Ω is also finite.*

Property P1 assures the existence of a feasible solution. Property P2 says that removing an object from a feasible bin always results in another feasible bin. P3 excludes *null* objects, i.e., objects that feasibly can be added to any feasible bin (e.g., a zero weight object in the classical bin packing problem), since an infinite number of distinct feasible bins arises by adding an arbitrary number of null objects to any given feasible bin. P3 is w.l.o.g. since all null objects can be eliminated from a problem instance and added arbitrarily to the solution of the remaining problem.

For problems in which the attributes take on a continuum of values, we will need one additional property:

P4. *Let $B \in \Omega$ be a feasible bin, $w \in [0, 1]^d$ be any object in B, and $0 \leq \alpha \leq 1$ and j be a continuous attribute. Let $\hat{w}$ denote the vector defined by $\hat{w}^j = \alpha w^j$ and $\hat{w}^i = w^i, i \neq j$. Then, $B - \{w\} + \{\hat{w}\} \in \Omega$.*

Property P4 says that replacing an object in a feasible bin by one with smaller attribute values preserves feasibility. (P4 is related to the definition of a *lower set* in Rhee and Talagrand 1991.) For physical attributes (e.g., size, weight, service time) P4 usually holds; however, this can depend on the choice of variables used to represent the problem (see, e.g., Section 4). In cases where some of the attributes are discrete and others are continuous, P4 needs to hold for the latter only.

For probabilistic analyses, we shall assume instances are formed by drawing objects independently with attribute vectors distributed according to a probability measure μ having bounded support. We emphasize that μ is a *joint* measure on the attributes; specifically, we shall not require independence among the various attributes.

Frequently, we need to examine the case where w is restricted to a finite set. In this case, any set of objects S can be represented by a vector $z = (z_1, \ldots, z_I) \in \mathfrak{X}^I_+$ where $i = 1, \ldots, I$ indexes the *type* of object (i.e., a type is a particular vector of discrete attribute values $(w_1, \ldots,$

w_d)) and z_i denotes the number of objects of type-i contained in the set S. Also in this case, μ is a probability mass function, denoted by $\pi = (\pi_1, \ldots, \pi_I)$.

The *generalized bin packing problem* (GBPP) is the problem of finding $b^*(S)$, the minimum number of feasible bins necessary to pack a given set of objects S. There are many applications of generalized bin packing problems that satisfy P1–P4:

Example 1. Vector Packing. Karp et al. (1984) (see also Rhee and Talagrand 1991) generalize the classical bin packing problem as follows. An object i is described by a d-vector of nonnegative weights a_{ij}, $j = 1, \ldots, d$. A bin B is feasible if it satisfies the inequalities $\Sigma_{i \in B} a_{ij} \leq 1$, $j = 1, \ldots, d$.

Example 2. Rectangular Packing. Object i is a rectangle described by two dimensions, (x_i, y_i). These rectangles are to be packed into a minimum number of unit squares so that they do not overlap and are contained by the squares. Generalizations include packing objects in a variety of shapes or into d-dimensional unit cubes (Karp et al. 1984).

Example 3. Automobile Hauling. A collection of automobiles has to be distributed from assembly plants to individual dealer locations by a minimum number of trailers. Automobiles have a size index, small (1), mid-size (2) and full-size (3). Trailers for hauling the automobiles can be configured in one of a finite number of ways by adjusting various platforms. Each configuration specifies which size automobile can fit in any particular slot on the trailer. The geometry of the platforms is complex, and thus feasibility is determined by more than a simple total size or total weight constraint. In general, one has to enumerate the various platform settings to generate the set of feasible configurations.

Example 4. A Lower Bound for a Simplified VRP with Time Windows. Consider a version of the time window model of Bramel et al. (1993). Items represent locations where deliveries on a given day need to be made. A service time is associated with each location and the day is divided into P disjoint time intervals of unit length. Items are thus described by a weight w_k, a service time s_k and an index p_k of the time period $p = 1, \ldots, P$ during which the location must be served. A feasible bin B is one that satisfies $\Sigma_{i \in S: p_i = p}\, s_i \leq 1$, $p = 1, \ldots, P$ and $\Sigma_{i \in S} w_i \leq 1$.

The number of vehicles needed to service any given set of customers is no less than the optimal value of this GBPP. P4 is satisfied since only the weight and service time are continuous. See Section 4 for a more general version of the VRP with time windows.

1.2. Known Results for the GBPP

Many probabilistic analysis results from classical bin packing can be generalized to the GBPP. In particular, let $S^{(n)}$ be a set formed by taking n i.i.d. objects distributed according to μ. Since the GBPP is subadditive, Kingman's (1976) theory of subadditive processes implies

$$\lim_{n \to \infty} \frac{b^*(S^{(n)})}{n} = \gamma \quad \text{(a.s.)}, \tag{1}$$

where the constant γ depends on μ and $1/\gamma$ is interpreted as the average number of objects per bin in the optimal GBPP solution. Since $0 \leq b^*/n \leq 1$, by the bounded convergence theorem (Royden 1968) we also have

$$\lim_{n \to \infty} \frac{E[b^*(S^{(n)})]}{n} = \gamma, \tag{2}$$

as well. These results show that the number of bins used in the optimal solution is highly predictable for large problem instances.

The optimal solution to the GBPP has the *1-conservative* property (Coffman and Lueker 1991), i.e., adding an object to a set cannot increase the optimal number of bins by more than one. (Note P1 and repeated application of P2 imply that every single item bin is feasible.) Rhee and Talagrand (1987) showed, using Azuma's Lemma for Martingale difference sequences, that if b^* is 1-conservative, then

$$P\{|b^*(S^{(n)}) - E[b^*(S^{(n)})]| > t\} \leq 2e^{-2t^2/n}. \tag{3}$$

(The above result includes sharper constants due to McDiarmid 1989.) Combining (2) and (3), they showed the following theorem:

Theorem 1 (Rhee and Talagrand 1987). *For every $\epsilon > 0$, $\alpha > 0$, there exists a constant $n_0 = n_0(\epsilon, \alpha)$ such that,*

$$P\left\{\left|\frac{b^*(S^{(n)})}{n} - \gamma\right| > \epsilon\right\} \leq 2e^{-2n\epsilon^2/(1+\alpha)},$$

for all $n \geq n_0$.

Note that Theorem 1 implies the complete convergence of b^*/n to γ by the Borel-Cantelli Lemma and thus is a stronger result than (1).

However, these results give few insights into heuristic algorithms. The constant γ also remains unknown. As shown below, we can use discrete approximation and the theory of linear programming to address both of these issues.

2. THE DISCRETE GBPP

In the discrete GBPP, each attribute can take on only a finite number of values. Thus, there are a finite number of possible vectors w or object *types*, indexed by i, $i = 1, \ldots, I$. We can represent any set S by a vector $z \in \mathcal{Z}_+^I$ where z_i denotes the number of type i objects in S. By Property P3, let j, $j = 1, \ldots, J$ index the finitely many feasible bins and let a_{ij} denote the number of type i objects in a type j bin. Define the matrix $A = [a_{ij}]$, and let A^j denote its jth column. Though I and J can be quite large, this problem is still polynomial in the number of objects $n = \Sigma_i z_i$ for fixed

I and J, which one can show using the dynamic program $b^*(z) = \min\{b^*(z - A^j) + 1\}$. The GBPP is equivalent to:

$$b^*(z) = \min \sum_{j=1}^{J} x_j \quad \text{s.t.} \quad Ax \geq z,\ x \geq 0, \quad x \text{ integer}. \tag{4}$$

We will frequently use the linear programming relaxation of (4),

$$b^{LP}(z) = \min \sum_{j=1}^{J} x_j, \quad \text{s.t.} \tag{5}$$

$$Ax \geq z,\ x \geq 0. \tag{6}$$

The next lemma summarizes some useful properties about b^* and b^{LP}. We provide a brief proof; parts (d) and (e) are proved in greater generality by Blair and Jeroslow (1982, Proposition 2.17 and Corollary 4.7), showing that b^* is a subadditive, Gomory function. Throughout the paper, $|x|$ denotes the l_1-norm of the vector x.

Lemma 1. (a) *$b^*(z)$ and $b^{LP}(z)$ are, component-wise, nondecreasing in z.*

(b) *$b^{LP}(z)$ is convex in z.*

(c) *$b^{LP}(\alpha z) = \alpha b^{LP}(z)$ for all $z \geq 0$ and $\alpha \geq 0$.*

(d) *$b^{LP}(z) - |z_0| \leq b^{LP}(z + z_0) \leq b^{LP}(z) + |z_0|$ and $b^*(z) - |z_0| \leq b^*(z + z_0) \leq b^*(z) + |z_0|$ for all nonnegative z, z_0.*

(e) *$b^{LP}(z) \leq b^*(z) \leq b^{LP}(z) + I$.*

Proof. Part (a) follows from (4) and (5) since $A \geq 0$; (b) is a standard result for right-hand-side parametric programming, while (c) follows easily from (5) by considering the change of variables $x_j' = \alpha x_j$. We get (d) from the 1-conservative property. The first inequality in (e) is obvious, while the second one is obtained by noting that if x is the optimal solution to (5), then $\{y^j = \lceil x^j \rceil, j = 1, \ldots, J\}$ is a feasible solution for (4). Since there are at most I nonzero (basic) variables in (5), and these are increased by at most one, the result follows. □

2.1. Almost Sure Convergence

Lemma 1 suggests a simple proof of (1) which also characterizes the constant γ. Recall $\pi = (\pi_1, \ldots, \pi_I)$ denotes the probability measure on the set of possible item types. Let $S^{(n)}$ denote a set of n i.i.d. objects distributed according to π and $z^{(n)}$ denote its vector representation. Then,

Theorem 2.

$$\lim_{n\to\infty} \frac{b^*(z^{(n)})}{n} = b^{LP}(\pi) \quad (a.s.).$$

Proof. For the stochastic sequence $z^{(n)}$ we have by the strong law of large numbers that $z^{(n)}/n \to \pi$ (a.s.). Thus, by the continuity of $b^{LP}(\cdot)$, $b^{LP}(z^{(n)}/n) \to b^{LP}(\pi)$ (a.s.). By Lemma 1, we have

$$\frac{1}{n} b^{LP}(z^{(n)}) \leq \frac{1}{n} b^*(z^{(n)}) \leq \frac{1}{n} b^{LP}(z^{(n)}) + \frac{I}{n},$$

so $\lim_{n\to\infty} 1/n\ b^*(z^{(n)}) = \lim_{n\to\infty} 1/n\ b^{LP}(z^{(n)}) = b^{LP}(\pi)$ (a.s.). □

Lemma 1(a) implies that solving the linear program (5) and rounding its solution up is an asymptotically optimal heuristic. Moreover, this linear program takes constant time under the uniform model of computation since its dimensions are fixed (and is $O(\log(n))$ on a machine with finite word length since the magnitude of the numbers $z^{(n)}$ increases linearly with n (a.s.)).

Note that the sequence of attribute vectors need only satisfy the strong law of large numbers and therefore the assumption that the sequence is i.i.d. can be weakened (Revesz 1968). Finally, one can obtain a tail probability result analogous to Theorem 1 using a direct analysis of the mathematical program (4) as well; see Federgruen and van Ryzin (1994).

2.2. The Asymptotic Distribution of b^*

In the discrete case, it is possible to characterize the full asymptotic distribution of $b^*(z^{(n)})$ and hence find bounds for its moments. Let V denote the $I \times I$ matrix with $V_{ii} = \pi_i(1 - \pi_i)$ and $V_{ij} = -\pi_i\pi_j$ for $i \neq j$.

Theorem 3. *Assume the linear program $b^{LP}(\pi)$ has a unique optimal dual vector y. Then $(b^*(z^{(n)}) - ny^T\pi)/\sqrt{ny^TVy}$ converges in distribution to a standard normal random variable, i.e.,*

$$\lim_{n\to\infty} P\left\{\frac{b^*(z^{(n)}) - ny^T\pi}{\sqrt{ny^TVy}} \leq x\right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\xi^2/2}\, d\xi, \tag{7}$$

$$ny^T\pi \leq E[b^*(z^{(n)})] \leq ny^T\pi + O(1), \tag{8}$$

and

$$|\text{var}[b^*(z^{(n)})] - ny^TVy| = O(1). \tag{9}$$

Moreover, (7) continues to hold with $b^*(z^{(n)})$ replaced by $b^H(z^{(n)})$, the value of the heuristic solution obtained by rounding up the solution of the linear programming relaxation (5).

Proof. Since the linear program $b^{LP}(\pi)$ has a unique, optimal dual solution y, there exists a non-empty polyhedron P in $\Re^I$ with π an interior point of P, such that $b^{LP}(\pi) = y^T\pi$ for all $\pi \in P$. Since $z^{(n)}/n \to \pi$ (a.s.), with probability one there exists an integer n_0 such that for all $n \geq n_0$, $z^{(n)}/n \in P$; hence,

$$b^{LP}(z^{(n)}) = nb^{LP}(z^{(n)}/n) = ny^Tz^{(n)}/n = y^Tz^{(n)} \quad \forall n \geq n_0. \tag{10}$$

Let u^k denote the unit I-vector indicating the identity of object k, i.e., $u_i^k = 1$ if object k is a type i object and $u_i^k = 0$ otherwise. By assumption, we have that the vectors $\{u^1, \ldots, u^n\}$ are i.i.d. with mean π and variance-covariance matrix V. Let $N(0, V)$ denote a multivariate normal with mean 0 and variance-covariance matrix V. It then follows from the multivariate central limit theorem (Theorem 30, Chapter 3 of Pollard 1984) that $(z^{(n)} - n\pi)/\sqrt{n} \to N(0, V)$ in distribution. This implies that

$$\frac{y^T z^{(n)} - ny^T \pi}{\sqrt{ny^T V y}} \to N(0, 1). \tag{11}$$

Note from Lemma 1 that

$$b^{LP}(z^{(n)}) \le b^*(z^{(n)}) \le b^{LP}(z^{(n)}) + I. \tag{12}$$

The results of the theorem then follow from (10), (11), and (12). □

Remark. The requirement that y be unique is essential for the analysis above. It is well known that $b^{LP}(\cdot)$ is a piece-wise linear function. As demonstrated in the above proof, $z^{(n)}/n$ is almost surely confined to the interior of *one* of the linear domains for all but finitely many n, provided the linear program $b^{LP}(\pi)$ has a unique dual solution. On the other hand, if the dual solution to $b^{LP}(\pi)$ is not unique, then infinitely many values of $z^{(n)}/n$ may be contained in two or more linear domains. If so, $b^{LP}(z^{(n)}/n)$ and $b^*(z^{(n)})$ would converge to a piece-wise linear transformation of a multivariate normal random variable, a distribution which is difficult to characterize.

3. THE CONTINUOUS GBPP

We next consider the case where the attribute vector $w \in [0, 1]^d$ is continuous, Property P4 holds and the probability measure μ has a continuous density almost everywhere. The case where μ consists of a mixture of continuous and discrete distributions is handled by combining the results for the purely discrete and continuous GBPP.

Our approach is to discretize the unit d-cube using a grid and to round the attributes to its points. We then analyze a sequence of discretized problems letting the grid size tend to zero.

3.1. Discretization Based on the Cubic Histogram

To form a cubic histogram, we divide each coordinate interval [0, 1] into I intervals, each of length $h = 1/I$, for I a nonnegative integer. This produces I^d subcubes with sides of length h. Henceforth, interpret $h \to 0$ to mean $I \to \infty$. This scheme induces a set of grid points $\{0, 1, \ldots, I\}^d$. Index these points by i, with $w(i)$ the ith grid point, $i = 1, \ldots, I'$, where $I' = (I + 1)^d = O(1/h^d)$.

Let $\bar{\mathcal{A}}_i(\underline{\mathcal{A}}_i)$ denote the subcube of all points which are transformed into grid point i when their components are rounded up (down) to the nearest multiple of h. That is,

$$\bar{\mathcal{A}}_i = \{w = (w^1, \ldots, w^d): h(\lceil w^1/h \rceil, \ldots, \lceil w^d/h \rceil) = w(i)\},$$

and

$$\underline{\mathcal{A}}_i = \{w = (w^1, \ldots, w^d): h(\lfloor w^1/h \rfloor, \ldots, \lfloor w^d/h \rfloor) = w(i)\}.$$

We extend this definition to include subcubes that are not subsets of $[0, 1]^d$ (see Figure 1).

We define two probability measures, $\bar{\pi}$ and $\underline{\pi}$, on the I' grid points by $\bar{\pi}_i = \mu(\bar{\mathcal{A}}_i)$ and $\underline{\pi}_i = \mu(\underline{\mathcal{A}}_i)$, $i = 1, \ldots, I'$. We write $\underline{\pi}(h)$ and $\bar{\pi}(h)$ when we want to make the dependence on h explicit. Note that this scheme defines a discretized version of the continuous GBPP which, by Property P3, has a finite set of feasible bins in Ω and can thus be solved using the integer program (4).

We will need a lemma relating these two probability distributions. First, define a density to be *Lipschitz continuous of order s* on a set $\mathcal{A} \subseteq \mathfrak{R}^d$ if

$$|f(x) - f(y)| \le C|x - y|^s \quad \forall x, y \in \mathcal{A},$$

where $f(w) = 0$ for $w \notin [0, 1]^d$. For example, any triangular density on [0, 1] is Lipschitz continuous of order one everywhere; the uniform density is Lipschitz continuous of all orders on the interior of the unit cube but not on $\mathfrak{R}^d$ since it is discontinuous at the boundary of the unit cube.

Lemma 2. (a) *If the probability measure μ has a density f that is bounded and continuous almost everywhere, then*

$$|\bar{\pi}(h) - \underline{\pi}(h)| \to 0 \quad \text{as } h \to 0.$$

(b) *If f is Lipschitz continuous of order one on the interior of* $[0, 1]^d$, *then*

$$|\bar{\pi}(h) - \underline{\pi}(h)| = O(h).$$

(c) *If f is Lipschitz continuous of order s everywhere on $\mathfrak{R}^d$ then*

$$|\bar{\pi}(h) - \underline{\pi}(h)| = O(h^s).$$

Proof. If f is continuous everywhere, the mean value theorem implies $\underline{\pi}_i = h^d f(\underline{x}_i)$ and $\bar{\pi}_i = h^d f(\bar{x}_i)$ for some $\underline{x}_i \in \underline{\mathcal{A}}_i$ and some $\bar{x}_i \in \bar{\mathcal{A}}_i$. Thus, $|\underline{\pi}_i - \bar{\pi}_i| = h^d|f(\bar{x}_i) - f(\underline{x}_i)|$. Since continuity of f implies uniform continuity over the bounded region $[-h, 1 + h]^d$ and $|\bar{x}_i - \underline{x}_i| \le dh$, for every $\epsilon > 0$ we can select a $\delta > 0$ such that $|f(\bar{x}_i) - f(\underline{x}_i)| \le \epsilon/(1 + h)^d$ for all i if $h < \delta$. Thus, for $h < \delta$, $|\underline{\pi} - \bar{\pi}| = \sum_{i=1}^{I'} |\underline{\pi}_i - \bar{\pi}_i| \le \epsilon h^d(1/h + 1)^d/(1 + h)^d = \epsilon$. Thus, (a) follows if f is continuous everywhere.

Suppose now that f has discontinuities on a set $D \subset [0, 1]^d$ with Lebesgue measure zero. Define $\mathscr{C}_1$ to be the set of indices i for which either $\underline{\mathcal{A}}_i$ or $\bar{\mathcal{A}}_i$ intersect $\mathcal{D}$ and $\mathscr{C}_2 = \{1, \ldots, I'\} - \mathscr{C}_1$. Then we can write

$$|\underline{\pi} - \bar{\pi}| = \sum_{i \in \mathscr{C}_1} |\underline{\pi}_i - \bar{\pi}_i| + \sum_{i \in \mathscr{C}_2} |\underline{\pi}_i - \bar{\pi}_i|.$$

Since f is continuous over both $\underline{\mathcal{A}}_i$ and $\bar{\mathcal{A}}_i$ for $i \in \mathscr{C}_2$, by the same argument as above we can select an h such that the second sum above is no more than $\epsilon/2$. Further, since f is bounded, $\sum_{i \in \mathscr{C}_1} |\underline{\pi}_i - \bar{\pi}_i| \le Mh^d|\mathscr{C}_1|$ for some finite M. But the fact that $\mathcal{D}$ has measure zero implies that for sufficiently small h, $h^d|\mathscr{C}_1|$ can be made arbitrarily small since this quantity is the area of a set circumscribing $\mathcal{D}$ in a partition of mesh size $h/\sqrt{d}$ (see Buck 1978, Theorem 2, p. 172). Thus, h can be chosen small enough so that $Mh^d|\mathscr{C}_1| \le \epsilon/2$ as well, which together with the bound on the second sum above, implies $|\underline{\pi} - \bar{\pi}| \le \epsilon$. This completes the proof of part (a).

To prove part (b), note that if f is Lipschitz continuous of order one on the interior of $[0, 1]^d$, then for all points i

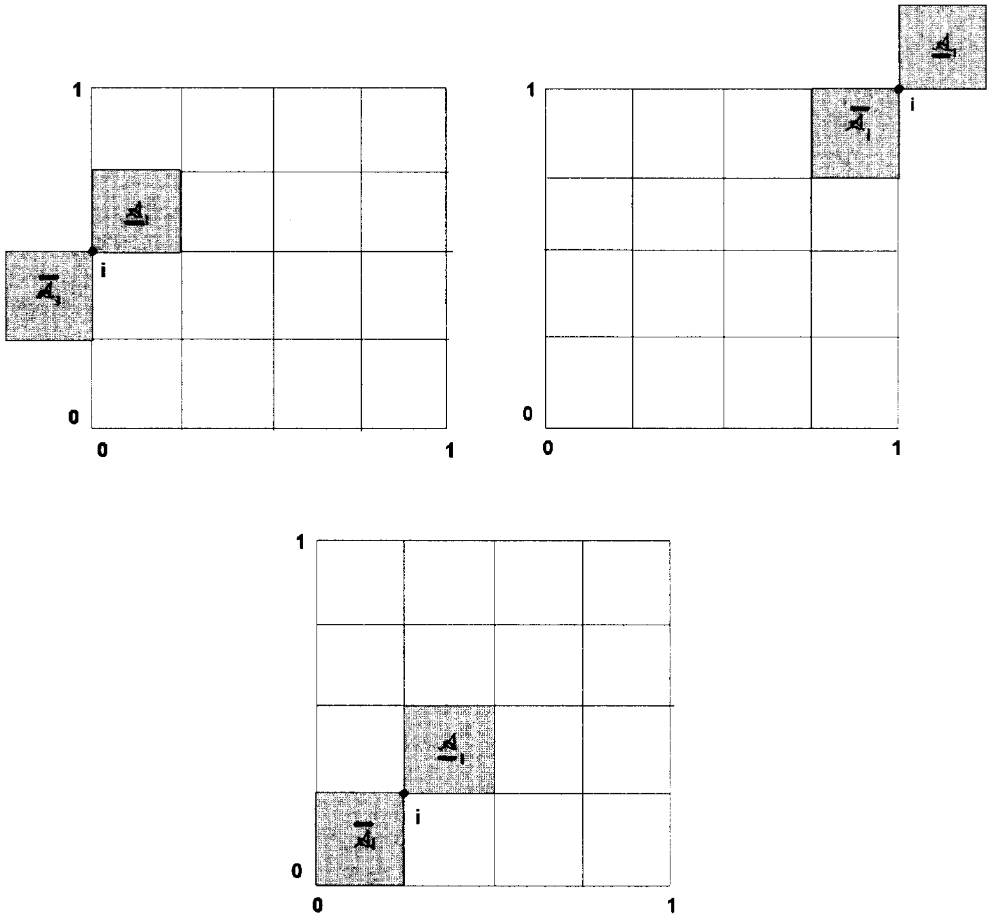

Figure 1. Examples of the regions $\underline{\mathcal{A}}_i$ and $\bar{\mathcal{A}}_i$.

with both $\underline{\mathcal{A}}_i$ and $\bar{\mathcal{A}}_i$ subsets of $[0, 1]^d$, we have as before $|\underline{\pi}_i - \bar{\pi}_i| = h^d |f(\bar{x}_i) - f(\underline{x}_i)| \le Ch^{d+1}$. Since there are at most $I' = O(1/h^d)$ interior points, the total contribution from these points to $|\underline{\pi} - \bar{\pi}|$ is $O(h)$. For points on the boundary, the boundedness of f implies $|\underline{\pi}_i - \bar{\pi}_i| \le Mh^d$ and since there are $O(I^{d-1}) = O(1/h^{d-1})$ boundary points, these contribute at most $O(h)$ as well.

For part (c), note that if f is Lipschitz continuous of order s everywhere, then $|\underline{\pi}_i - \bar{\pi}_i| \le Ch^{d+s}$ for all i and the result follows as in the previous cases. □

Remark. The condition of part (b) is satisfied whenever the density f is continuously differentiable on the interior of the unit cube or even when f has discontinuities on the interior, provided they occur only on a finite number of sufficiently smooth surfaces (Buck 1978).

Remark. Part (c) shows that the rate of convergence is a function of the smoothness of the density f, paralleling the results in Devroye and Gyorfi (1985) for the convergence rate of density estimators.

3.2. Almost Sure Convergence

We next characterize the asymptotic value of the GBPP.

Theorem 4. *Let $S^{(n)}$ denote a set of n i.i.d. objects distributed according to μ. If μ has a density f that is bounded and continuous almost everywhere, then* $\lim_{n\to\infty} b^*(S^{(n)})/n = \gamma$, *a.s. where* $\gamma \equiv \lim_{h\to 0} b^{LP}(\underline{\pi}(h))$.

Proof. For a given $h > 0$ consider the pair of discretization schemes, with $\bar{\pi}(h)$ and $\underline{\pi}(h)$ the associated probability distributions on $\{i : i = 1, \ldots, I'\}$. For $i = 1, \ldots, I'$, define $\bar{z}_i^{(n)}$ ($\underline{z}_i^{(n)}$) as the number of objects in $S^{(n)}$ which are rounded up (down) to the ith grid point, and let $\bar{z}^{(n)} = (\bar{z}_i^{(n)})$ and $\underline{z}^{(n)} = (\underline{z}_i^{(n)})$. In the upper (lower) discretization scheme every object is replaced by one with larger (smaller) attribute values; thus, by P4 and Lemma 1, $b^*(\underline{z}^{(n)}) \le b^*(S^{(n)}) \le b^*(\bar{z}^{(n)}) \le b^*(\underline{z}^{(n)}) + |\underline{z}^{(n)} - \bar{z}^{(n)}|$. Dividing these inequalities by n, taking limits, and using Theorem 2 and the strong law of large numbers we obtain

$$b^{LP}(\underline{\pi}(h)) \le \liminf_{n\to\infty} \frac{b^*(S^{(n)})}{n}$$

$$\le \limsup_{n\to\infty} \frac{b^*(S^{(n)})}{n} \le b^{LP}(\underline{\pi}(h))$$

$$+ |\underline{\pi}(h) - \bar{\pi}(h)| \quad (a.s.). \tag{13}$$

We know from Lemma 2 that if f is bounded and continuous almost everywhere, then for every $\epsilon > 0$ we can find a $\delta_1 > 0$ such that for $h < \delta_1$, $|\underline{\pi}(h) - \bar{\pi}(h)| \leq \epsilon/2$. It therefore suffices to verify that $b^{LP}(\underline{\pi}(h))$ converges monotonically to a constant γ as $h \to 0$ so that we can find a δ_2 with $|b^{LP}(\underline{\pi}(h)) - \gamma| \leq \epsilon/2$ for all $h < \delta_2$ and then take $h = \min\{\delta_1, \delta_2\}$. To show this, note that for $h_1 < h_2$, Property P4 implies $b^*(\underline{z}^{(n)}(h_2)) \leq b^*(\underline{z}^{(n)}(h_1))$. Dividing by n, taking $n \to \infty$ and invoking Theorem 2 we obtain $b^{LP}(\underline{\pi}(h_2)) \leq b^{LP}(\underline{\pi}(h_1))$, i.e. $b^{LP}(\underline{\pi}(h))$ is nondecreasing in h and converging since $b^{LP}(\underline{\pi}(h)) \leq 1$. □

Observe from the proof that an asymptotically ϵ-optimal heuristic H can be designed by applying the upper discretization grid size h, solving the linear program and rounding, and, as in the discrete case, H runs in linear time. However, we can say more. Lemma 2 shows that the error term $|\underline{\pi}(h) - \bar{\pi}(h)|$ in (13) approaches zero as $h \to 0$. Moreover, one can show that $\underline{z}^{(n)}/n$ converges to $\underline{\pi}(h)$ as $h \to 0$ provided $nh^d \to \infty$. (See Devroye and Gyorfi 1985 for a proof in the context of convergence of histogram estimators.) Therefore, it follows that one can choose h to decrease as a function of n slowly enough so that the heuristic H is asymptotically optimal (a.s.). Of course, as h decreases the size of the linear program increases, so the running time of H will increase with n. The rate of increase depends on the problem structure (i.e., how many columns are introduced as h becomes smaller). However, we have the following general result:

Theorem 5. *Suppose μ has a density f that is bounded and continuous almost everywhere. Then the heuristic H is both polynomial in n and asymptotically optimal (a.s.) as long as $nh^d \to \infty$.*

Proof. Let M denote the maximum number of objects with the attribute vector 0 that can be feasibly packed in one bin. By Property P3, $M < +\infty$. By Property P4, M is also an upper bound for the number of objects in any feasible bin. Recall that $I' = O(h^{-d})$. The number of columns in the linear program, J, is bounded by the number of ways in which M objects can be spread over I' types, i.e. I'^M. Then using any of a number of interior point methods (see Goldfarb and Todd 1989) the dual of the linear program can be solved in $O(I'^4L)$, where L is the length of the input string, which is $O(I'^{M+1})$. Therefore, the complexity is $O(I'^{M+5}) = O(h^{-d(M+5)}) = o(n^{M+5})$ since $nh^d \to \infty$. □

Note, however, convergence to optimality could be slow and/or the complexity (though polynomial) could be high because M is usually large. In the next subsection, we obtain bounds on the rate of convergence. In Section 5, we use these convergence bounds together with a column generation approach to show that the heuristic H has both fast convergence and reasonable running times for some special cases of the GBPP. Lastly, we note that using Theorems 2 and 4 these results can extend to finite mixtures of discrete distributions and distributions with a density which is continuous almost everywhere.

3.3. Rates of Convergence

In this section, we derive upper bounds for the rate of convergence of $E[b^*(S^{(n)})]$ to its asymptotic limit $n\gamma$.

Theorem 6. (a) *If μ has a density f that is Lipschitz of order one on the interior of* $[0, 1]^d$, *then*

$$n\gamma \leq E[b^*(S^{(n)})] \leq n\gamma + O(n^{d/(d+1)}).$$

(b) *If μ has a density f that is Lipschitz continuous of order s everywhere, then*

$$n\gamma \leq E[b^*(S^{(n)})] \leq n\gamma + O(n^{\max(1/2, d/(d+s))}).$$

Proof. For the lower bounds, fix h and note that $b^*(S^{(n)}) \geq b^{LP}(\underline{z}^{(n)})$ and since $b^{LP}(\cdot)$ is convex, Jensen's inequality implies $E[b^*(\underline{z}^{(n)})] \geq b^{LP}(E[\underline{z}^{(n)}]) = nb^{LP}(\underline{\pi}(h))$. The result then follows by letting $h \to 0$.

To prove the upper bounds, we employ a modification of a construction originally proposed by Coffman and Lueker (1991). Consider the upper discretization corresponding with the grid size h, let F denote the c.d.f. associated with p.d.f. $\bar{\pi}$ and $\bar{F} = 1 - F$ its complement. Also, let

$$F_n(w) = \frac{1}{n}|\{W_k \in S^{(n)}: W_k^1 \leq w^1, \ldots, W_k^d \leq w^d\}|$$

$$= \frac{1}{n} \sum_{i: w(i) \leq w} \bar{z}_i^{(n)},$$

denote the associated empirical distribution of the set $S^{(n)}$. Define the Kolmogorov-Smirnov statistic, D_n, by $D_n = \max_{w \in [0,1]^d} |\bar{F}_n(w) - \bar{F}(w)|$. It is well known (see Kiefer 1961) that a constant K exists, independent of F (and hence h) such that $E[D_n] \leq K/\sqrt{n}$. (See also Serfling 1986, Massart 1986, and Talagrand 1994 for improved bounds for the tail behavior of D_n.)

Now, for each attribute j, remove the $\lceil nD_n \rceil$ largest objects with respect to that attribute. Put each of these items into its own bin and let $\tilde{S}$ denote the set of remaining objects and $\tilde{z}$ denote its vector representation. Note this procedure removes at most $d\lceil nD_n \rceil \leq dnD_n + d$ objects. Further, for all $w = (w^1, \ldots, w^d)$, $|\{W_k \in \tilde{S}: h\lceil W_k^1/h \rceil \leq w^1, \ldots, h\lceil W_k^d/h \rceil \leq w^d\}| \leq n\bar{F}(w)$. This implies by P4 and Lemma 1(a), that $b^{LP}(\tilde{z}) \leq b^{LP}(n\bar{\pi})$. Therefore we have, $b^{LP}(\bar{z}^{(n)}) \leq nb^{LP}(\bar{\pi}) + dnD_n + d$, and hence by Lemma 1(e)

$$b^*(\bar{z}^{(n)}) \leq nb^{LP}(\bar{\pi}) + dnD_n + d + I'. \tag{14}$$

Taking expectations and using Lemma 1(d) we obtain

$$E[b^*(S^{(n)})] \leq nb^{LP}(\underline{\pi}) + n|\underline{\pi} - \bar{\pi}| + E[dnD_n] + O(1/h^d).$$

Note that $b^{LP}(\underline{\pi}) \leq \gamma$ and $E[dnD_n] \leq K\sqrt{n}$ for every h. Also, by Lemma 2, if μ has a density that is Lipschitz of order one on the interior, $|\underline{\pi} - \bar{\pi}| = O(h)$. Combining these bounds we have $E[b^*(S^{(n)})] \leq n\gamma + O(nh) +$

$O(n^{1/2}) + O(1/h^d)$. Choosing $h = O(n^{-1/(d+1)})$ minimizes the right-hand side asymptotically and proves the first part of the theorem. Part (b) follows in a similar manner by noting from Lemma 2 that $|\underline{\pi} - \bar{\pi}| = O(h^s)$ and taking $h = O(n^{-1/(d+s)})$. □

The bounds in part (a) continue to apply even when the distribution μ does not possess a Lipschitz continuous density provided the attributes are *independently* distributed. Consider the following *equal-probability* discretization of a GBPP with d attributes, which is a generalization of a construction in Coffman and Lueker (1991): If the attributes are independent, $F(w) = \Pi_{j=1}^d F_j(w^j)$ where $F_j(w^j)$ denotes the marginal distribution of attribute j. Let $F_j^{-1}(x) = \min\{w : F_j(w) = x\}$ define the inverse of F_j. Consider now the discretization scheme defined by the $I' = (1 + I)^d$ (perhaps nondistinct) grid points $\Pi_{j=1}^d \{F_j^{-1}(ih) : i = 0, 1, \ldots, I\}$. We index these points by $i = 1, \ldots, I'$.

As we did for the cubic histogram, for a set of objects $S^{(n)}$ let $\bar{z}^{(n)}$ (resp. $\underline{z}^{(n)}$) denote the set of objects formed by rounding up (resp. down) all attributes to the nearest grid point and let $\bar{\pi}$ (resp. $\underline{\pi}$) denote the corresponding probability vectors. A nice property of this discretization scheme is that $|\bar{\pi} - \underline{\pi}| = O(h)$ for any product-form distribution F:

Lemma 3. *If* $F = F_1 F_2 \cdots F_d$, *then for the equal-probability discretization scheme defined above* $|\bar{\pi} - \underline{\pi}| = O(h)$.

Proof. Note that selecting a random object W from a distribution $F = F_1 F_2 \cdots F_d$ and rounding up its attributes to the nearest grid point is equivalent to taking an independent sequence of $U[0, 1]$ random variables, $\{U^j; j \geq 1\}$, and forming the vector $(F_1^{-1}(h\lceil U^1/h \rceil), \ldots, F_d^{-1}(h\lceil U^d/h \rceil))$; likewise, rounding down to the nearest grid point is equivalent to forming the vector $(F_1^{-1}(h\lfloor U^1/h \rfloor), \ldots, F_d^{-1}(h\lfloor U^d/h \rfloor))$. Thus, it follows that $\bar{\pi}_i = \underline{\pi}_i = h^d$ except perhaps for grid points i on the boundary (i.e., points with values $w^j = 0$ or $w^j = 1$ along some dimension j) for which $|\bar{\pi}_i - \underline{\pi}_i| = h^d$. Since there are only $O(I^{d-1}) = O(1/h^{d-1})$ of these boundary points, the lemma follows. □

Again, solving the integer program (4) with $z = \bar{z}$ gives an upper bound on b^*. Lemma 3 implies that Theorem 4 holds as well. Finally, using the fact that $|\bar{\pi} - \underline{\pi}| = O(h)$ and repeating the argument in Theorem 6, we obtain the following result:

Theorem 7. *If* $F = F_1 F_2 \cdots F_d$, *then* $n\gamma \leq E[b^*(S^{(n)})] \leq n\gamma + O(n^{d/(d+1)})$.

Remark. Note this equal probability scheme requires *explicit knowledge* of the distribution F and independence among the attributes, neither of which is needed for the cubic histogram scheme.

Remark. If F has a density f that is Lipschitz continuous of order $s > 1$ everywhere, then the bound in Theorem 6 is stronger than that of Theorem 7.

3.4. Dual Feasible Functions

Coffman (1982) introduced the notion of dual feasible functions for the classical bin packing problem. A dual feasible function is a function $u : [0, 1] \to [0, 1]$ satisfying

$$\sum_{i=1}^{k} x_i \leq 1 \Rightarrow \sum_{i=1}^{k} u(x_i) \leq 1, \tag{15}$$

for all finite sequences $\{x_1, \ldots, x_k\}$ of positive real numbers. If object weights are distributed like some nonnegative random variable X, one can easily show that

$$\gamma \geq E[u(X)], \tag{16}$$

for all dual feasible functions u, where γ is the optimal bin packing constant in (1). Rhee and Talagrand (1989a and 1989b) subsequently showed that when X is distributed according to an arbitrary probability measure μ, then there is a dual feasible function u for which

$$\gamma(\mu) = \int u(x)\mu\{dx\} = E[u(x)]. \tag{17}$$

One can think of u as a function that transforms sizes so that the resulting problem is perfectly packable. Thus, knowing u is equivalent to knowing the packing constant γ. The proof of this result requires a fairly sophisticated argument based on the Hahn-Banach theorem. A generalization to vector and rectangular packing is given in Rhee and Talagrand (1991), where the authors also show that

$$\gamma(\mu) = \sup\left\{ \int f(x)\mu\{dx\} \right\}, \tag{18}$$

with the supremum taken over all *continuous*, dual feasible functions f. In this subsection, we generalize the notion of dual feasible functions to the GBPP and give simple proofs of the corresponding results for the GBPP.

We say a function $u : [0, 1]^d \to [0, 1]$ is dual feasible for the GBPP if for any finite collection of objects $\{w_1, \ldots, w_k\}$,

$$\{w_1, \ldots, w_k\} \in \Omega \Rightarrow \sum_{i=1}^{k} u(w_i) \leq 1. \tag{19}$$

One can view u as a function that maps the GBPP into a classical bin packing problem. We will show that such a mapping exists into a classical bin packing problem that is perfectly packable, but first we need the following lemma due to Rhee (1993).

Lemma 4. *Consider a sequence* $\{f_n\}$ *of dual feasible functions. If* $\int f_n(w)\mu\{dw\}$ *is nondecreasing in* n *and* $\lim_{n\to\infty} \int f_n(w)\mu\{dw\} = \gamma(\mu)$, *then there exists a dual feasible function* f *such that* $\int f(w)\mu\{dw\} = \gamma(\mu)$.

Proof. Let $L_2(\mu)$ denote the space of all real-valued functions f on $[0, 1]^d$ such that $|f(w)|^2$ is μ-measurable. Endow this space with the norm $\|f\| = (\int |f(w)|^2 \mu\{dw\})^{1/2}$. Define the sets $C_n = \{f : f = \sum_{r=n}^{\infty} \alpha_r f_r, \sum_{r=n}^{\infty} \alpha_r = 1, \alpha_r \geq 0 \ \forall r\}$, i.e. C_n is the closed convex hull of the functions f_n, $f_{n+1}, \ldots$. It is easy to verify that every function f in C_n is dual feasible; The sets C_n are subsets of the unit sphere, which is compact in the weak topology on $L_2(\mu)$ (Alaoglu's Theorem, Royden 1968, p. 202). One can verify that the sets C_n are closed and therefore compact in the weak topology as well. Note that $\{C_n\}$ has the finite intersection property, i.e., any finite subfamily of the sequence has a nonempty intersection. Since C_1 is compact, this implies there exists a function $f \in \cap_{n=1}^{\infty} C_n$ (Royden 1968, Proposition 1, p. 158). Since $f \in C_n$, it is dual feasible and $f = \sum_{r=n}^{\infty} \alpha_r f_r$ for some vector α, which implies, by the monotone convergence theorem that for all n

$$\gamma(\mu) \geq \int f\mu\{dw\} = \sum_{r=n}^{\infty} \alpha_r \int f_r \mu\{dw\} \geq \int f_n \mu\{dw\}.$$

The lemma now follows by letting $n \to \infty$. □

Theorem 8. *If* (i) μ *has a density f that is bounded and continuous almost everywhere, or* (ii) μ *has a finite distribution, or* (iii) μ *is a finite mixture of measures of the types in* (i) *and* (ii), *then a dual feasible function* $u : [0, 1]^d \to [0, 1]$ *exists with* $\int u(w)\mu\{dw\} = \gamma(\mu)$.

Proof. We first consider the problem in which μ is a probability mass function on a finite set of points, whose values are denoted $w(1), \ldots, w(I)$. Let y_i denote the optimal dual variables of (5) with $z = \pi$. By dual feasibility, these satisfy, with $A = [a_{ij}]$ (see (5)) $\sum_{i=1}^{I} y_i a_{ij} \leq 1 \ \forall j$, and by dual optimality they satisfy $\sum_{i=1}^{I} y_i \pi_i = b^*(\pi) = \gamma(\mu)$. Thus, if the objects have a discrete distribution, the dual function for the GBPP is exactly $u(w) = y_i$, $w = w(i)$.

Now consider the case where μ has only a continuous part with a bounded, continuous density f. Form the cubic histogram with a given grid size h and let $y(h)$ denote the optimal dual variables in the solution to (5) with $z = \underline{\pi}$. Define $u_h(w) = y_i(h)$, $w \in \mathcal{A}_i(h)$. Recall that $\mathcal{A}_i(h)$ denotes the set of points in $\Re^d$ that, when rounded down, correspond to grid point i. We claim that $u_h(x)$ is dual feasible for all $h > 0$. To see this, take any finite collection $\{w_1, \ldots, w_k\}$ of objects that form a feasible bin and form a new set of objects $\underline{w}_i = h\lfloor w_i/h \rfloor$, $i = 1, \ldots, k$. By Property P4, this collection of rounded down objects also forms a feasible bin, which corresponds to some column A^r in the linear program (5) for grid size h. Further, $\sum_{i=1}^{k} u_h(w_i) = y^T(h)A^r \leq 1$ by the dual feasibility of $y(h)$. Thus, u_h is a dual feasible function.

Let $\{h_n\}_{n=1}^{\infty}$ be a sequence which decreases to zero. Consider the sequence of dual feasible functions $\{u_{h_n}\}_{n=1}^{\infty}$. Now, $\int u_{h_n}(w)\mu\{dw\} = b^{LP}(\underline{\pi}(h_n))$, which is nondecreasing in n as shown in the proof of Theorem 4. Thus, the sequence $\{u_{h_n}\}_{n=1}^{\infty}$ satisfies the properties of Lemma 4.

The case where μ is a mixture of a finite, discrete part and continuous parts follows by defining a fixed number of types of objects corresponding to the discrete sizes, applying the discretization to the remaining items and combining the above arguments. □

Remark. Equation (18) can be obtained in the case where μ is continuous with a bounded density. The idea is to decrease the dual feasible functions u_h to zero near the boundaries of the subregions $\mathcal{A}_i$ to make them continuous. The resulting function is dual feasible, and a sequence of such functions establishes (18).

4. THE CAPACITATED VEHICLE ROUTING PROBLEM WITH TIME WINDOWS

We next apply the GBPP results to the probabilistic analysis of the vehicle routing problem (VRP) with time window constraints, where customer k has a location x_k, demand w_k, a service time s_k for loading/unloading the vehicle, and a time window specified by an earliest delivery time e_k and a latest delivery time l_k, $k = 1, \ldots, n$. A planning period of T units of time is available during which all on-site service must start and end. (Vehicle travel is not necessarily restricted to this window, and can begin and end trips outside the interval.) An unlimited number of vehicles is available, each of which has capacity C and travels at a constant velocity v. The problem is to find a feasible collection of tours of minimum total length that visit all customers. We assume w.l.o.g. that distance, time, and size are scaled so that $T = 1$, $v = 1$ and $C = 1$.

To guarantee the existence of a feasible solution, we must have $0 \leq e_k + s_k \leq l_k \leq 1$. So that all attribute values satisfy property P4, we make a change of variables and specify the end of a window by a variable $r_k = 1 - l_k$, i.e., the time between the latest completion time l_k and the end of the planning period. Note reducing the values e_k, s_k, and r_k relaxes the problem. With this transformation, we must have $s_k + e_k + r_k \leq 1$ to guarantee feasibility.

We first consider a somewhat simpler scheduling problem in which the travel time between customers is ignored (assumed to be zero). The minimum number of vehicles in this case is a lower bound on the number needed when travel time is included. Further, this scheduling problem is an instance of a GBPP, where Ω is the collection of all sets of objects $B = \{(w_k, s_k, e_k, r_k) : k = 1, \ldots, K\}$ that satisfy $\sum_{k=1}^{K} w_k \leq 1$, and whose processing can be scheduled within one unit of time. That is, there exists some set of start times $\{t_1, \ldots, t_K\}$ such that processing takes place in the allowable time windows, i.e., $e_k \leq t_k$ and $t_k + s_k + r_k \leq 1$ for all $k = 1, \ldots, K$, and the processing is nonoverlapping, i.e., if customer i precedes customer j, $t_i + s_i \leq t_j$. We associate one such *feasible schedule* with every feasible bin. Observe that a customer's attributes are represented by a point in the four-dimensional compact set $C = \{(w, s, e, r) : 0 \leq w \leq 1, s \geq 0, e \geq 0, r \geq 0, s + e + r \leq 1\}$. As in the general GBPP, we shall apply a discretization scheme to solve this continuous problem.

With these definitions, we are now ready to state our main theorem:

Theorem 9. *Let $\{x_k : k = 1, \ldots, n\}$ be a sequence of independent random variables in $\Re^2$ with distribution μ_0 that has compact support. Let $d(y)$ denote the Euclidean distance from a point $y \in \Re^2$ to the depot and let*

$$E[d] = \int_{\Re^2} d(y)\mu_0\{dy\}.$$

Let $S^{(n)} = \{(w_k, s_k, e_k, r_k) : k = 1, \ldots, n\}$ be a set of independent vectors with distribution μ, which possesses a bounded density f that is continuous almost everywhere on the interior of C. Let b^* denote the optimal value of the associated GBPP, and $\gamma(\mu) = \lim_{n\to\infty} b^*(S^{(n)})/n$. Finally, let $Z^{(n)}$ denote the optimal value of the time-window VRP for n customers with locations $\{x_k : k = 1, \ldots, n\}$ and attributes $S^{(n)}$ such that x_k and (w_k, s_k, e_k, r_k) are independent of each another. Then,

$$\lim_{n\to\infty} \frac{Z^*(S^{(n)})}{n} = 2\gamma(\mu)E[d] \quad (a.s.).$$

Our proof is based on a modification of the analysis of Bramel et al. (1992) for the Capacitated VRP with unsplit demands. Indeed, we have the following lower bound, which we state without proof since it follows directly from Theorem 4 and Bramel et al. (1992):

Lemma 5. *Under the conditions of Theorem 9, $\lim_{n\to\infty} Z^{(n)}/n \geq 2\gamma(\mu)E[d]$ (a.s.).*

For the upper bound, we consider the region partitioning scheme described in Bramel et al. (1992). Fix $h > 0$ and let $G(h)$ denote the grid of squares with sides $h/\sqrt{2}$ and edges parallel to the coordinate system. Let $j, j = 1, \ldots, t(h)$ index those squares in the grid that have nonzero probability of containing a point drawn from the distribution μ_0, and let $n(j)$ be the number of customers in region j. Bramel et al. (1992) show the following:

Lemma 6. *Suppose we have some upper bound, $\bar{b}(j)$, on the minimum number of vehicles needed to service the customers in region j that satisfies $\lim_{n\to\infty} \bar{b}(j)/n(j) = \bar{\gamma}(\mu)$ (a.s.), $j = 1, \ldots, t(h)$. Then,*

$$\lim_{n\to\infty} \frac{Z^*(S^{(n)})}{n} \leq 2\bar{\gamma}(\mu)E[d] + 2h \quad (a.s.).$$

Using Lemmas 5 and 6, we can prove the theorem provided we can find a $\bar{\gamma}(\mu)$ that is arbitrarily close to $\gamma(\mu)$ for small values of h. To do this, consider the following heuristic to construct feasible tours in each subregion j:

GBPP Heuristic. Apply the region partitioning scheme using a partition with squares of sides $h/\sqrt{2}$, for a given $h > 0$ as described above. Discretize the attributes (w, s, e, r) of customers using a cubic histogram of the same grid size h. Let $i, i = 1, \ldots, I' = (1 + 1/h)^4$ denote the types of objects formed by this discretization. For each subregion j, form tours as follows:

1. Remove customers k with $s_k + e_k + r_k \in (1 - 4h, 1]$. Assign each of these customers to a separate vehicle.
2. For customers k with $s_k + e_k + r_k \in [0, 1 - 4h]$, modify the service time to $s'_k \equiv s_k + h$.
3. Form the vector $\bar{z}'$ where $\bar{z}'_i$ denotes the number of customers in the subregion for which $(h\lceil w_k/h \rceil, h\lceil s'_k/h \rceil, h\lceil e_k/h \rceil, h\lceil r_k/h \rceil)$ corresponds to a type i object in the associated GBPP.
4. Solve the linear program $b^{LP}(\bar{z}')$ and round the solution up to obtain an integer vector. Assign one vehicle to each resulting bin.
5. Have the vehicle visit the customers in each bin according to one of its corresponding feasible schedules.

We are now ready to prove Theorem 9:

Proof. First, we claim that the GBPP heuristic produces a feasible collection of tours. To see this, note that the tours that serve individual customers are clearly feasible. For the remaining customers, we have that $h\lceil s'_k/h \rceil + h\lceil e_k/h \rceil + h\lceil r_k/h \rceil) \leq 1$ so the associated GBPP is always feasible. Also, for each tour that serves a set of customers B, $\sum_{k\in B} w_k \leq 1$, and since $s'_k = s_k + h$ and h is the maximum distance between any two points in a subregion, by following the feasible schedule associated with each bin we can visit and service all demands in no more than one unit of time. Hence, the tours associated with any feasible bin in the associated GBPP with shifted service times are feasible.

Let $b^H(S(j))$ denote the total number of tours the heuristic generates for the set of customers $S(j)$ in region j and let $\delta_j(h) = |\{k \in S(j) : s_k + e_k + r_k > 1 - 4h\}|$ be the number of customers in region j placed in their own tours by Step 1. Then $b^H(S(j))$ satisfies

$$\begin{aligned} b^H(S(j)) &= b^*(\bar{z}') + \delta_j(h) \\ &\leq b^{LP}(\bar{z}') + I' + \delta_j(h) \\ &\leq n(j)b^{LP}(\bar{\pi}') + n(j)|\bar{z}'/n(j) - \bar{\pi}'| \\ &\quad + I' + \delta_j(h), \end{aligned}$$

and therefore

$$\frac{b^H(S(j))}{n(j)} \leq b^{LP}(\underline{\pi}) + |\underline{\pi} - \bar{\pi}'| + |\bar{z}'/n(j) - \bar{\pi}'| + \frac{I'}{n(j)} + \frac{\delta_j(h)}{n(j)}.$$

Noting that as $n \to \infty$, $n(j) \to \infty$ (a.s.) and thus by the strong law of large numbers $|\bar{z}'/n(j) - \bar{\pi}'| \to 0$ (a.s.) and

$$\frac{\delta_j(h)}{n(j)} = \frac{|\{k \in S(j) : s_k + e_k + r_k > 1 - 4h\}|}{n(j)} \to P\{s_k + e_k + r_k > 1 - 4h\} \quad \text{(a.s.)},$$

we obtain

$$\lim_{n\to\infty} \frac{b^H(S(j))}{n(j)} = \lim_{n(j)\to\infty} \frac{b^H(S(j))}{n(j)} \leq b^{LP}(\underline{\pi}) + |\underline{\pi} - \bar{\pi}'|$$
$$+ P\{s_k + e_k + r_k > 1 - 4h\} \quad \text{(a.s.)}.$$

Now $|\underline{\pi} - \bar{\pi}'| \leq |\underline{\pi} - \bar{\pi}| + |\bar{\pi} - \bar{\pi}'|$. If μ has a bounded density f that is continuous almost everywhere, we have from Lemma 2 that for every $\epsilon > 0$, we can choose h so that $|\underline{\pi} - \bar{\pi}| \leq \epsilon/3$. Similarly, noting that $\bar{\pi}_i$ and $\bar{\pi}_i'$ are obtained by integrating f over adjacent subregions $\mathcal{A}_i$ and $\mathcal{A}_i'$, we can make $|\bar{\pi} - \bar{\pi}'| \leq \epsilon/3$ for all sufficiently small h as well. Also, the boundedness of f and the fact that f has its support on C implies $P\{s_k + e_k + r_k > 1 - 4h\} = P\{1 \geq s_k + e_k + r_k > 1 - 4h\} = O(h)$ so this term too can be made less than $\epsilon/3$ for all sufficiently small h. Combining these observations, we have that for every $\epsilon > 0$, there exists a δ, $\epsilon \geq \delta > 0$ such that for all $h \leq \delta$,

$$\lim_{n\to\infty} \frac{b^H(S(j))}{n(j)} = \lim_{n(j)\to\infty} \frac{b^H(S(j))}{n(j)} \leq b^{LP}(\underline{\pi}) + \epsilon \quad \text{(a.s.)}.$$

Now, using the fact that $b^{LP}(\underline{\pi}) \leq \gamma(\mu)$ and combining this with Lemmas 5 and 6, we obtain

$$2\gamma(\mu)E[d] \leq \lim_{n\to\infty} \frac{Z^*(S^{(n)})}{n}$$
$$\leq 2(\gamma(\mu) + \epsilon)E[d] + 2\epsilon \quad \text{(a.s.)}.$$

Since this holds for every $\epsilon > 0$, Theorem 9 follows. □

The running time of the heuristic for a fixed $\epsilon > 0$ (equivalently, fixed $h > 0$) is $O(n)$, since it takes $O(n)$ time to determine each of the n customers' subregion j and object type i and construct the vector $\bar{z}^j$, the linear program is solved in $O(\log(n))$ time and the routes are constructed from this solution in linear time. We discuss complexity and convergence rates further in Section 5.

One can view the GBPP heuristic as an efficient aggregation scheme for solving VRPs as set covering problems, which is one of the oldest approaches to these problems due originally to Balinski and Quandt (1964) (see also Magnanti 1981). In the set covering approach, each of the n customers is treated as a separate "type," and the objective is to cover these n types by a subset of *all* feasible tours. In our scheme, we use cubic histograms to drastically reduce the number of customer types, and, by only considering tours of customers in the same region, we similarly reduce the number of feasible tours to be considered. Thus, the GBPP heuristic can be viewed as a combined row aggregation *and* column restriction scheme.

Lastly, Theorem 9 remains valid for measures μ that are a mixture of a continuous distribution and a finite distribution, provided the marginal distribution for the service time has only a continuous density. This is needed to ensure that $|\bar{\pi} - \bar{\pi}'| \to 0$ as $h \to 0$ in the proof of Theorem 9. In particular, this generalization includes as a subcase models in which only a finite set of possible time windows is allowed; see, e.g., Bramel et al. 1993.

5. COMPUTATIONAL ISSUES IN SOLVING THE LINEAR PROGRAMMING RELAXATION OF A GBPP

We next analyze and discuss column generations techniques for solving the linear programming relaxation (5) and analyze the complexity of the resulting heuristic H.

5.1. Column Generation Techniques

Consider first the classical bin packing problem. In a discretization of this problem, each type i corresponds to a weight $w_i = i/h$, $i = 0, 1, \ldots, I = 1/h$ and a feasible bin $A_j = (a_{1j}, \ldots, a_{Ij})$ is one that satisfies

$$\sum_{i=1}^{I} a_{ij} w_i \leq 1.$$

As mentioned in Section 1, we can ignore zero weight (null) objects; thus, we shall consider only weights $w_i = i/h$, $i = 1, \ldots, I = 1/h$.

Suppose we have an initial feasible basis for the linear program (5). (Recall that putting each type of item in its own bin is always feasible; thus the identity matrix is always a feasible basis.) Associated with this basis is a dual price y. The basis is optimal if and only if $y^T A^j \leq 1$ for all feasible bins j. One can check this condition by solving the knapsack problem, $\max\{y^T x : \sum_{i=1}^{I} x_i w_i \leq 1, x \in \mathcal{Z}_+^I\}$. This can be solved by the dynamic program, $V(k) = \max_{i \leq k}\{V(k - i)) + y_i\}$, $0 < k \leq I$ with boundary condition $V(0) = 0$, where $V(k)$ is the maximum value of a set of objects whose total weight is no more than kh. If $V(I) \leq 1$, the current basis is optimal; if $V(I) > 1$, the solution provides a column of A with a negative reduced cost that can be brought into the basis. Repeating this process, one can generate columns until an optimal basis is found. This *column generation* approach for solving the relaxation of classical bin packing problems dates back to the work of Gilmore and Gomory (1961) and has proved to be quite efficient in practice. Note the complexity of generating a column (finding $V(I)$) is only $O(I^2) = O(h^{-2})$. It is therefore polynomial in the *discretization level* $1/h$.

A similar approach can be used to generate columns for the vector packing problem. For this we let $V(k_1, \ldots, k_d)$ denote the maximum value of a set of objects whose total weight in the jth dimension does not exceed $k_j h$, $j = 1, \ldots, d$. $V(\cdot)$ is found by the recursion,

$$V(k_1, \ldots, k_d) = \max_{\{i_j \leq k_j, 1 \leq j \leq d\}} \{V(k_1 - i_1, \ldots, k_d - i_d)$$
$$+ y(i_1, \ldots, i_d)\},$$

with boundary condition $V(0, \ldots, 0) = 0$, where $0 \leq k_i \leq I$, $i = 1, \ldots, d$ and $y(i_1, \ldots, i_d)$ denotes the current dual value for an object with attributes $(i_1 h, \ldots, i_d h)$. This dynamic program takes $O(I^{\prime 2}) = O(h^{-2d})$ time.

One can also generate columns for the time window VRP in time polynomial in $1/h$. Let $V(k, u)$ be the maximum value of a set of objects which can be scheduled in kh units of time and whose total weight does not exceed uh. The decision variables in the recursion are t, the delay until the start of the next job after time k, and (w, s, e, r), the type of job started at time t. Then $V(k, u)$ satisfies

$$V(0, 0) = 0, \tag{20}$$

$$V(k, u) = \max_{\mathcal{U}(k,u)} \{V(k - t - s, u - w) + y(w, s, e, r)\}, \tag{21}$$

$$0 \leq k \leq I, \quad 0 \leq u \leq I,$$

where $\mathcal{U}(k, u)$ denotes the set of values (t, w, s, e, r) satisfying

$$t + I - k \geq e,$$
$$t + s + r \leq k,$$
$$Tw \leq u,$$
$$(t, w, s, e, r) \in \mathcal{Z}^5_+,$$

and $y(w, s, e, r)$ denotes the current dual value for objects with attributes (wh, sh, eh, rh). The running time for this procedure is $O(I^7) = O(h^{-7})$.

Other schemes can be employed to speed up the solution of the linear program. For example, one need not solve the column generating problems exactly. In practice, it may be better to use a fast heuristic to generate feasible solutions and resort to an exact algorithm only when the heuristic fails to produce a column with positive reduced cost. If the problem has to be solved repeatedly for many instances drawn from the same known distribution μ, e.g., as in the partitioning scheme used for the time window problem, we can use a *hierarchical*, two-stage approach. In stage one, we solve the full linear program once *off-line* using the stationary vector $\underline{\pi}$. This generates an optimal basis, which one can interpret as a set of bins that are "good" candidates to use when packing objects drawn from the distribution μ. At the second level, each instance is solved *on-line* using the considerably reduced set of columns (equivalently, using only the restricted collection of "good" bins). It is not hard to show that this approach is also asymptotically optimal and has the same convergence rates as before.

5.2. Complexity of the Heuristic *H*

In addition to providing a practical approach to solving the linear programming relaxation of a GBPP, the column generation approach can be used to establish complexity bounds on the linear program (5) as a function of the discretization level $1/h$. Combining these bounds with Theorem 6, we are able to provide attractive bounds on both the convergence and complexity of the heuristic H for some specific cases of the GBPP. The analysis follows the approach used by Karmarkar and Karp (1982) for the worst-case analysis of classical bin packing problems.

We use the Grötschel-Lovasz-Schrijver (GLS) (Grötschel et al. 1981 and 1988) central-cut ellipsoid algorithm in our analysis. The GLS algorithm takes as input a rational number $\epsilon > 0$, a polyhedron $P \subset \Re^k$, cost vector c, balls $B(x_1, r)$, $B(x_2, R)$ satisfying $B(x_1, r) \subseteq P \subseteq B(x_2, R)$ and a *separating hyperplane oracle* that, given a vector y, either (i) determines that $y \in P$ or (ii) provides a vector a such that $a^T y > a^T x$ for all $x \in P$, and generates as output a point $y^* \in P$ satisfying $c^T y^* \geq \max_{y \in P} c^T y - \epsilon$. The algorithm makes at most $N = O(k^2 \log(R^2/(r\epsilon))$ calls to the separating hyperplane oracle. Thus, if the oracle runs in polynomial time, then so does the GLS algorithm.

We apply the GLS algorithm to the dual of the linear programming relaxation (5):

$$\begin{aligned} \max \quad & y^T z \\ \text{s.t.} \quad & \\ & y^T A \leq \mathbf{1}, \\ & y \geq 0, \end{aligned}$$

where $\mathbf{1}$ denotes the vector of ones. For a given discretization, the dimension of y is $k = O(1/h^d)$ and in the heuristic H, we would have $z = \bar{z}^{(n)}$, the vector of discrete sizes produced by rounding up. In this case, the separating hyperplane oracle corresponds exactly to a column generation routine; that is, given a price vector y, if the optimal value obtained from the subproblem is at most one, then y is a feasible price ($y \in P$); otherwise, the solution vector a satisfies $a^T y > 1 \geq a^T x$, $\forall x \in P$, so a is a separating hyperplane. Since the number of calls to the hyperplane oracle is polynomial in $1/h$, polynomiality (in $1/h$) of the column generating procedure implies polynomiality (in $1/h$) of the overall linear program.

The precise complexity and convergence rate depends on the problem. For example, consider the classical bin packing problem. Suppose the distribution μ of the sizes has a density that is Lipschitz of order one on the interior of [0, 1]. Then if we select $h = n^{-1/2}$, Theorem 6 shows the error in our heuristic H is $O(\sqrt{n})$, so that

$$\frac{E[b^H(n)]}{n} \leq \gamma(\mu) + O(n^{-1/2}).$$

For this problem, one can show that $h \leq y_i \leq 1$ so that the balls $B(x_1, r) = B(0, 1)$ and $B(x_2, R) = B(0, \sqrt{1/h})$ satisfy $B(x_1, r) \subseteq P \subseteq B(x_2, R)$. Thus, since the dimension $k = 1/h = n^{1/2}$, the GLS algorithm makes at most $O(n \log n)$ calls to the column generating routine. Since the complexity of the dynamic program to generate columns is $O(h^{-2}) = O(n)$, as shown above, the overall complexity of the heuristic H is $O(n^2 \log n)$. This complexity and convergence rate compare favorably with more specialized approaches (Coffman and Lueker 1991).

As the dimension of the problem, d, grows both the convergence and complexity worsen. For example, the same analysis of the vector packing problem in the case where the density is Lipschitz of order one shows that if $h = n^{-1/(1+d)}$, then the error term is $O(n^{d/(d+1)})$ and the running time is $O(n^{4d/(d+1)} \log n)$. If the density is smoother, both convergence and complexity improve. For example, if the density is Lipschitz order $s < d$ everywhere in the vector packing problem, then choosing $h = n^{-1/(d+s)}$ yields an error term $O(n^{d/(d+s)})$ and a complexity of $O(n^{4d/(d+s)}$ $\log n)$; if $s \geq d$, the error is $O(\sqrt{n})$ with the same complexity.

ACKNOWLEDGMENTS

We are indebted to Wan Soo Rhee for pointing out McDiarmid's bounds, for suggesting Lemma 4 used in the proof of Theorem 8 and for many other helpful comments on our initial draft of this paper. We also thank three anonymous referees and an associate editor, who provided many valuable comments and references. The research of both authors was supported by an internal grant from the Columbia University Graduate School of Business.

REFERENCES

BALINSKI, M. AND R. QUANDT. 1964. On an Integer Program for a Delivery Problem. *Opns. Res.* **12,** 300–304.

BLAIR, C. AND R. JEROSLOW. 1982. The Value Function of an Integer Program. *Math. Programming*, **23,** 237–273.

BUCK, R. C. 1978. *Advanced Calculus*. Third Edition, McGraw-Hill, New York.

BRAMEL, J., E. G. COFFMAN, P. W. SHOR, AND D. SIMCHI-LEVI. 1992. Probabilistic Analysis of the Capacitated Vehicle Routing Problem with Unsplit Demands. *Opns. Res.* **40,** 1095–1106.

BRAMEL, J., C. L. LI, AND D. SIMCHI-LEVI. 1993. Probabilistic Analysis of a Vehicle Routing Problem with Time Windows. *Amer. J. Math. Mgmt. Sci.* **13,** 267–322.

BRAMEL, J. AND D. SIMCHI-LEVI. 1997. Probabilistic Analysis and Practical Algorithms for the Vehicle Routing Problem with Time Windows *Opns. Res.* **44,** 501–509.

COFFMAN, E. G. 1982. An Introduction to Proof Techniques for Bin-Packing Algorithms. In *Deterministic and Stochastic Scheduling*. M. A. H. Dempster, J. K. Lenstra and A. H. G. Rinnooy Kan (eds.), Dordrecht, Holland.

COFFMAN, E. G., M. GAREY, AND D. JOHNSON. 1984. Approximation Algorithms for Bin-Packing: An Updated Survey, In *Algorithm Design for Computer System Design*. G. Ausiello, M. Lucertini and P. Serafini (eds.), 49–106, Springer-Verlag.

COFFMAN, E. G., G. LUEKER, AND A. RINNOOY KAN. 1988. Asymptotic Methods in the Probabilistic Analysis of Sequencing and Packing Heuristics. *Mgmt. Sci.* **34,** 266–290.

COFFMAN, E. G. AND G. S. LUEKER. 1991. *Probabilistic Analysis of Packing and Partitioning Algorithms*. John Wiley, New York.

COURCOUBETIS, C. AND R. R. WEBER. 1986. Necessary and Sufficient Conditions for Stability of a Bin Packing System. *J. Appl. Probability*, **23,** 989–999.

DEVROYE, L. AND L. GYORFI. 1985. *Nonparametric Density Estimation, The L_1 View*. John Wiley, New York.

EISEMANN, K. 1957. The Trim Problem. *Mgmt. Sci.* **3,** 279–283.

FEDERGRUEN, A. AND G. J. VAN RYZIN. 1994. Probabilistic Analysis of a Generalized Bin Packing Problem and Applications (unabridged version). Working Paper, Columbia Graduate School of Business.

FREDERICKSON, G. N. 1980. Probabilistic Analysis for Simple One and Two-Dimensional Bin Packing Algorithms. *Information Processing Letters*, **11,** 156–161.

GILMORE, P. C. AND R. E. GOMORY. 1961. A Linear Programming Approach to the Cutting-Stock Problem. *Opns. Res.* **9,** 849–859.

GOLDFARB, D. AND M. J. TODD. 1989. Linear Programming. In *Optimization*, G. L. Nemhauser, A. H. G. Rinnooy Kan, and M. J. Todd (eds.). Handbooks in Operations Research, Elsevier, Amsterdam.

GRÖTSCHEL, M., L. LOVASZ, AND A. SCHRIJVER. 1981. The Ellipsoid Method and its Consequences in Combinatorial Optimization. *Combinatorica*, **1,** 169–197.

GRÖTSCHEL, M., L. LOVASZ, AND A. SCHRIJVER. 1988. *Geometric Algorithms and Combinatorial Optimization*. Springer-Verlag, Berlin.

HAIMOVICH, M. AND A. H. G. RINNOOY KAN. 1985. Bounds and Heuristics for Capacitated Routing Problems. *Math. O. R.* **10,** 527–541.

KARMARKAR, N. AND R. M. KARP. 1982. An Efficient Approximation Scheme for the One-Dimensional Bin-Packing Problem. *Proc. IEEE 23rd Annual Symposium on the Foundations of Computer Science*. Chicago, November 3–5.

KARP, R. M. 1972. Reducibility Among Combinatorial Problems. In *Complexity of Computer Computations*, R. E. Miller and J. W. Thatcher (eds.). 85–103, Plenum Press, New York.

KARP, R. M. 1982. Lecture Notes. Computer Science Division EECS, University of California, Berkeley.

KARP, R. M., M. LUBY, AND A. MARCHETTI-SPACCAMELA. 1984. A Probabilistic Analysis of Multidimensional Bin Packing Problems. Sixteenth ACM Symposium on the Theory of Computing, Washington, DC, April 30–May 2.

KIEFER, J. 1961. On Large Deviations of the empirical D.F. of Vector Chance Variables and a Law of Iterated Logarithm. *Pacific J. Math*, **11,** 649–660.

KINGMAN, J. F. C. 1976. *Subadditive Processes*. Lecture Notes in Mathematics, **539,** Springer-Verlag, Berlin.

KNODEL, W. 1981. A Bin Packing Algorithm with Complexity $O(n \log n)$ and Performance 1 in the Stochastic Limit, published as *Mathematical Foundations of Computer Science 1981, Proceedings, 10th Symposium*, J. Gruska and M. Chytil, eds., Czechoslovakia, August 31–September 4, 1981. *Lecture Notes in Computer Science*, 118, 369–378. Springer-Verlag, Berlin, 1981.

KOLEN, A., A. RINNOOY KAN, AND H. TRIENEKENS. 1987. Vehicle Routing with Time Windows. *Opns. Res.* **35,** 266–273.

LUEKER, G. S. 1982. An Average-Case Analysis of Bin Packing with Uniformly Distributed Item Sizes. Technical Report 181, University of California at Irvine, Department of Information and Computer Science.

MAGNANTI, T. 1981. Combinatorial Optimization and Vehicle Fleet Planning: Perspectives and Prospects. *Networks*, **11,** 179–213.

MASSART, P. 1986. Rates of Convergence in the Central Limit Theorem for Empirical Processes. *Ann. Inst. H. Poincare*, **22,** 381–423.

MCDIARMID, C. 1989. The Method of Bounded Difference Sequences. *Surveys in Combinatorics*. London Mathematical Society Lecture Notes, **141,** Cambridge University Press, Cambridge, 148–188.

POLLARD, D. 1984. *Convergence of Stochastic Processes*. Springer-Verlag, New York.

ROYDEN, H. 1968. *Real Analysis*. MacMillan Company, London.

REVESZ, P. 1968. *The Laws of Large Numbers*. Academic Press, New York.

RHEE, W. T. 1993. Personal communication.

RHEE, W. T. AND M. TALARAND. 1987. Martingale Inequalities and NP-Complete Problems. *Math. O. R.* **12,** 177–181.

RHEE, W. T. AND M. TALAGRAND. 1989a. Optimal Bin Packing with Items of Random Size—II. *SIAM J. Comput.* **18,** 139–498.

RHEE, W. T. AND M. TALAGRAND. 1989b. Optimal Bin Packing with Items of Random Size—III. *SIAM J. Comput.* **18,** 473–486.

RHEE, W. T. AND M. TALAGRAND. 1991. Multidimensional Optimal Bin Packing with Items of Random Size. *Math. O. R.* **16,** 490–503.

SERFLING, R. 1986. *Approximation Theorems in Mathematical Statistics*. John Wiley, New York.

SOLOMON, M. M. 1987. Algorithms for Vehicle Routing and Scheduling Problems with Time Window Constraints. *Opns. Res.* **35,** 254–265.

SOLOMON, M. M. AND J. DEROSIERS. 1988. Time Window Constrained Routing and Scheduling Problems. *Transp. Sci.* **22,** 1–13.

TALAGRAND, M. 1994. Sharper Bounds for Gaussian and Empirical Processes. *Ann. Probability*, **22,** 28–76.