New Bounds for the Stochastic Knapsack Problem

Ekesh Kumar

June 5, 2020

Abstract

The knapsack problem is a problem in combinatorial optimization that seeks to maximize the objective function $\sum_{i=1}^{n} v_i x_i$ subject to the constraints $\sum_{i=1}^{n} w_i x_i \leq W$ and $x_i \in \{0, 1\}$, where $\mathbf{x}, \mathbf{v} \in \mathbb{R}^n$ and W are provided. We consider the stochastic variant of this problem in which $\mathbf{v}$ remains deterministic, but $\mathbf{x}$ is an n-dimensional vector drawn uniformly at random from $[0, 1]^n$. We establish a sufficient condition under which the summation-bound condition is almost surely satisfied. Furthermore, we discuss the implications of this result on the deterministic problem.

1 Introduction

The classical knapsack problem involves choosing a subset S of n items with values $v_1, \ldots, v_n$ and weights $w_1, \ldots, w_n$ such that the inequality $\sum_{i \in S} w_i \leq W$ is satisfied, and the summation $\sum_{i \in S} v_i$ is maximal. It is well-known that the knapsack problem is weakly NP complete due to a polynomial-time reduction from PARTITION and the existence of an $\mathcal{O}(nW)$ pseudopolynomial dynamic programming algorithm. The stochastic variant of the problem is at least NP-hard since it can easily be generalized to the classical deterministic problem.

In [1], Morton and Wood introduce the same stochastic variant of the knapsack problem. Most notably, they provide various ways to maximize the probability that a total return threshold is met. In [2], Nagarajan summarizes recent results on the stochastic knapsack, and he discusses various extensions to the problem. Importantly, his discussion demonstrates the need to discover links between the deterministic and stochastic knapsack problems since there are concerning their hardness of each of them.

In this paper, we adopt the standard convention of normalizing our weights to satisfy the sum constraint $\sum_{i=1}^{n} w_i = 1$. This is done by simply performing componentwise division by $\sum_{i=1}^{n} w_i$ on the random vector $\mathbf{w}$.

2 Results

Theorem 1. *Let $X_1, X_2, \ldots, X_n \sim \mathcal{U}(0, 1)$ be independent random variables, and define the random sum $X = \sum_{i=1}^{n} X_i$. Furthermore, let $\mu = \mathbb{E}[X]$ denote the expectation of the random sum. For any $0 \leq \delta \leq 1$, we have*

$$P(X > (1 + \delta)\mu) \leq e^{-\delta^2 \mu/3}.$$

Proof. A standard application of the multiplicative Chernoff bound yields

$$P(X > (1 + \delta)\mu) \leq \left(\frac{e^\delta}{(1+\delta)^{1+\delta}} \right)^\mu.$$

However, since this bound is difficult to work with, we will loosen it. A Taylor series expansion of $\log(1 + \delta)$ yields $\frac{2\delta}{2+\delta} \leq \log(1 + \delta)$, from which it immediately follows that

$$\left(\frac{e^\delta}{(1+\delta)^{1+\delta}} \right)^\mu \leq e^{-\delta^2 \mu/(2+\delta)} \leq e^{-\delta^2 \mu/3}.$$

This is the desired result, so we are done. $\square$

Theorem 2. *Let $\mathcal{I} = \{1, 2, \ldots, n\}$ be an indexing set, and let $\mathcal{S} \subseteq \{1, 2, \ldots, n\}$ be a random set of indices drawn uniformly from $\mathcal{P}(\mathcal{I})$. For all $0 \leq \epsilon \leq 1$, the inequality*

$$\left| \sum_{i \in \mathcal{S}} w_i - \frac{1}{2} \right| \geq \epsilon \tag{1}$$

holds with probability p satisfying $p \leq \dfrac{2}{e^{n\epsilon^2/4}}$.

Proof. Let $\mathbf{w}$ be fixed random vector in $\mathbb{R}^n$, and let S be a random sum defined by $S \stackrel{\text{def}}{=} \sum_{i=1}^{n} w_i$, and define $\mu \stackrel{\text{def}}{=} \mathbb{E}[S]$. By linearity of the expectation,

$$\mu = \mathbb{E}[S] = \mathbb{E}\left[\sum_{i=1}^{n} w_i \right] = \sum_{i=1}^{n} \mathbb{E}[w_1] = \frac{n}{2}.$$

Thus, by the preceding theorem, we obtain $P(S < n/4) \leq e^{-n^2/6}$, or equivalently,

$$P(S \geq n/4) \geq 1 - e^{-n^2/6}. \tag{2}$$

Now for a fixed random variable S satisfying Equation (2), the proportion of subsets for which Equation (1) holds is equal exactly to the probability that Equation (1) holds provided that we draw a subset from $\mathcal{P}(\mathcal{I})$ uniformly at random. Now since $\mathbb{E}\left[\sum_{i \in S} w_i \right]$ is equal to the random variable $S/2$, we can apply the preceding theorem a second time in order to deduce

$$P\left(\left| \sum_{i \in S} w_i - \frac{S}{2} \right| \geq \epsilon S \right) \leq 2 e^{-S\epsilon^2}.$$

But since $S \geq n/4$ by assumption, the monotonicity of $f(x) = e^x$ implies the desired result. $\qquad \square$

3   Discussion

The implications of this result are fairly easy to see. Interestingly, this result demonstrates that, for an instance of the stochastic knapsack problem, a random subset of $\{1, 2, \ldots, n\}$ will satisfy the knapsack sum constraint with probability 1 provided that $W \geq 1/2$. Surprisingly, the converse also holds.

This result also has implications in the classical version of the knapsack problem. Particularly, even if $\mathbf{w}$ isn't chosen uniformly at random, our result would still hold as long as our $w_i's$ satisfy $w_i = \mathcal{O}(1/n)$ for every index i. This is a direct consequence of an optimization of the generic Chernoff bound whose derivation is provided in [3].

References

[1] Morton D.P., Wood R.K. (1998) On a Stochastic Knapsack Problem and Generalizations. In: Woodruff D.L. (eds) Advances in Computational and Stochastic Optimization, Logic Programming, and Heuristic Search. Operations Research/Computer Science Interfaces Series, vol 9. Springer, Boston, MA.

[2] Nagarajan, V. (2014, August). Stochastic Knapsack.
`https://theory.stanford.edu/ jvondrak/data/knapsack_focs.pdf`

[3] Wainright, M. (2015) Basic tail and concentration bounds.
`https://www.stat.berkeley.edu/ mjwain/stat210b/Chap2_TailBounds_Jan22_2015.pdf`