AVERAGE-CASE ANALYSIS OF A GREEDY ALGORITHM FOR THE 0/1 KNAPSACK PROBLEM

JAMES M. CALVIN AND JOSEPH Y-T. LEUNG

Abstract. We consider the average-case performance of a well-known approximation algorithm for the 0/1 knapsack problem, the Decreasing Density Greedy (DDG) algorithm. Let $U_n = \{u_1, \dots, u_n\}$ be a set of n items, with each item u_i having a size s_i and a profit p_i, and K_n be the capacity of the knapsack. Given an instance of the 0/1 knapsack problem, let P_L denote the total profit of an optimal solution of the linear version of the problem (i.e., a fraction of an item can be packed in the knapsack) and P_{DDG} denote the total profit of the solution obtained by the DDG algorithm. Assuming that U_n is a random sample from the uniform distribution over $(0,1]^2$ and $K_n = \sigma n$ for some constant $0 < \sigma < 1/2$, we show that $\sqrt{n}(P_L - P_{DDG})$ converges in distribution.

Key words and phrases. 0/1 knapsack; NP-hard; Approximation algorithms; Probabilistic analysis

1. Introduction

The 0/1 knapsack problem has been studied extensively in the past two decades. The problem can be stated as follows: Given a knapsack with capacity K_n and a set $U_n = \{u_1, u_2, \dots, u_n\}$ of n items, with each item u_i having a size s_i and a profit p_i, find a subset $U_0 \subset U_n$ such that the total size of U_0 is no more than K_n and the total profit of U_0 is maximized. An instance of the 0/1 knapsack problem will be denoted by the ordered pair (U_n, K_n).

The 0/1 knapsack problem is known to be NP-hard [3], so it is unlikely that the problem can be solved in polynomial time. Because of the computational complexity of the problem, a lot of attention has been given to approximation algorithms [4, 5, 7, 9, 10, 11, 12, 13]. Sahni [13] gave a polynomial time approximation scheme and Ibarra and Kim [7] gave a fully polynomial time approximation scheme for the problem; the reader is referred to [3] for an exposition of these two schemes. An $O(n \lg n)$-time algorithm, called the Decreasing Density Greedy (DDG) algorithm in this article, has also been proposed [3], which works as follows: Scanning the items in nonincreasing order of their densities (profit versus size), pack as many items into the knapsack as possible. The DDG algorithm then outputs the better of this solution and the one obtained by merely taking the item

Date: February 4, 2003.

1991 *Mathematics Subject Classification.* Primary 60J65, 68Q25; Secondary 11Y16, 73K40.

The work of the first author was supported in part by NSF grant DMI-9900117.

with the largest profit. It is known that the DDG algorithm has a worst-case performance bound of 2 [3].

In the 0/1 knapsack problem an item is either completely packed in the knapsack, or not at all. A slight variation of this problem is to allow a fraction of an item to be packed in the knapsack. To distinguish between these two problems, we shall call the original problem the integer version and the other problem the linear version. Unlike the integer version, optimal solutions for the linear version can be found in $O(n \lg n)$ time by a slight variation of the DDG algorithm: Pack the items in nonincreasing order of their densities until an item cannot be completely packed in the knapsack, at which time pack the largest possible fraction of this item in the knapsack. Throughout this article we let P_L denote the total profit of an optimal solution of the linear version and P_{DDG} denote the total profit of the solution produced by the DDG algorithm.

Lueker [9] analyzed the expected difference between the total profits of the optimal solutions of the two versions. The probabilistic model he used is that the selection of the n items are viewed as a random placement of n points in the unit square, with the first (resp. second) component of each point representing the size (resp. profit) of the item, and the knapsack size $K_n = \sigma n$, $0 < \sigma < 1/2$. Under this model, Lueker [9] showed that the expected difference between the total profits of the optimal solutions of the two versions is $O(n^{-1} \log^2 n)$. Goldberg and Marchetti-Spaccamela [5] later showed that this bound is tight. Dyer and Frieze [2] extended Lueker's result to the m-dimensional case, $m \geq 1$.

Under a certain probabilistic model, Meanti et al. [12] studied the asymptotic behavior of the total profits of optimal (integer) solutions for the m-dimensional 0/1 knapsack problem. They showed that the sequence of total profits, properly normalized, converges almost surely to a value that is a function of the capacities of the knapsack. Using results from the theory of empirical processes, Geer and Stougie [4] established a convergence rate for the above result, and the convergence rate obtained is sharp in most cases.

There are also probabilistic analyses of algorithms for the 0/1 knapsack problem. Under Lueker's model [9], Goldberg and Marchetti-Spaccamela [5] showed that for any given $\epsilon > 0$, there is a polynomial-time algorithm that finds an optimal (integer) solution with probability no less than $1-\epsilon$. Meanti et al. [12] proposed a greedy algorithm for the m-dimensional 0/1 knapsack problem. Under a certain probabilistic model, they showed that the ratio of the total profit of an optimal (integer) solution versus that obtained by the greedy algorithm converges to one, almost surely. We note that their algorithm is exactly the DDG algorithm when $m = 1$. Marchetti-Spaccamela and Vercellis [11] proposed a linear-time, on-line algorithm. Under Lueker's model [9], they showed that the expected difference between the total profit of an optimal (integer) solution and that obtained by their algorithm is $O(\log^{3/2} n)$. Recently, Lueker [10] proposed another on-line algorithm, and showed that the expected difference between the total profit of an optimal (integer) solution and that obtained by his algorithm is $\Theta(\log n)$. He also showed that no other on-line algorithm can improve over his algorithm by more than $o(\log n)$. Finally, he gave a description of the expected value of the optimal (integer) solution that is the most precise so far.

In this article we follow the probabilistic model in [9]; i.e., we assume that the set $U_n = \{u_1, \ldots, u_n\}$ of n items is a random sample from the uniform distribution over $(0,1]^2$ and that the capacity K_n of the knapsack is σn for some positive constant $0 < \sigma < 1/2$. Note that in our model the size and profit attributes are independent. We shall show that

$$\lim_{n\to\infty} P\left(\left(\frac{n}{\gamma(\sigma)}\right)^{1/2}(P_L - P_{DDG}) \le x\right) = F(x) \tag{1.1}$$

for a continuous distribution function F, where

$$\gamma(\sigma) = \begin{cases} (6\sigma)^{-1/2} & \text{if } 0 < \sigma \le \frac{1}{6}, \\ 3(1/2-\sigma) & \text{if } \frac{1}{6} < \sigma < \frac{1}{2}. \end{cases} \tag{1.2}$$

We conjecture that $F(x) = 1 - \exp(-x^2/2)$.

2. Preliminaries

We assume that each problem instance $\{(s_i, p_i) : 1 \le i \le n\}$ is arranged in nonincreasing order of the densities p_i/s_i. Following the DDG algorithm, let δ_0 be the density of the first item that will not fit (the *critical density*), and let κ_0 be the remaining capacity when the first item is encountered that will not fit. After this point, let δ_i be the density of the ith item to be packed, and κ_i the remaining capacity after it is packed, for $i \ge 1$. Let N be the index of the last item packed. The $\{\delta_i\}$ is a nonincreasing sequence, and

$$\begin{aligned} P_L - P_{DDG} &= \delta_0\kappa_0 - \sum_{i=1}^{N} \delta_i(\kappa_{i-1} - \kappa_i) \\ &= \sum_{i=1}^{N}(\delta_0 - \delta_i)(\kappa_{i-1} - \kappa_i) \\ &= \sum_{i=0}^{N-1} \kappa_i(\delta_i - \delta_{i+1}) + \kappa_N\delta_N. \end{aligned}$$

The difference in profits is represented as the shaded area in Figure 1.

To determine recursive relations satisfied by the κ_i and the δ_i, refer to Figure 2 and recall that the points below the line with slope δ_i are uniformly distributed. We can express the κ_i's by the relation

$$\kappa_{i+1} = (1-\zeta_i)\kappa_i \tag{2.1}$$

for $i \ge 1$, where $\{\zeta_i\}$ is a (doubly infinite) sequence of independent random variables on $[0,1]$ with $P(\zeta_1 \le z) = z^2$. That is, the weight of the $(i+1)$st item taken is $\zeta_i\kappa_i$.

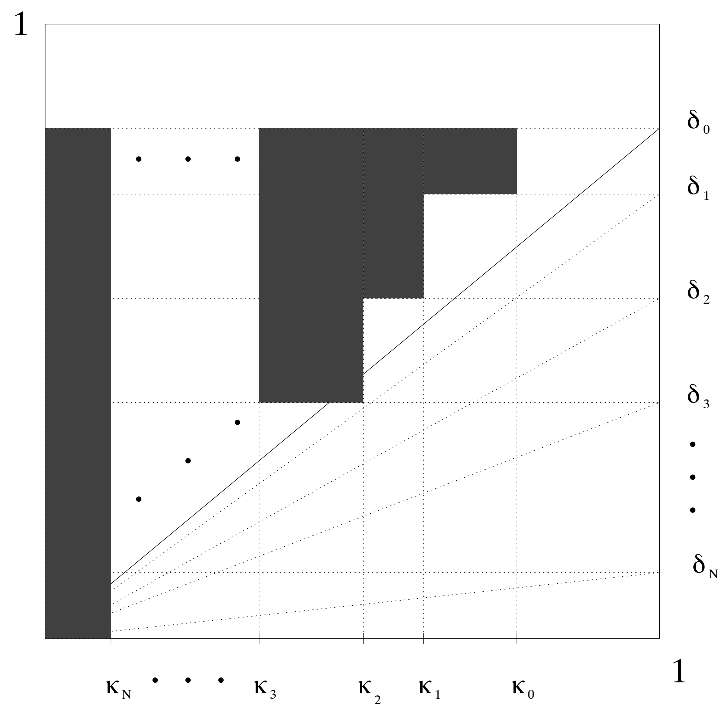

FIGURE 1. Representation of the difference in profits.

We use the notation $X \stackrel{\mathcal{D}}{=} Y$ (resp. $X \stackrel{\mathcal{D}}{\leq} Y$) to indicate that random variables X and Y have the same distribution (resp. X is stochastically smaller than Y). $X_n \xrightarrow{P} X$ indicates that X_n converges in probability to X; that is, for any $\epsilon > 0$, $P(|X_n - X| > \epsilon) \to 0$. If $Ef(X_n) \to Ef(X)$ for every bounded, continuous function f, then we write $X_n \xrightarrow{\mathcal{D}} X$ (X_n converges in distribution to X).

Conditioning on δ_i, κ_i, we obtain

$$\begin{aligned} P\left(\frac{\delta_{i+1}}{\delta_i} \leq x \,\middle|\, \delta_i, \kappa_i\right) &= P(\text{triangle of area } \frac{1}{2}\kappa_i(\delta_i - \delta_i x)\kappa_i \text{ is empty}) \\ &= \left(1 - \frac{1}{2}\kappa_i^2\delta_i(1-x)\right)^n, \end{aligned}$$

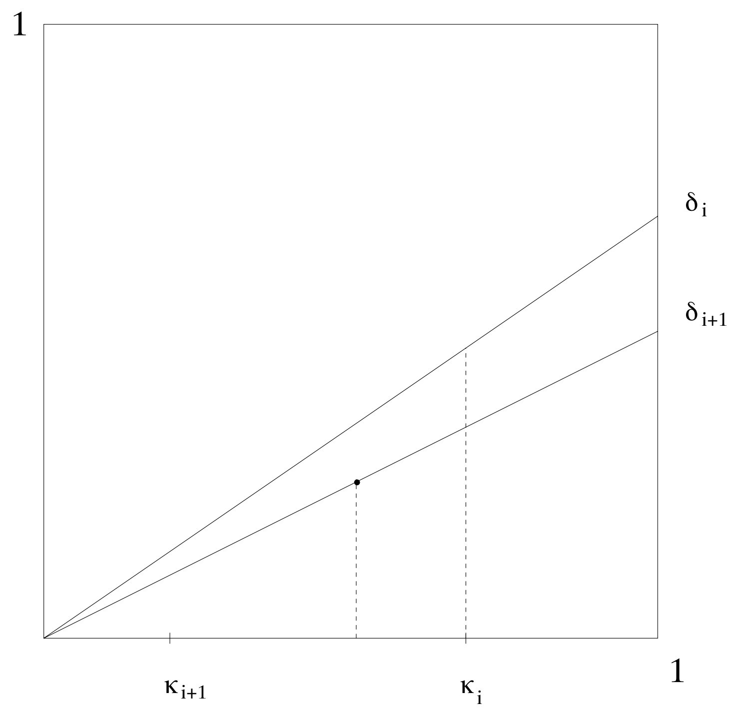

FIGURE 2. Relation between (δ_i, κ_i) and $(\delta_{i+1}, \kappa_{i+1})$.

and so given (δ_i, κ_i),

$$\frac{\delta_{i+1}}{\delta_i} \stackrel{\mathcal{D}}{=} \left(1 - \frac{2(1 - V_i^{1/n})}{\kappa_i^2 \delta_i}\right)^+$$

where V_i is uniformly distributed over $[0, 1]$, and $x^+ = \max(0, x)$. Alternatively,

$$\delta_{i+1} = \left(\delta_i - \frac{2(1 - V_i^{1/n})}{\kappa_i^2}\right)^+$$

or

$$\delta_{i+1} = \left(\delta_0 - \sum_{j=0}^{i} \frac{2(1 - V_j^{1/n})}{\kappa_j^2}\right)^+ . \tag{2.2}$$

Let N be such that

$$\sum_{j=0}^{N-1} \frac{2(1-V_j^{1/n})}{\kappa_j^2} < \delta_0 \quad \text{and} \quad \sum_{j=0}^{N} \frac{2(1-V_j^{1/n})}{\kappa_j^2} \geq \delta_0.$$

Then using (2.2) the difference in profits is

$$\begin{aligned} P_L - P_{DDG} &= \sum_{i=0}^{N-1} \kappa_i(\delta_i - \delta_{i+1}) + \kappa_N \delta_N \\ &= \sum_{i=0}^{N-1} \frac{2(1-V_i^{1/n})}{\kappa_i} + \kappa_N \left(\delta_0 - \sum_{j=0}^{N-1} \frac{2(1-V_j^{1/n})}{\kappa_j^2} \right). \end{aligned}$$

Therefore,

$$\left(\frac{n}{\delta_0}\right)^{1/2} (P_L - P_{DDG}) = \sum_{j=0}^{N-1} \frac{2n(1-V_j^{1/n})}{\sqrt{n\delta_0}\kappa_j} + \sqrt{n\delta_0}\,\kappa_N \left(1 - \sum_{j=0}^{N-1} \frac{2n(1-V_j^{1/n})}{n\delta_0\kappa_j^2} \right). \tag{2.3}$$

For our main result we will need the following lemma.

Lemma 2.1. *For $A > 0$, let j_n^A be the index such that*

$$\kappa_{j_n^A+1} < \frac{A}{\sqrt{n}} \leq \kappa_{j_n^A}. \tag{2.4}$$

As $A \uparrow \infty$,

$$\sum_{i=0}^{j_n^A} \frac{2n(1-V_i^{1/n})}{\sqrt{n\delta_0}\kappa_i} \xrightarrow{P} 0, \quad \textit{and} \quad \sum_{i=0}^{j_n^A} \frac{2n(1-V_i^{1/n})}{\sqrt{n\delta_0}\kappa_i^2} \xrightarrow{P} 0.$$

Proof. Since

$$\kappa_{j_n^A} \geq A/\sqrt{n},$$

it follows that

$$\sum_{i=0}^{j_n^A} \frac{2n(1-V_i^{1/n})}{\sqrt{n\delta_0}\kappa_i} = \sum_{i=0}^{j_n^A} \frac{2n(1-V_{j_n^A-i}^{1/n})}{\sqrt{n\delta_0}\,\kappa_{j_n^A}} \prod_{l=1}^{i}(1-\zeta_{j_n^A-l})$$

$$\leq \sum_{i=0}^{j_n^A} \frac{2n(1-V_{j_n^A-i}^{1/n})}{\sqrt{n\delta_0}\,A/\sqrt{n}} \prod_{l=1}^{i}(1-\zeta_{j_n^A-l})$$

$$\stackrel{\mathcal{D}}{\leq} \sum_{i=0}^{\infty} \frac{2n(1-V_i^{1/n})}{\sqrt{n\delta_0}A/\sqrt{n}} \prod_{l=1}^{i}(1-\zeta_l)$$

$$= \frac{2}{A\sqrt{\delta_0}} \sum_{i=0}^{\infty} n(1-V_i^{1/n}) \prod_{l=1}^{i}(1-\zeta_l). \tag{2.5}$$

Using the independence of the $\{\zeta_i\}$ and the facts that

$$E\,n(1-V_i^{1/n}) = \frac{n}{n+1}, \qquad E(1-\zeta_l) = \frac{1}{3},$$

taking expectations of (2.5) gives

$$E\sum_{i=0}^{j_n^A} \frac{2n(1-V_i^{1/n})}{\sqrt{n\delta_0}\kappa_i} \leq E\frac{2}{A\sqrt{\delta_0}} \sum_{i=0}^{\infty} n(1-V_i^{1/n}) \prod_{l=1}^{i}(1-\zeta_l)$$

$$= EE\left(\frac{2}{A\sqrt{\delta_0}} \sum_{i=0}^{\infty} n(1-V_i^{1/n}) \prod_{l=1}^{i}(1-\zeta_l) \,\Big|\, \{\zeta_i\}\right)$$

$$= E\left(\frac{2}{A\sqrt{\delta_0}} \sum_{i=0}^{\infty} \frac{n}{n+1} \prod_{l=1}^{i}(1-\zeta_l)\right)$$

$$\leq E\left(\frac{2}{A\sqrt{\delta_0}} \sum_{i=0}^{\infty} \prod_{l=1}^{i}(1-\zeta_l)\right)$$

$$= \frac{2}{A\sqrt{\delta_0}} \sum_{i=0}^{\infty} \left(\frac{1}{3}\right)^i = \frac{3}{A\sqrt{\delta_0}},$$

which converges to 0 as $A \uparrow \infty$. Since convergence in expectation to 0 implies convergence in probability to 0 for nonnegative random variables, the first statement is proved. The second statement is proved in a similar fashion. □

We record here the limiting distribution of the undershoot and overshoot of a fixed level by the sequence $\{\kappa_i\}$.

Lemma 2.2. *For fixed $a > 0$, let*

$$K_- = \frac{\max\{\kappa_i : \kappa_i < a\}}{a}, \qquad K_+ = \frac{\min\{\kappa_i : \kappa_i \geq a\}}{a}.$$

Then for $x \geq 1, y \leq 1$,

$$\lim_{\kappa_0 \to \infty} P(K_+ > x, K_- \leq y) = \frac{1}{3}\left(4\frac{y}{x} - \left(\frac{y}{x}\right)^2\right). \tag{2.6}$$

In particular,

$$\lim_{\kappa_0 \to \infty} P(K_- \leq s) = s(4-s)/3, \qquad s \leq 1,$$

and

$$\lim_{\kappa_0 \to \infty} P(K_+ \leq s) = 1 + \frac{1}{3s^2} - \frac{4}{3s}, \qquad s \geq 1.$$

Proof. Consider the random walk $S_n = \log(\kappa_n) = \log(\kappa_0) + \sum_{i=0}^{n} \log(1-\zeta_i)$. Let N be such that $S_{N-1} \geq \log(a)$ and $S_N < \log(a)$. Then (see Karlin and Taylor [8], p. 193)

$$(S_{N-1} - \log(a), \log(a) - S_N) \xrightarrow{\mathcal{D}} (\gamma, \nu), \tag{2.7}$$

where

$$P(\gamma > y, \nu > x) = \frac{4}{3}e^{-(x+y)} - \frac{1}{3}e^{-2(x+y)}. \tag{2.8}$$

Therefore,

$$\begin{aligned}\left(\frac{k_{N-1}}{a}, \frac{k_{N+1}}{a}\right) &= \left(\frac{e^{S_{N-1}}}{a}, \frac{e^{S_{N+1}}}{a}\right) \\ &= \left(e^{S_{N-1}-\log(a)}, e^{S_{N+1}-\log(a)}\right) \\ &\xrightarrow{\mathcal{D}} \left(e^{\gamma}, e^{-\nu}\right),\end{aligned}$$

and (2.6) is obtained from (2.8). □

3. Limit Random Variable

We now define a limiting random variable that appears in our main convergence theorem.

Let $\{\xi_i\}$ be a sequence of independent exponential random variables with mean 2. For $k = 1, 2$ let W_k be a random variable with the distribution of

$$\begin{aligned}&\xi_0 + \xi_1(1-\zeta_1)^k + \xi_2(1-\zeta_1)^k(1-\zeta_2)^k + \cdots \\ &\qquad = \sum_{i=0}^{\infty} \xi_i \prod_{j=1}^{i}(1-\zeta_j)^k, \qquad \left(\prod_{\emptyset} \triangleq 1\right).\end{aligned}$$

W_1 and W_2 have finite second moments, with

$$\begin{aligned} EW_1 &= 3, \qquad & EW_1^2 &= 72/5, \\ EW_2 &= 12/5, \qquad & EW_2^2 &= 72/7. \end{aligned}$$

We now define a doubly infinite sequence $\{\cdots, k_{-1}, k_0, k_1, \cdots\}$ as follows. Let (k_{-1}, k_0) be chosen according to the joint distribution given at (2.6), and for $i \neq 0$ set $k_{i+1} = (1-\zeta_{i+1})^{-1} k_i$. For $m > 1$ define $N(m)$ by

$$\sum_{j=N(m)}^{m} \frac{\xi_j}{k_j^2} < 1 \quad \text{and} \quad \sum_{j=N(m)-1}^{m} \frac{\xi_j}{k_j^2} \geq 1, \tag{3.1}$$

and set

$$\Delta(m) = \sum_{j=N(m)}^{m} \frac{\xi_j}{k_j} + k_{N(m)-1}\left(1 - \sum_{j=N(m)}^{m} \frac{\xi_j}{k_j^2}\right).$$

Proposition 3.1. As $m \uparrow \infty$, $\Delta(m)$ converges in distribution.

Denote the limit random variable by Δ, and its distribution function by F.

Proof. Since $N(m)$ is nondecreasing, there is a random variable $N(\infty)$ such that $N(m) \uparrow N(\infty)$. The limit is a proper random variable, since

$$P(N(\infty) > n) = P\left(\sum_{j=n}^{\infty} \frac{\xi_j}{k_j^2} \geq 1\right) \to 0$$

as $n \to \infty$.

Since $N(m)$ is increasing in m, for any $i > 0$,

$$N(m) \leq N(m+i) \leq m \leq m+i.$$

Then $\Delta(m+i) - \Delta(m) \xrightarrow{P} 0$ as $m \to \infty$, since

$$\begin{aligned}
\Delta(m+i) - \Delta(m) &= \sum_{j=N(m+i)}^{m+i} \frac{\xi_j}{k_j} + k_{N(m+i)-1}\left(1 - \sum_{j=N(m+i)}^{m+i} \frac{\xi_j}{k_j^2}\right) \\
&\quad - \sum_{j=N(m)}^{m} \frac{\xi_j}{k_j} - k_{N(m)-1}\left(1 - \sum_{j=N(m)}^{m} \frac{\xi_j}{k_j^2}\right) \\
&= \sum_{j=m+1}^{m+i} \frac{\xi_j}{k_j} - \sum_{j=N(m)}^{N(m+i)-1} \frac{\xi_j}{k_j} \\
&\quad + \left(k_{N(m+i)-1} - k_{N(m)-1}\right)\left(1 - \sum_{j=N(m+i)}^{m} \frac{\xi_j}{k_j^2}\right) \\
&\quad - k_{N(m+i)-1} \sum_{j=m+1}^{m+i} \frac{\xi_j}{k_j^2} + k_{N(m)-1} \sum_{j=N(m)}^{N(m+i)-1} \frac{\xi_j}{k_j^2}.
\end{aligned}$$

Each of the terms in the last expression converge in probability to zero, because $N(m) \to N(\infty)$, $k_{N(m)} \to k_{N(\infty)}$, and

$$\sum_{j=m+1}^{m+i} \frac{\xi_j}{k_j^l} \xrightarrow{P} 0$$

as $m \to \infty$ for $l = 1, 2$.

Therefore, $\lim_{m\to\infty} P(\Delta(m) \le x) = F(x)$ for some subdistribution function F. To show that F is a distribution function (that is, $F(x) \to 1$ as $x \to \infty$), it is necessary to show that the $\{\Delta(m) : m \ge 1\}$ is a tight family, for which it suffices to show that $\sup_m E\Delta(m)^2 < \infty$ (see Billingsley [1]). To this end, note that

$$\begin{aligned}
\Delta(m) &= \sum_{j=N(m)}^{m} \frac{\xi_j}{k_j} + k_{N(m)-1}\left(1 - \sum_{j=N(m)}^{m} \frac{\xi_j}{k_j^2}\right) \\
&\le \sum_{j=N(m)}^{m} \frac{\xi_j}{k_j} I_{\{k_j \ge 1\}} + \sum_{j=N(m)}^{m} \frac{\xi_j}{k_j} I_{\{k_j < 1\}} + k_{N(m)-1} \\
&\le \sum_{j=N(m)}^{m} \frac{\xi_j}{k_j} I_{\{k_j \ge 1\}} + \sum_{j=N(m)}^{m} \frac{\xi_j}{k_j^2} I_{\{k_j < 1\}} + k_{N(m)-1} \\
&\le \sum_{j=N(m)}^{m} \frac{\xi_j}{k_j} I_{\{k_j \ge 1\}} + 1 + k_{N(m)-1} \\
&\le \sum_{j=0\vee N(m)}^{m} \frac{\xi_j}{k_j} + 1 + k_{N(m)-1} \\
&\triangleq 1 + \Psi(m).
\end{aligned}$$

Let $N(m)^+ = 0 \vee N(m)$, so that

$$\begin{aligned}
\Psi(m) &= \sum_{j=N(m)^+}^{m} \frac{\xi_j}{k_j} + k_{N(m)-1} \\
&= \frac{1}{k_{N(m)^+}} \sum_{j=N(m)^+}^{m} \xi_j \prod_{i=1+N(m)^+}^{j} (1 - \zeta_i) + k_{N(m)-1} \\
&\triangleq \frac{1}{k_{N(m)^+}} Z(m) + k_{N(m)-1}.
\end{aligned}$$

Thus

$$\begin{aligned}
\Psi(m)^2 &= \frac{1}{k_{N(m)^+}^2} Z(m)^2 + k_{N(m)-1}^2 + 2\frac{k_{N(m)-1}}{k_{N(m)^+}} Z(m) \\
&\le Z(m)^2 + k_{N(m)-1}^2 + 2Z(m),
\end{aligned}$$

since $k_{N(m)^+} \geq 1$. Because $Z(m) \overset{\mathcal{D}}{\leq} W_1$, the first and last terms of the previous expression have bounded expectations. By (3.1),

$$\sum_{j=N(m)-1}^{m} \frac{\xi_j}{k_j^2} = \frac{1}{k_{N(m)-1}^2} \sum_{j=N(m)-1}^{m} \xi_j \prod_{i=N(m)}^{j} (1-\zeta_i)^2 \geq 1$$

and so it follows that

$$k_{N(m)-1}^2 \leq \sum_{j=N(m)-1}^{m} \xi_j \prod_{i=N(m)}^{j} (1-\zeta_i)^2 \overset{\mathcal{D}}{\leq} W_2. \tag{3.2}$$

Therefore, $k_{N(m)-1}$, $\Psi(m)$ and thus $\Delta(m)$ has bounded second moments. This implies that $\lim_{x\to\infty} F(x) = 1$, that is, that $\Delta(m) \xrightarrow{\mathcal{D}} \Delta$, where $P(\Delta \leq x) = F(x)$. □

4. Main Result

We are now prepared to show that the difference in profits, suitably normalized, converges to the random variable Δ defined in the previous section.

Theorem 4.1. *As $n \to \infty$,*

$$\left(\frac{n}{\delta_0}\right)^{1/2} (P_L - P_{DDG}) \xrightarrow{\mathcal{D}} \Delta. \tag{4.1}$$

Proof. Scale distances by $\sqrt{n}$, and fix $A > 0$. Let N_n be the point process on $[0, A]^2$ induced by the scaled points; i.e., for Borel sets $B \subset [0, A]^2$,

$$N_n(B) = \sum_{i=1}^{n} I_{\{\sqrt{n}(p_i, s_i) \in B\}}.$$

It is well known that $N_n \xrightarrow{\mathcal{D}} N$, where N is a Poisson point process with unit intensity on $[0, A]^2$. Therefore,

$$\{2n(1 - V_i^{1/n}); j_n^A \leq i \leq N\} \xrightarrow{\mathcal{D}} \{\xi_{j_n^A}, \dots, \xi_N\}, \tag{4.2}$$

and

$$\left(\sqrt{n\delta_0}\kappa_N, \sqrt{n\delta_0}\kappa_{N-1}, \cdots, \sqrt{n\delta_0}\kappa_{j_n^A+1}\right) \xrightarrow{\mathcal{D}} \left(k_{N(\infty)-1}, k_{N(\infty)}, \cdots, k_{m^A}\right), \tag{4.3}$$

where m^A is the random index defined by

$$k_{m^A} < A\sqrt{\delta_0} \leq k_{m^A+1}.$$

Therefore,

$$
\begin{aligned}
&\left(\frac{n}{\delta_0}\right)^{1/2}(P_L - P_{DDG}) = \\
&\sum_{i=0}^{N-1}\frac{2n(1-V_i^{1/n})}{\sqrt{n\delta_0}\kappa_i} + \sqrt{n\delta_0}\kappa_N\left(1-\sum_{i=0}^{N-1}\frac{2n(1-V_i^{1/n})}{n\delta_0\kappa_i^2}\right) \\
&= \sum_{i=j_n^A+1}^{N-1}\frac{2n(1-V_i^{1/n})}{\sqrt{n\delta_0}\kappa_i} + \sqrt{n\delta_0}\kappa_N\left(1-\sum_{i=j_n^A+1}^{N-1}\frac{2n(1-V_i^{1/n})}{n\delta_0\kappa_i^2}\right) \\
&+ \sum_{i=0}^{j_n^A}\frac{2n(1-V_i^{1/n})}{\sqrt{n\delta_0}\kappa_i} - \sqrt{n\delta_0}\kappa_N\sum_{i=0}^{j_n^A}\frac{2n(1-V_i^{1/n})}{n\delta_0\kappa_i^2}.
\end{aligned}
\tag{4.4}
$$

By (4.2) and (4.3),

$$
\sum_{i=j_n^A+1}^{N-1}\frac{2n(1-V_i^{1/n})}{\sqrt{n\delta_0}\kappa_i} + \sqrt{n\delta_0}\kappa_N\left(1-\sum_{i=j_n^A+1}^{N-1}\frac{2n(1-V_i^{1/n})}{n\delta_0\kappa_i^2}\right) \xrightarrow{\mathcal{D}} \Delta(m^A).
$$

The last two terms of (4.4) each converges to 0 in probability as $A \to \infty$ by Lemma 2.1, since

$$
\sqrt{n\delta_0}\kappa_N \xrightarrow{\mathcal{D}} k_{N(\infty)-1}
$$

which is bounded in probability by (3.2). Letting $A \uparrow \infty$, $\Delta(m^A) \xrightarrow{\mathcal{D}} \Delta$, which completes the proof. □

To complete the proof of (1.1), we need the following Lemma.

Lemma 4.2. *As $n \to \infty$, $\delta_0 \xrightarrow{P} \gamma(\sigma)$, where γ is defined at (1.2).*

Proof. Let $W_n(d)$ denote the combined weights of all items with densities at least d:

$$
W_n(d) = \sum_{i=1}^{n} s_i I_{\{p_i/s_i \ge d\}}.
$$

The moments of the summands are given by

$$
\begin{aligned}
E\left(s_i I_{\{p_i/s_i\ge d\}}\right)^k &= \int_{s=0}^{1}\int_{p=0}^{1}\left(sI_{\{p/s\ge d\}}\right)^k\,dp\,ds \\
&= \int_{s=0}^{1\wedge d^{-1}}\int_{p=sd}^{1} s^k\,dp\,ds \\
&= \int_{s=0}^{1\wedge d^{-1}}(s^k - ds^{k+1})\,ds \\
&= \frac{(1\wedge d^{-1})^{k+1}}{k+1} - \frac{d(1\wedge d^{-1})^{k+2}}{k+2}.
\end{aligned}
$$

In particular,

$$E\left(s_i I_{\{p_i/s_i \geq d\}}\right) = \begin{cases} \frac{1}{6d^2}, & \text{if } d > 1, \\ \frac{1}{2} - \frac{d}{3}, & \text{if } d \leq 1, \end{cases}$$

and

$$E\left(s_i I_{\{p_i/s_i \geq d\}}\right)^2 = \begin{cases} \frac{1}{12d^3}, & \text{if } d > 1, \\ \frac{1}{3} - \frac{d}{4}, & \text{if } d \leq 1. \end{cases}$$

The second moment is bounded in d, and so the law of large numbers implies that

$$\frac{W_n(d)}{n} \xrightarrow{P} \begin{cases} \frac{1}{6d^2}, & \text{if } d > 1, \\ \frac{1}{2} - \frac{d}{3}, & \text{if } d \leq 1. \end{cases}$$

Let

$$d_n^* = \max\{d : W_n(d) \geq \sigma n\} = \max\{d : W_n(d)/n \geq \sigma\}.$$

Then

$$d_n^* \xrightarrow{P} \begin{cases} (6\sigma)^{-1/2}, & \text{if } 0 < \sigma \leq \frac{1}{6}, \\ 3(1/2 - \sigma), & \text{if } \frac{1}{6} < \sigma < \frac{1}{2}. \end{cases}$$

□

We have now established (1.1), since

$$\left(\frac{n}{\gamma(\sigma)}\right)^{1/2} (P_L - P_{DDG}) = \left(\frac{\delta_0}{\gamma(\sigma)}\right)^{1/2} \left(\frac{n}{\delta_0}\right)^{1/2} (P_L - P_{DDG}) \xrightarrow{\mathcal{D}} \Delta,$$

because

$$\left(\frac{\delta_0}{\gamma(\sigma)}\right)^{1/2} \xrightarrow{P} 1$$

by Lemma 4.2 and

$$\left(\frac{n}{\delta_0}\right)^{1/2} (P_L - P_{DDG}) \xrightarrow{\mathcal{D}} \Delta$$

by Theorem 4.1.

We have no explicit expression for the distribution function of Δ. In Figure 3 we plot an empirical distribution function for Δ (based on 100,000 samples).

5. Conclusions

The Decreasing Density Greedy algorithm is an $O(n \lg n)$-time algorithm that produces results that differ from the optimum by order $n^{-1/2}$ under the probabilistic model considered in this paper. It would be interesting to explore heuristics that improve on the $n^{-1/2}$ error rate while maintaining reasonable time complexities.

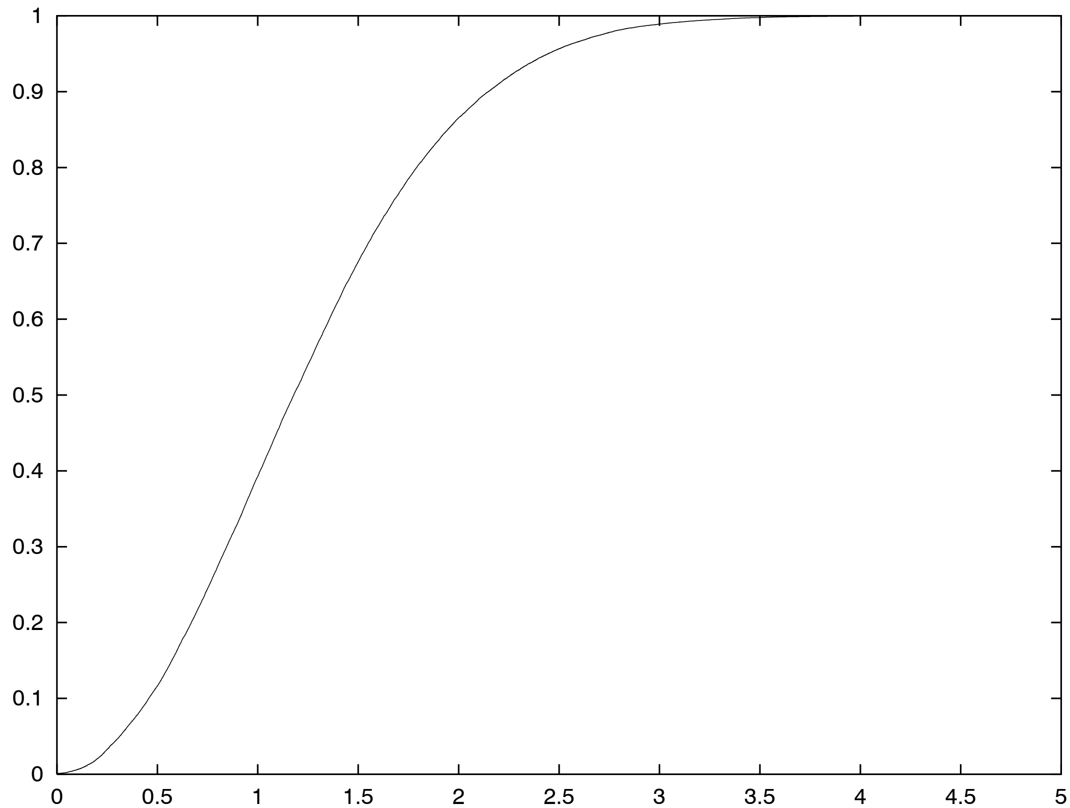

FIGURE 3. Empirical cumulative distribution function for Δ.

It would also be interesting to extend the analysis to the m-dimensional knapsack problem.

REFERENCES

[1] P. Billingsley, *Convergence of Probability Measures*, (Wiley, New York, 1968).

[2] M.E. Dyer and A.M. Frieze, Probabilistic analysis of the multidimensional knapsack problem, *Math. Oper. Res.* **14** (1989) 162–176.

[3] M.R. Garey and D.S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*, (Freeman, San Francisco, 1979).

[4] S. van de Geer and L. Stougie, On rate of convergence and asymptotic normality in the multiknapsack problem, *Mathematical Programming* 51 (1993) 349–358.

[5] A.V. Goldberg and A. Marchetti-Spaccamela, On finding the exact solution of a zero-one knapsack problem, in: *Proc. XVI STOC*, 1984, 359–368.

[6] W. Hoeffding, Probability inequalities for sums of bounded random variables, *American Statistical Association Journal* (1963) 13–30.

[7] O.H. Ibarra and C.E. Kim, Fast approximation algorithms for the knapsack and sum of subset problems, *J. Assoc. Comput. Mach.* 22 (1975) 463–468.

[8] S. Karlin and H. Taylor, *A First Course in Stochastic Processes (2nd ed.)*, (Academic Press, San Diego, 1975).

[9] G.S. Lueker, On the average difference between the solutions to linear and integer knapsack problems, in *Applied Probability - Computer Science: The Interface,* vol. I, R.L. Disney and T.J. Offa (Eds.), Birkhauser, Boston, 1982, 489–504.

[10] G.S. Lueker, Average-case analysis of off-line and on-line knapsack problems, *Journal of Algorithms* **29** (1998) 277–305.
[11] A. Marchetti-Spaccamela and C. Vercellis, Efficient on-line algorithms for the knapsack problem, *Proc. XIV ICALP*, Lecture Notes in Computer Science, vol. 267, Springer Verlag, 1987, 445–456.
[12] M. Meanti, A.H.G. Rinnooy Kan, L. Stougie and C. Vercellis, A probabilistic analysis of the multiknapsack value function, *Mathematical Programming* 46 (1990) 237–247.
[13] S. Sahni, Approximate algorithms for the 0/1 knapsack problem, *J. Assoc. Comput. Mach.* 22 (1975) 115–124.

Department of Computer Science, New Jersey Institute of Technology, Newark, NJ 07102
E-mail address: `calvin@njit.edu, leung@njit.edu`